A. Preliminaries of Differential Inclusion

Recall that we denote $F(x) = f(x) + g(Ax)$, and Assumption 4 holds. To transit from the smooth case to the nonsmooth case, we use the tool of differential inclusion to build the connection between subdifferentiable $F$ and differentiable functions. One basic example of a differential inclusion takes the form of:

$$\dot{x}(t) \in \partial F(x(t))$$

To bridge the gap between differentiable objective functions and nondifferentiable objective functions, we follow Vassilis et al. (2018) and consider the Moreau-Yosida Approximation, which is a standard tool in convex analysis.

**Definition 15 (Moreau-Yosida Approximation).** Moreau-Yosida Approximation of a convex function $F$ with parameter $\mu > 0$ is defined as

$$F_\mu(x) := \inf_y \left\{ F(y) + \frac{1}{2\mu} \|y - x\|^2 \right\}$$

Use $J_\mu(x)$ to denote the unique point that achieves the infimum above, then $\nabla F_\mu(x) = \frac{1}{\mu}(x - J_\mu(x))$ by the Envelope Theorem (Afriat, 1971; Takayama, 1985). For any $\mu > 0$, $F_\mu$ is a convex, continuously differentiable function.

We take the Definition 3.1 in Vassilis et al. (2018) of a shock solution to define a solution of a differential inclusion. The existence of a shock solution are described in Section 3 of Vassilis et al. (2018). More specifically, we can build a sequence $x_\mu(t)$ such that its subsequence converges, where $x_\mu(t)$ are the solutions to the Approximate Differential Equation (ADE) defined below:

**Approximate Differential Equation (ADE)**

We consider the Moreau-Yosida approximation $F_\mu(x)$ of the objective $F(x)$ with $\mu > 0$. We consider the following approximating ODE:

$$\begin{cases}
\dot{x}_\mu(t) + \nabla F_\mu(x_\mu(t)) = 0 \\
x_\mu(0) = x_0
\end{cases}$$

Here $\nabla F_\mu$ can approximate $\partial F$ and $F_\mu$ is differentiable as is shown in the theory of Moreau-Yosida approximation.

The convergence to a shock solution is described as the Approximation Scheme (AS):

**Approximation Scheme (AS)**

Let $\{F_\mu\}_{\mu > 0}$ be a family of functions such that $F_\mu$ is the Moreau–Yosida approximation of $F$ for all $\mu > 0$. Then there exists a subsequence $\{x_\mu\}_{\mu > 0}$ of solutions of (ADE) that converges to a shock solution $x$ of differential inclusion in the following sense:

- $x_\mu \to x$ uniformly on $[0, T]$ for all $T > 0$ as $\mu \to 0$
- $\dot{x}_\mu \to \dot{x}$ in $L^p([0, T]; \mathbb{R}^d)$ for all $p \in [1, \infty)$ for all $T > 0$ as $\mu \to 0$
- $F_\mu(x_\mu) \to F(x)$ in $L^p([0, T]; \mathbb{R}^d)$ for all $p \in [1, \infty)$ for all $T > 0$ as $\mu \to 0$

B. Proofs of the Theorems Related to Linearized ADMM and Gradient-Based ADMM

In this following sections, we prove the main results provided in Section 2.1. Sections B.1, B.2 and B.3 prove Theorems 5, 6 and 8, respectively.
Differential Inclusions for Modeling Nonsmooth ADMM Variants: A Continuous Limit Theory

B.1. Proof of Theorem 5

Proof of Theorem 5. Due to the strong convexity of the optimization subproblems (5a) and (5b), it is easy to verify that the sequence \( \{x_k, z_k, u_k\} \) is unique. We have from the first-order optimality conditions of (5a) and (5b) that

\[
0 \in \partial f(x_{k+1}) + \tau_L \left[ x_{k+1} - \left( x_k - \frac{\rho}{\tau_L} A^T (A x_k - z_k + u_k) \right) \right], \tag{14a}
\]
\[
0 \in -\frac{1}{\rho} \partial g(z_{k+1}) - (\alpha A x_{k+1} + (1 - \alpha) z_k - z_{k+1} + u_k). \tag{14b}
\]

We detail the proof in the following:

(i) Adding up (14b) and (5c) eliminates the common term \((\alpha A x_{k+1} + (1 - \alpha) z_k - z_{k+1} + u_k)\) and reduces to a simple \(u\)-update:

\[
u_{k+1} + 1 \in \frac{1}{\rho} \partial g(z_{k+1}). \tag{15}\]

Taking the continuous limit \(\rho \to \infty\) gives \(U(t) = 0\), and hence \(\dot{U}(t) = 0\).\(^5\)

(ii) Reorganize (14a) into the following form:

\[
0 \in \partial f(x_{k+1}) + \tau_L (x_{k+1} - x_k) + \rho A^T (A x_k - z_k + u_k). \tag{16}
\]

Bringing (15) into (16) leads to:

\[
0 \in \partial f(x_{k+1}) + A^T \partial g(z_k) + \tau_L (x_{k+1} - x_k) + \rho A^T (A x_k - z_k). \tag{17}
\]

