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*

$$F_\mu(x) := \inf_y \left\{ F(y) + \frac{1}{2\mu}\|y - x\|_2^2 \right\}$$

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_μ 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 ∇F_μ can approximate ∂F and F_μ is differentiable as is shown in the theory of Moreau-Yosida approximation.

The convergence to a shock solution is described as the Approximation Scheme ($\mathcal{AS}$):

Approximation Scheme ($\mathcal{AS}$)

Let $\{F_\mu\}_{\mu>0}$ be a family of functions such that F_μ 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.

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

$$0 \in \partial f(x_{k+1}) + \tau_L \left[x_{k+1} - \left(x_k - \frac{\rho}{\tau_L} A^\top (Ax_k - z_k + u_k) \right) \right], \tag{14a}$$

$$0 \in \frac{1}{\rho} \partial g(z_{k+1}) - (\alpha A x_{k+1} + (1-\alpha) z_k - z_{k+1} + u_k). \tag{14b}$$

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:

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

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:

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

Bringing (15) into (16) leads to:

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

Again from (5c),

$$\begin{aligned} u_{k+1} - u_k &= \alpha A x_{k+1} + (1-\alpha) z_k - z_{k+1} \\ &= \alpha A (x_{k+1} - x_k) - (z_{k+1} - z_k) + \alpha (Ax_k - z_k), \end{aligned}$$

and hence

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

Plugging (18) into (17) gives

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

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

$$0 \in \partial f(X(t)) + A^\top \partial g(Z(t)) + \left(cI - A^\top A \right) \dot{X}(t) + \frac{1}{\alpha} A^\top \dot{Z}(t). \tag{20}$$

(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

$$Z(t) = AX(t), \qquad \dot{Z}(t) = A\dot{X}(t).$$

It is straightforward to check that

$$\partial f(X(t)) + A^\top \partial g(Z(t)) \subseteq \partial F(X(t)) \tag{21}$$

Combining the above and (20) concludes

$$0 \in \partial F(X(t)) + \left(cI + \frac{1-\alpha}{\alpha} A^\top A \right) \dot{X}(t),$$

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

$$0 = \nabla f(x_k) + \rho A^T(Ax_k - z_k + u_k) + \tau_G(x_{k+1} - x_k), \tag{22a}$$

$$0 \in \partial g(z_{k+1}) - \rho(\alpha A x_{k+1} + (1-\alpha)z_k - z_{k+1} + u_k), \tag{22b}$$

$$u_{k+1} = u_k + (\alpha A x_{k+1} + (1-\alpha)z_k - z_{k+1}). \tag{22c}$$

Seeing τ_L in the place of τ_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.

□

B.3. Proof of Theorem 8

Proof of Theorem 8. For notation simplicity, we choose a matrix B such that $B^\top B = cI + \frac{1-\alpha}{\alpha}A^\top A$. Recall that the largest and smallest singular value of B are κ_1 and κ_d. Note that when $0 < \alpha \le 1$, $\kappa_1 = \sqrt{c + \frac{1-\alpha}{\alpha}\sigma_1^2}$ and $\kappa_d = \sqrt{c + \frac{1-\alpha}{\alpha}\sigma_d^2}$, and when $1 < \alpha < 2$, $\kappa_1 = \sqrt{c + \frac{1-\alpha}{\alpha}\sigma_d^2}$ and $\kappa_d = \sqrt{c + \frac{1-\alpha}{\alpha}\sigma_1^2}$, where σ_1, σ_d are singular value of matrix A. Then the original differential inclusion becomes $0 \in \partial F(X(t)) + (B^\top B)\dot{X}(t)$. 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_μ^*.

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

$$\mathcal{E}_\mu(t) = t(F_\mu(X_\mu(t)) - F_\mu(x_\mu^*)) + \frac{\lambda}{2}\|B(X_\mu(t) - x_\mu^*)\|_2^2, \tag{23}$$

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

$$\dot{\mathcal{E}}_\mu(t) = (F_\mu(X_\mu(t)) - F_\mu(x_\mu^*)) + t\langle \nabla F_\mu(X_\mu(t)), \dot{X}_\mu(t)\rangle + \lambda\langle B^\top B(X_\mu(t) - x_\mu^*), \dot{X}_\mu(t)\rangle. \tag{24}$$

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

$$\dot{\mathcal{E}}_\mu(t) = -t\|B\dot{X}_\mu(t)\|_2^2 + (F_\mu(X_\mu(t)) - F_\mu(x_\mu^*)) - \lambda\langle (X_\mu(t) - x_\mu^*), \nabla F_\mu(X_\mu(t))\rangle \le 0, \tag{25}$$

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

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

$$\mathcal{E}(t) = t(F(X(t)) - F(x^*)) + \frac{\lambda}{2}\|B(X(t) - x^*)\|_2^2. \tag{26}$$

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

$$\mathcal{E}(0) = \frac{\lambda}{2}\|B(x_0 - x^*)\|_2^2 \le \frac{\lambda\kappa_1^2}{2}\|x_0 - x^*\|_2^2. \tag{27}$$