Again from (5c),

\[
u_{k+1} - u_k = \alpha A x_{k+1} + (1 - \alpha) z_k - z_{k+1}
= \alpha A (x_{k+1} - x_k) - (z_{k+1} - z_k) + (\alpha A x_k - z_k),
\]

and hence

\[
A x_k - z_k = \frac{1}{\alpha} [(u_{k+1} - u_k) + (z_{k+1} - z_k)] - A (x_{k+1} - x_k). \tag{18}\]

Plugging (18) into (17) gives

\[
0 \in \partial f(x_{k+1}) + A^T \partial g(z_k) + \tau_L (x_{k+1} - x_k) + \rho A^T \left( \frac{1}{\alpha} [(u_{k+1} - u_k) + (z_{k+1} - z_k)] - A (x_{k+1} - x_k) \right). \tag{19}\]

Taking the limit \(\rho \to \infty\) and letting \(\tau_L / \rho \to c\), using the fact that \(\dot{U}(t) = 0\), (19) reduces to

\[
0 \in \partial f(X(t)) + A^T \partial g(Z(t)) + (c I - A^T A) \dot{X}(t) + \frac{1}{\alpha} A^T \dot{Z}(t). \tag{20}\]

(iii) We directly take the \(\rho \to \infty\) limit in (5c) with the fact that \(u_{k_1} \to u_k\) and \(z_{k+1} \to z_k\), we conclude

\[
Z(t) = AX(t), \quad \dot{Z}(t) = A \dot{X}(t).
\]

It is straightforward to check that

\[
\partial f(X(t)) + A^T \partial g(Z(t)) \subseteq \partial F(X(t)) \tag{21}\]

Combining the above and (20) concludes

\[
0 \in \partial F(X(t)) + \left( c I + \frac{1 - \alpha}{\alpha} A^T A \right) \dot{X}(t),
\]

This completes the proof.

\(^5\)Although the continuous version of \(U(t)\) is constantly zero, it is different with \(u_k = 0\). One may regard \(u_k\) as an infinitesimal number that dynamically changes in the system.
B.2. Proof of Theorem 6

Proof of Theorem 6. Again the sequence \( \{x_k, z_k, u_k\} \) is unique due to the strong convexity of the optimization subproblem (5a) and (5b). It follows from the optimality conditions that

\[
\begin{align*}
0 &= \nabla f(x_k) + \rho A^T (A x_k - z_k + u_k) + \tau_G (x_{k+1} - x_k), \\
0 &\in \partial g(z_{k+1}) - \rho (\alpha A x_{k+1} + (1 - \alpha) z_k - z_{k+1} + u_k), \\
u_{k+1} &= u_k + (\alpha A x_{k+1} + (1 - \alpha) z_k - z_{k+1}).
\end{align*}
\]

(22a) (22b) (22c)

Seeing \( \tau_L \) in the place of \( \tau_G \), (22b) and (22c) are identical to (14b) and (5c), while (22a) is identical to (14a) with \( \partial f(x_{k+1}) \) replaced by \( \nabla f(x_k) \).

Carrying out the proof of Theorem 5 in §B.1 gives (19) with \( \partial f(x_{k+1}) \) replaced by \( \nabla f(x_k) \), and hence taking corresponding limits gives differential inclusion (20) with \( \partial f(X(t)) \) replaced by \( \nabla f(X(t)) \). The rest of the proof follows in the same fashion as Part (iii) in the proof of Theorem 5.

\[ \square \]

B.3. Proof of Theorem 8

Proof of Theorem 8. For notation simplicity, we choose a matrix \( B \) such that \( B^T B = cI + \frac{1 - \alpha}{\alpha} A^T A \). Recall that the largest and smallest singular value of \( B \) are \( \kappa_1 \) and \( \kappa_d \). Note that when \( 0 < \alpha \leq 1 \), \( \kappa_1 = \sqrt{c + \frac{1 - \alpha}{\alpha} \sigma_1^2} \) and \( \kappa_d = \sqrt{c + \frac{1 - \alpha}{\alpha} \sigma_d^2} \), and when \( 1 < \alpha < 2 \), \( \kappa_1 = \sqrt{c + \frac{1 - \alpha}{\alpha} \sigma_1^2} \) and \( \kappa_d = \sqrt{c + \frac{1 - \alpha}{\alpha} \sigma_d^2} \), where \( \sigma_1, \sigma_d \) are singular value of matrix \( A \). Then the original differential inclusion becomes \( 0 \in \partial F(X(t)) + (B^T B) \dot{X}(t) \). Because Moreau-Yosida approximation \( F_\mu(X_\mu(t)) \) is a continuously differentiable, convex function for all \( \mu > 0 \), we denote an arbitrary minimizer as \( x_\mu^* \).

For each \( \mu > 0 \), consider the energy functional of Moreau-Yosida approximation defined as

\[
\mathcal{E}_\mu(t) = t(F_\mu(X_\mu(t)) - F_\mu(x_\mu^*)) + \frac{\lambda}{2} \|B(X_\mu(t) - x_\mu^*)\|^2.
\]

(23)

where \( \lambda \) is an arbitrary constant greater than or equal to 1. Because \( F_\mu \) is a continuously differentiable function, we could write the time derivative of \( \dot{\mathcal{E}}_\mu(t) \) as

\[
\dot{\mathcal{E}}_\mu(t) = (F_\mu(X_\mu(t)) - F_\mu(x_\mu^*)) + t(\nabla F_\mu(X_\mu(t)), \dot{X}_\mu(t)) + \lambda (B^T B)(X_\mu(t) - x_\mu^*), \dot{X}_\mu(t)).
\]

(24)

By substituting \( B^T B \dot{X}_\mu(t) \) by \( -\nabla F_\mu(X_\mu(t)) \) and \( \nabla F_\mu(X_\mu(t)) \) by \( -B^T B \dot{X}_\mu(t) \) according to (9) and the definition of the shock solution \( X_\mu(t) \) in Appendix A, we have

\[
\dot{\mathcal{E}}_\mu(t) = -t\|B \dot{X}_\mu(t)\|^2 + (F_\mu(X_\mu(t)) - F_\mu(x_\mu^*)) - \lambda (\langle X_\mu(t) - x_\mu^* \rangle, \nabla F_\mu(X_\mu(t))) \leq 0,
\]

(25)

where we used the convexity of \( F_\mu \) and nonnegativity of \( (F_\mu(X_\mu(t)) - F_\mu(x_\mu^*)) \), \( \|B \dot{X}_\mu\|_2 \) in the last inequality.

Similar to \( \dot{\mathcal{E}}_\mu(t) \), we define the energy functional for \( F(X(t)) \) as

\[
\mathcal{E}(t) = t(F(X(t)) - F(x^*)) + \frac{\lambda}{2} \|B(X(t) - x^*)\|^2.
\]

(26)

At time \( t = 0 \), there is an upper bound on \( \mathcal{E}(0) \) as

\[
\mathcal{E}(0) = \frac{\lambda}{2} \|B(x_0 - x^*)\|^2 \leq \frac{\lambda \kappa_d^2}{2} \|x_0 - x^*\|^2.
\]

(27)

By applying the approximation scheme (AS) argument (details as in Appendix A) as \( \mu \to 0 \), we have for a.e. \( t \geq 0 \) that \( \mathcal{E}(t) \leq \mathcal{E}(0) \).

By non-negativity of \( F(X) - F(x^*) \) in (26), we find

\[
\frac{\lambda \kappa_d^2}{2} \|X(t) - x^*\|^2 \leq \mathcal{E}(0).
\]

(28)
Combining with the upper bound of $\mathcal{E}(0)$ in (27), we derive for a.e. $t \geq 0$ that

$$\|X(t) - x^*\|_2 \leq \frac{\kappa_1}{\kappa_d} ||x_0 - x^*||_2. \quad (29)$$

Using the nonnegativity of all terms in (26) and monotonicity of $\mathcal{E}(t)$ on a.e. $t \geq 0$, we have, for a.e. $t \geq 0$,

$$t(F(X(t)) - F(x^*)) \leq \mathcal{E}(t) \leq \mathcal{E}(0) \leq \frac{\kappa_2^2}{2t} \|x_0 - x^*\|_2^2 \quad (30)$$

Choosing $\lambda = 1$, we have the following result, for a.e. $t \geq 0$,

$$F(X(t)) - F(x^*) \leq \mathcal{E}(t) \leq \mathcal{E}(0) \leq \frac{\kappa_2^2}{2t} \|x_0 - x^*\|_2^2 \quad (31)$$

By applying convexity of $F_\mu$ to (25), we have

$$\dot{\mathcal{E}}(t) \leq (1 - \lambda)(F_\mu(X_\mu(t)) - F_\mu(x_\mu^*)) - t\|\dot{X}_\mu(t)\|_2^2. \quad (32)$$

Notice that the two terms in (32) are all negative, we find

$$F_\mu(X_\mu(t)) - F_\mu(x_\mu^*) \leq \frac{\dot{\mathcal{E}}(t)}{\lambda - 1} \quad \text{and} \quad t\|\dot{X}_\mu(t)\|_2^2 \leq -\frac{\dot{\mathcal{E}}(t)}{\kappa_d^2}. \quad (33)$$

By integrating over $(0, T)$, the inequalities above give for all $T > 0$ that

$$\int_0^T (F_\mu(X_\mu(t)) - F_\mu(x_\mu^*))dt \leq \frac{\mathcal{E}(0)}{\lambda - 1}, \quad \int_0^T t\|\dot{X}_\mu(t)\|_2^2 dt \leq \frac{\mathcal{E}(0)}{\kappa_d^2}. \quad (34)$$

By applying approximation scheme (AS), taking limit $T \to \infty$, choosing $\lambda \to \infty$ and $\lambda = 1$ respectively, and plugging in (27), we have

$$\int_0^\infty (F(X_\mu(t)) - F(x^*))dt \leq \frac{\kappa_1^2}{2} \|x_0 - x^*\|_2^2, \quad \int_0^\infty t\|\dot{X}(t)\|_2^2 dt \leq \frac{\kappa_1^2}{2\kappa_d^2} \|x_0 - x^*\|_2^2. \quad (35)$$

\[\square\]

**C. Proofs of the Theorems Related to G-ADMM and the Accelerated G-ADMM**

**C.1. Proof of Theorem 9**

**Proof of Theorem 9.** Proof of Theorem 9 uses idea similar to the proof of Theorem 5 to analyze G-ADMM updates. By strong convexity of the optimization subproblems (10a) and (10b), we could verify that the sequence $\{x_k, z_k, u_k\}$ is unique. Together with (10c), we have from the first-order optimality conditions of (10a) and (10b) that

$$\partial f(x_{k+1}) + \rho A^T(Ax_{k+1} - z_k + u_k) \ni 0, \quad (36a)$$

$$\frac{1}{\rho}\partial g(z_{k+1}) - (\alpha Ax_{k+1} + (1 - \alpha)z_k - z_{k+1} + u_k) \ni 0, \quad (36b)$$

$$u_{k+1} = \alpha Ax_{k+1} + (1 - \alpha)z_k - z_{k+1} + u_k = 0. \quad (36c)$$

Adding up (36b) and (36c) eliminates the common term $(\alpha Ax_{k+1} + (1 - \alpha)z_k - z_{k+1} + u_k)$ and reduces to a simple $u$-update:

$$u_{k+1} \in \frac{1}{\rho}\partial g(z_{k+1}). \quad (37)$$
Taking the continuous limit \( \rho \to \infty \) gives \( U(t) = 0 \), and hence \( \dot{U}(t) = 0 \).

Bringing (37) into (36a) leads to:

\[
0 \in \partial f(x_{k+1}) + A^T \partial g(z_k) + \rho A^T (Ax_{k+1} - z_k),
\]

where again from (36c),

\[
u_{k+1} - u_k = \alpha (Ax_{k+1} - z_k) - (z_{k+1} - z_k),
\]

and hence

\[
Ax_{k+1} - z_k = \frac{1}{\alpha} [(u_{k+1} - u_k) + (z_{k+1} - z_k)]
\]

Plugging (39) into (38) gives

\[
0 \in \partial f(x_{k+1}) + A^T \partial g(z_k) + \rho A^T \left( \frac{1}{\alpha} [(u_{k+1} - u_k) + (z_{k+1} - z_k)] \right).
\]

Taking the limit \( \rho \to \infty \), using the fact that \( \dot{U}(t) = 0 \), (40) reduces to

\[
0 \in \partial f(X(t)) + A^T \partial g(Z(t)) + \frac{1}{\alpha} A^T (\dot{Z}(t)).
\]

We directly take the \( \rho \to \infty \) limit in (36c) and conclude

\[
Z(t) = AX(t), \quad \dot{Z}(t) = A\dot{X}(t).
\]

Recalling (21) and combining the above with (41) concludes

\[
0 \in \partial F(X(t)) + \left( \frac{1}{\alpha} A^T A \right) \dot{X}(t),
\]

Thus we complete the proof.

\[ \Box \]

C.2. Proof of Theorem 10

Proof of Theorem 10. For each \( \mu > 0 \), consider the energy functional of Moreau-Yosida approximation defined as

\[
\mathcal{E}_\mu(t) = \alpha t(F_\mu(X_\mu(t)) - F_\mu(x_\mu^*)) + \frac{\lambda}{2} \| A(X_\mu(t) - x_\mu^*) \|^2,
\]

where \( \lambda \) is an arbitrary constant chosen as \( \lambda \geq 1 \) and \( x_\mu^* \) denotes the minimizer of \( F_\mu \). Because \( F_\mu \) is a continuously differentiable function, we could write the time derivative of \( \mathcal{E}_\mu(t) \) as

\[
\dot{\mathcal{E}}_\mu(t) = \alpha (F_\mu(X_\mu(t)) - F_\mu(x_\mu^*)) + \alpha \langle \nabla F_\mu(X_\mu(t)), \dot{X}_\mu(t) \rangle + \lambda \langle A^T A(X_\mu(t) - x_\mu^*), \dot{X}_\mu(t) \rangle.
\]

By using the equality of \( A^T A \dot{X}_\mu(t) \) and \( -\alpha \nabla F_\mu(X_\mu(t)) \), we have

\[
\dot{\mathcal{E}}_\mu(t) = -t \| A\dot{X}_\mu(t) \|^2 + \alpha (F_\mu(X_\mu(t)) - F_\mu(x_\mu^*)) - \lambda \alpha (\langle X_\mu(t) - x_\mu^*, \nabla F_\mu(X_\mu(t)) \rangle) \leq 0,
\]

where we used the convexity of \( F_\mu \) and nonnegativity of \( (F_\mu(X_\mu(t)) - F(x_\mu^*)), \| A\dot{X}_\mu \|_2 \) in the last inequality.

Similar to \( \mathcal{E}_\mu(t) \), we define the energy functional for \( F(X(t)) \) as

\[
\mathcal{E}(t) = \alpha t(F(X(t)) - F(x^*)) + \frac{\lambda}{2} \| A(X(t) - x^*) \|^2.
\]

At time 0, there is an upper bound on \( \mathcal{E}(0) \) as

\[
\mathcal{E}(0) = \frac{\lambda}{2} \| A(X(0) - x^*) \|^2 \leq \frac{\lambda\sigma^2}{2} \| x_0 - x^* \|^2.
\]
By applying the approximation scheme (AS) argument (details as in Appendix A) as $\mu \to 0$ to equation (44), we have for a.e. $t \geq 0$, $\mathcal{E}(t) \leq 0$ and that $\mathcal{E}(t) \leq \mathcal{E}(0)$.

In (45), by non-negativity of $F(X(t)) - F(x^*)$ and $\|X(t) - x^*\|^2_2$, we find

$$\frac{\lambda}{2}\|A(X(t) - x^*)\|^2_2 \leq \mathcal{E}(0).$$

(47)

Combining with the upper bound of $\mathcal{E}(0)$ in (46), and by taking $\lambda = 1$, we derive for a.e. $t \geq 0$ that

$$\|X(t) - x^*\|^2_2 \leq \frac{\sigma_d^2}{2}\|x_0 - x^*\|^2_2.$$

(48)

Using the nonnegativity of all terms in (45) and the convexity of $\mathcal{E}(t)$ on a.e. $t \geq 0$, we have

$$\alpha t(F(X(t)) - F(x^*)) \leq \mathcal{E}(t) \leq \mathcal{E}(0) \leq \frac{\lambda\sigma_d^2}{2}\|x_0 - x^*\|^2_2 \text{ for a.e. } t,$$

(49)

which is given by (46). Thus $(F(X(t)) - F(x^*)) \leq \frac{\sigma_d^2}{2\alpha^2}\|x_0 - x^*\|^2_2$ by taking $\lambda = 1$.

From (44) and using the convexity of $F_\mu$, we have

$$\dot{\mathcal{E}}_\mu(t) \leq \alpha(1 - \lambda)(F_\mu(X_\mu(t)) - F_\mu(x^*_\mu)) - t\|A\dot{X}_\mu(t)\|^2_2.$$

(50)

Notice that the two terms in (50) are all negative, we find

$$F_\mu(X_\mu(t)) - F_\mu(x^*_\mu) \leq -\frac{\dot{\mathcal{E}}_\mu(t)}{\alpha(1 - \lambda)} \quad \text{and} \quad t\|A\dot{X}_\mu(t)\|^2_2 \leq -\dot{\mathcal{E}}_\mu(t).$$

(51)

By integrating over $(0, T)$, the inequalities above give

$$\int_0^T (F_\mu(X_\mu(t)) - F_\mu(x^*_\mu))dt \leq \frac{\mathcal{E}_\mu(0)}{\alpha(1 - \lambda)}, \quad \int_0^T t\|A\dot{X}_\mu(t)\|^2_2dt \leq \mathcal{E}_\mu(0).$$

(52)

By applying approximation scheme (AS) and plugging in (46), we have

$$\int_0^T (F_\mu(X_\mu(t)) - F_\mu(x^*_\mu))dt \leq \frac{\lambda\sigma_d^2}{2\alpha^2}\|x_0 - x^*\|^2_2, \quad \int_0^T t\|A\dot{X}_\mu(t)\|^2_2dt \leq \frac{\lambda\sigma_d^2}{2}\|x_0 - x^*\|^2_2.$$

(53)

Taking the limit when $\mu \to 0, T \to \infty$ and choosing $\lambda \to \infty$ and $\lambda = 1$ respectively, we get

$$\int_0^\infty (F(X(t)) - F(x^*))dt \leq \frac{\sigma_d^2}{2\alpha}\|x_0 - x^*\|^2_2, \quad \int_0^\infty t\|\dot{X}(t)\|^2_2dt \leq \frac{\sigma_d^2}{2}\|x_0 - x^*\|^2_2.$$

(54)

This completes our proof.

C.3. Proof of Theorem 11

Proof of Theorem 11. To analyze Accelerated G-ADMM, we adopt idea similar to proof of Theorem 5. Using strong convexity of the optimization subproblems (12a) and (12b), we know that the sequence $\{x_k, z_k, u_k, \hat{u}_k, \hat{z}_k\}$ is unique. Together with (12c), we have from the first-order optimality conditions of (12a) and (12b) that

$$\partial f(x_{k+1}) + \rho A^T(Ax_{k+1} - \hat{z}_k + \hat{u}_k) \geq 0,$$

$$\frac{1}{\rho} \partial g(z_{k+1}) - (\alpha Ax_{k+1} + (1 - \alpha)\hat{z}_k - z_{k+1} + \hat{u}_k) \geq 0,$$

$$u_{k+1} - (\alpha Ax_{k+1} + (1 - \alpha)\hat{z}_k - z_{k+1} + \hat{u}_k) = 0.$$

(55a, 55b, 55c)
Yosida approximation defined as

\[ u_{k+1} = \frac{1}{\rho} \partial g(z_{k+1}). \] (56)

Taking the continuous limit \( \rho \to \infty \) gives \( U(t) = 0 \), and hence \( \dot{U}(t) = 0, \ddot{U}(t) = 0 \). The idea is similar to the proof of Theorem 5.

Bringing (56) and equation (12d) which is the definition of \( \ddot{u} \) into (55a) leads to:

\[ \dot{f}(x_{k+1}) + A^T \partial g(z_k) + \rho A^T (Ax_{k+1} - z_k) + \rho g_k + A^T (u_{k+1} - u_k) \geq 0, \] (57)

where again from (55c),

\[ (Ax_{k+1} - z_k) = \frac{1}{\alpha}[(u_{k+1} - \hat{u_k}) + (z_{k+1} - \hat{z}_k)]. \] (58)

In addition, from equation (12d) and equation (12e), we find that

\[ u_{k+1} - \hat{u_k} = u_{k+1} - (1 + \gamma_k)u_k + \gamma_k u_{k-1} \]

\[ z_{k+1} - \hat{z}_k = z_{k+1} - (1 + \gamma_z)z_k + \gamma_z z_{k-1}. \]

For \( u_{k+1} - \hat{u_k} \), we add the term \( u_k - \hat{u_k} + u_{k-1} - u_{k-1} \) to the right hand side, the resulting equation is a combination of the second order difference and first order difference of the sequence \( \{u_k\} \):

\[ u_{k+1} - \hat{u_k} = (u_{k+1} - 2u_k + u_{k-1}) + (1 - \gamma_k)(u_k - u_{k-1}). \] (59)

Similarly, the equation holds that:

\[ z_{k+1} - \hat{z}_k = (z_{k+1} - 2z_k + z_{k-1}) + (1 - \gamma_z)(z_k - z_{k-1}). \] (60)

We note that \( 1 - \gamma_k = 1 - \frac{k}{k+r} = \frac{\rho \alpha}{\rho \alpha t+1} \). Taking the limit \( \rho \to \infty \), under infinitesimal step sizes, using relationships (58), (59), (60) and the fact that \( \ddot{U}(t) = 0, \dot{U}(t) = 0 \), equation (57) becomes:

\[ \dot{f}(X(t)) + A^T \partial g(Z(t)) + \frac{1}{\alpha} A^T (\frac{\rho}{t} \dot{Z}(t) + \ddot{Z}(t)) \geq 0. \] (61)

We directly take the \( \rho \to \infty \) limit in (55c) and conclude

\[ Z(t) = AX(t), \quad \dot{Z}(t) = A\dot{X}(t), \quad \ddot{Z}(t) = A\ddot{X}(t). \]

Recalling (21) and combining the above with (61) concludes

\[ 0 \in \partial F(X(t)) + \left( \frac{1}{\alpha} A^T A \right) (\dot{X}(t) + \frac{\rho}{t} \ddot{X}(t)), \]

\[ \square \]

C.4. Proof of Theorem 12

Proof of Theorem 12. Recall that \( x^*_\mu \) is the minimizer of \( F_\mu \). For each \( \mu > 0 \), consider the energy functional of Moreau-Yosida approximation defined as

\[ E_\mu(t) = t^2 (F_\mu(X_\mu(t)) - F_\mu(x^*_\mu)) + \frac{1}{2\alpha} \left\| A \left( \lambda (X_\mu(t) - x^*_\mu) + \lambda t \dot{X}_\mu(t) \right) \right\|_2^2 + \frac{\lambda(r-\lambda-1)}{2\alpha} \| A(X_\mu(t) - x^*_\mu) \|_2^2 \] (62)

where \( \lambda \) is a constant chosen within \( 2 \leq \lambda \leq r - 1 \). Because \( F_\mu \) is a continuously differentiable function, we could write the time derivative of \( E_\mu(t) \) as

\[ \dot{E}_\mu = 2t(F_\mu(X_\mu(t)) - F_\mu(x^*_\mu)) + t^2 \nabla F_\mu(X_\mu(t))^T \dot{X}_\mu + \left( \lambda (X_\mu - x^*_\mu) + t \dot{X}_\mu \right)^T \left( \frac{1}{\alpha} A^T A \right) \left( \lambda (X_\mu - x^*_\mu) + t \dot{X}_\mu \right) + \lambda(r-\lambda-1)(X_\mu - x^*_\mu)^T \left( \frac{1}{\alpha} A^T A \right) \dot{X}_\mu \]
By using the equality of $tA^\top AX_\mu$ and $-rA^\top AX_\mu - \alpha t\nabla F_\mu(X_\mu(t))$, we have
\[\dot{E}_\mu = -\lambda t (F_\mu(x^*_\mu) - F_\mu(X_\mu(t))) - (x^*_\mu - X_\mu)^\top \nabla F_\mu(X_\mu(t))) - (r-2)t(F_\mu(X_\mu(t)) - F_\mu(x^*_\mu)) - \frac{(r-1-\lambda)t}{\alpha} \|A\dot{X}_\mu\|_2^2 \leq 0\] (63)
where we used the convexity of $F_\mu$ and nonnegativity of $F_\mu(X_\mu(t)) - F_\mu(x^*_\mu)$, $\|A\dot{X}_\mu\|_2$ in the last inequality.

Similar to $E_\mu(t)$, we define the energy functional for $F(X(t))$ as
\[E(t) = t^2(F(X(t)) - F(x^*)) + \frac{1}{2\alpha} \|A(\lambda(X(t) - x^*) + t\dot{X}(t))\|^2_2 + \frac{\lambda(r-1)}{2\alpha} \|A(X(t) - x^*)\|^2_2\]
At time $t_0$, there is an upper bound on $E(t_0)$ as
\[E(t_0) = t_0^2(F(x_0) - F(x^*)) + \frac{\lambda(r-1)}{2\alpha} \|A(x_0 - x^*)\|^2_2 \leq \frac{2\alpha + \lambda(r-1)\sigma_d^2}{2\alpha} \Delta_0^2\] (64)
By non-negativity of $F_\mu(X_\mu(t)) - F_\mu(x^*_\mu)$, $\|X_\mu - x^*_\mu\|^2_2$ and $\|\dot{X}_\mu\|^2_2$, we find for all $r \geq 3$ and $t \geq t_0$ that
\[\frac{d}{dt}(t\|X_\mu - x^*_\mu\|^2_2) = \|X_\mu - x^*_\mu\|^2_2 + 2t(X_\mu - x^*_\mu)^\top \dot{X}_\mu \leq \frac{1}{2} \|2(X_\mu - x^*_\mu) + t\dot{X}_\mu\|^2_2 \leq \frac{\alpha E_\mu(t_0)}{\sigma_d^2} \leq \frac{\alpha E_\mu(t_0)}{\sigma_d^2}\]
By integrating over $(t_0, t)$, this gives us
\[t\|X_\mu - x^*_\mu\|^2_2 - t_0\|x_0 - x^*_\mu\|^2_2 \leq \frac{\alpha E_\mu(t_0)}{\sigma_d^2}\] (65)
By applying the approximation scheme (AS) argument (details as in Appendix A) as $\mu \to 0$, we have for a.e. $t \geq t_0$ that
\[\|X - x^*\|^2_2 \leq \frac{\alpha E(t_0)}{\sigma_d^2} + \|x_0 - x^*\|^2_2\]
Combining with the upper bound of $E(t_0)$ in (64), we derive for a.e. $t \geq t_0$ that
\[\|X(t) - x^*\|^2_2 \leq C_1\Delta_0\]
with factor $C_1 = \sqrt{\frac{\alpha + (r-1)\sigma_d^2}{\sigma_d^2}}$. Here we choose $\lambda = 2$ to minimize $C_1$.

From (63), we know that $E_\mu(t)$ is nonincreasing for $t \geq t_0$, for all $\mu > 0$. By applying (AS) we find that $E(t)$ is nonincreasing for a.e. $t \geq t_0$. Using the nonnegativity of all three terms in (62) and monotonicity of $E(t)$ on a.e. $t \geq t_0$, we have for a.e. $t \geq t_0$ that
\[F(X(t)) - F(x^*) \leq \frac{1}{t^2} E(t) \leq \frac{1}{t^2} E(t_0) \leq \frac{C_2}{t^2} \Delta_0^2\]
where factor $C_2 = 1 + (r-1)\sigma_d^2/\alpha$ is given by (64) with $\lambda = 2$, and
\[\|\lambda(X(t) - x^*) + t\dot{X}(t)\|^2_2 \leq \frac{2\alpha}{\sigma_d^2} E(t) \leq \frac{2\alpha}{\sigma_d^2} E(t_0) \leq \frac{\alpha E(t_0)}{\sigma_d^2} \leq \frac{\alpha E(t_0)}{\sigma_d^2} \Delta_0^2\]
Therefore, by triangle inequality and (65),
\[\|\dot{X}(t)\|^2_2 \leq \frac{1}{t} \|\lambda(X(t) - x^*) + t\dot{X}(t)\|^2_2 + \frac{1}{t} \|\lambda(X(t) - x^*)\|^2_2 \leq C_3 \Delta_0\]
with factor $C_3 = \sqrt{\frac{2\alpha + 2(r-1)\sigma_d^2}{\sigma_d^2}} + 2C_1$. Here we choose $\lambda = 2$ to minimize $C_3$.

From (63), we have
\[\dot{E}_\mu \leq -(\lambda - 2)t(F_\mu(X_\mu(t)) - F_\mu(x^*_\mu)) - \frac{(r-1-\lambda)t}{\alpha} \|A\dot{X}_\mu\|^2_2\]
when $r = 3$, we could only choose $\lambda = 2 = r - 1$ and the right hand side of the inequality above is always zero. However, if we further assume $r > 3$, then we could choose $\lambda = r - 1$ and $\lambda = 2$ respectively, such that

$$t(F'_{\mu}(x\mu(t)) - F'_{\mu}(x^*\mu)) \leq -\frac{1}{r - 3} \dot{\mu} \text{ and } t\|\dot{X}_\mu\|^2 \leq -\frac{\alpha}{(r - 3)\sigma_d^2} \dot{\mu}$$

By integrating over $(t, \infty)$, the inequalities above give

$$\int_{t_0}^{\infty} t(F'_{\mu}(X(t)) - F'_{\mu}(x^*))(t)dt \leq \frac{\dot{\mu}(t_0)}{r - 3} \text{ and } \int_{t_0}^{\infty} t\|X(t)\|^2dt \leq \frac{\alpha\dot{\mu}(t_0)}{r - 3}\sigma_d^2$$

By applying (A5) and plugging in (64), we have

$$\int_{t_0}^{\infty} t(F(X(t)) - F(x^*))dt \leq C_4\Delta_0^2 \text{ and } \int_{t_0}^{\infty} t\|\dot{X}(t)\|^2dt \leq C_5\Delta_0^2$$

with factors $C_4 = \frac{2\alpha + (r-1)^2}{2(r-3)\alpha}$ and $C_5 = \frac{\alpha + (r-1)^2}{(r-3)\sigma_d^2}$. \hfill \Box

C.5. Proof of Theorem 14

Proof of Theorem 14. The energy functional we used in Theorem 12 is no longer applicable, because we can not find $\lambda$ satisfying $\lambda - 2 \geq 0$ and $r - 1 - \lambda \geq 0$ simultaneously when $0 < r < 3$. Here we consider a new energy functional for the Moreau-Yosida approximation

$$E_{\mu}(t) = t^2(F'_{\mu}(X(t)) - F'_{\mu}(x^*)) + \frac{1}{2\alpha} \left\| \frac{2r}{3} A(X(t) - x^*) + tA\dot{X}_\mu(t) \right\|^2 + \frac{r(3 - r)}{9\alpha} \|A(X(t) - x^*)\|^2$$

By taking its time derivative, we have

$$\dot{E}_{\mu} = 2t(F'_{\mu}(X(t)) - F'_{\mu}(x^*)) + t^2\nabla F'_{\mu}(X(t))\dot{X}_\mu + \left( \frac{2r}{3}(X(t) - x^*) + t\dot{X}_\mu \right) \frac{1}{\alpha A^\top A} \left( \frac{2r}{3} + 1 \right) X_\mu + t\dot{X}_\mu$$

By using the equality of $tA^\top A\dot{X}_\mu$ and $-rA^\top A\dot{X}_\mu - \alpha t\nabla F'_{\mu}(X(t))$ and applying the convexity of $F'_{\mu}$, we have

$$\dot{E}_{\mu} \leq 2\frac{(3 - r)}{3} t(F'_{\mu}(X(t)) - F'_{\mu}(x^*)) + \frac{4r(3 - r)}{9\alpha} (X(t) - x^*)^\top A^\top A\dot{X}_\mu + \frac{3 - r}{3\alpha} t\|A\dot{X}_\mu\|^2$$

Although this energy functional does not have nonnegative derivative, there is a special relationship between it and its derivative. We notice that

$$\dot{E}_{\mu} - 2\frac{(3 - r)}{3t} E_{\mu} \leq -\frac{2r(3 - r)(3 + r)}{27\alpha t} \|A(X(t) - x^*)\|^2 \leq 0$$

This implies that, for $H_{\mu}(t) := t^{-\frac{2(3-r)}{3}} E_{\mu}(t)$, for all $t \geq t_0$,

$$H_{\mu} = t^{-\frac{2(3-r)}{3}} \cdot (\dot{E}_{\mu} - 2\frac{(3 - r)}{3t} E_{\mu}) \leq 0$$

Therefore, $H_{\mu}(t)$ is nonincreasing over $t \geq t_0$, for all $\mu > 0$. By making similar definition as $H(t) := t^{-\frac{2(3-r)}{3}} E(t)$ and applying the approximation scheme, we have that $H(t)$ is nonincreasing for a.e. $t \geq t_0$. At time $t_0$,

$$H(t_0) \leq t_0^{-\frac{2(3-r)}{3}} \left( 1 + \frac{r(3 + r)}{9\alpha} \sigma_d^2 \right) \Delta_0^2$$

By the nonnegativity of all terms in (66) and the monotonicity of $H(t)$, we have for a.e. $t \geq t_0$ that

$$F(X(t)) - F(x^*) \leq \frac{1}{t^{\frac{2(3-r)}{3}}} H(t) \leq \frac{1}{t^{\frac{2(3-r)}{3}}} H(t_0) \leq C_6 t_0^{-\frac{2(3-r)}{3}} \Delta_0^2$$
Differential Inclusions for Modeling Nonsmooth ADMM Variants: A Continuous Limit Theory

with factor $C_6 = 1 + \frac{r(3+r)\sigma^2}{9\alpha}$.

Similarly, we have for a.e. $t \geq t_0$ that

$$\left\| \frac{2r}{3} (X(t) - x^*) + t\dot{X} \right\|^2 \leq \frac{2\alpha}{\sigma_d^2} t^{\frac{2(3-r)}{3}} \mathcal{H}(t) \leq \frac{2\alpha}{\sigma_d^2} t^{\frac{2(3-r)}{3}} \mathcal{H}(t_0) \leq \frac{2\alpha}{\sigma_d^2} C_6 t^{\frac{2(3-r)}{3}} \Delta_0^2$$

If we also assume the trajectory $\{X(t)\}_{t \geq t_0}$ is bounded, then by adopting the same interpretation as in Theorem 12, there exists some positive factor $C_0$ such that, for a.e. $t \geq t_0$, $\|X(t) - x^*\| \leq C_0 \Delta_0$. Then triangle inequality gives us, for a.e. $t \geq t_0$, that

$$\|\dot{X}\| \leq \frac{1}{t} \left\| \frac{2r}{3} (X(t) - x^*) + t\dot{X} \right\| + \frac{2r}{3t} \|X(t) - x^*\| \leq \frac{C_7 t_0^{-\frac{3-r}{2}} \Delta_0}{t^2}$$

with factor $C_7 = \sqrt{\frac{2\alpha C_6}{\sigma_d^2}} + \frac{2r}{3} C_0$. \qed