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

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

$$\frac{\lambda\kappa_d^2}{2}\|(X(t) - x^*)\|_2^2 \le \mathcal{E}(0). \tag{28}$$

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

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

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. $t \geq 0$,

$$t(F(X(t)) - F(x^*)) \leq \mathcal{E}(t) \leq \mathcal{E}(0) \leq \frac{\lambda\kappa_1^2}{2}\|x_0 - x^*\|_2^2 \tag{30}$$

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

$$F(X(t)) - F(x^*) \leq \mathcal{E}(t) \leq \mathcal{E}(0) \leq \frac{\kappa_1^2}{2t}\|x_0 - x^*\|_2^2 \tag{31}$$

By applying convexity of F_μ to (25), we have

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

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

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

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

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

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

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

□

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

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

$$\frac{1}{\rho}\partial g\left(z_{k+1}\right) - (\alpha A x_{k+1} + (1-\alpha)z_k - z_{k+1} + u_k) \ni 0, \tag{36b}$$

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

Adding up (36b) and (36c) 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:

$$u_{k+1} \in \frac{1}{\rho}\partial g(z_{k+1}). \tag{37}$$

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

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

$$0 \in \partial f(x_{k+1}) + A^\top \partial g(z_k) + \rho A^\top (Ax_{k+1} - z_k), \tag{38}$$

where again from (36c),

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

and hence

$$Ax_{k+1} - z_k = \frac{1}{\alpha}[(u_{k+1} - u_k) + (z_{k+1} - z_k)] \tag{39}$$

Plugging (39) into (38) gives

$$0 \in \partial f(x_{k+1}) + A^\top \partial g(z_k) + \rho A^\top \left(\frac{1}{\alpha}[(u_{k+1} - u_k) + (z_{k+1} - z_k)] \right). \tag{40}$$

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

$$0 \in \partial f(X(t)) + A^T \partial g(Z(t)) + \frac{1}{\alpha} A^T(\dot{Z}(t)). \tag{41}$$

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

$$Z(t) = AX(t), \qquad \dot{Z}(t) = A\dot{X}(t).$$

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

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

Thus we complete the proof.

□

C.2. Proof of Theorem 10

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

$$\mathcal{E}_\mu(t) = \alpha t(F_\mu(X_\mu(t)) - F_\mu(x^*_\mu)) + \frac{\lambda}{2}\|A(X_\mu(t) - x^*_\mu)\|_2^2, \tag{42}$$

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

$$\dot{\mathcal{E}}_\mu(t) = \alpha(F_\mu(X_\mu(t)) - F_\mu(x^*_\mu)) + \alpha t\langle \nabla F_\mu(X_\mu(t)), \dot{X}_\mu(t)\rangle + \lambda\langle A^\top A(X_\mu(t) - x^*_\mu), \dot{X}_\mu(t).\rangle \tag{43}$$

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

$$\dot{\mathcal{E}}_\mu(t) = -t\|A\dot{X}_\mu(t)\|_2^2 + \alpha(F_\mu(X_\mu(t)) - F_\mu(x^*_\mu)) - \lambda\alpha\langle (X_\mu(t) - x^*_\mu), \nabla F_\mu(X_\mu(t))\rangle \leq 0, \tag{44}$$

where we used the convexity of F_μ 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

$$\mathcal{E}(t) = \alpha t(F(X(t)) - F(x^*)) + \frac{\lambda}{2}\|A(X(t) - x^*)\|_2^2. \tag{45}$$

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

$$\mathcal{E}(0) = \frac{\lambda}{2}\|A(X(0) - x^*)\|_2^2 \leq \frac{\lambda\sigma_1^2}{2}\|x_0 - x^*\|_2^2. \tag{46}$$

By applying the approximation scheme ($\mathcal{AS}$) argument (details as in Appendix A) as $\mu \to 0$ to equation (44), we have for a.e. $t \geq 0$, $\dot{\mathcal{E}}(t) \leq 0$ 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

$$\frac{\lambda}{2}\|A(X(t) - x^*)\|_2^2 \leq \mathcal{E}(0). \tag{47}$$

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

$$\|X(t) - x^*\|_2 \leq \frac{\sigma_1}{\sigma_d}\|x_0 - x^*\|_2. \tag{48}$$

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

$$\alpha t(F(X(t)) - F(x^*)) \leq \mathcal{E}(t) \leq \mathcal{E}(0) \leq \frac{\lambda\sigma_1^2}{2}\|x_0 - x^*\|_2^2 \quad for \quad a.e. \quad t, \tag{49}$$

which is given by (46). Thus $(F(X(t)) - F(x^*)) \leq \frac{\sigma_1^2}{2\alpha t}\|x_0 - x^*\|_2^2$ by taking $\lambda = 1$.

From (44) and using the convexity of F_μ, we have

$$\dot{\mathcal{E}}_\mu(t) \leq \alpha(1-\lambda)(F_\mu(X_\mu(t)) - F_\mu(x_\mu^*)) - t\|A\dot{X}_\mu(t)\|_2^2. \tag{50}$$

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

$$F_\mu(X_\mu(t)) - F_\mu(x_\mu^*) \leq \frac{-\dot{\mathcal{E}}_\mu(t)}{\alpha(\lambda-1)} \qquad \text{and} \qquad t\|A\dot{X}_\mu(t)\|_2^2 \leq -\dot{\mathcal{E}}_\mu(t). \tag{51}$$

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

$$\int_0^T (F_\mu(X_\mu(t)) - F_\mu(x_\mu^*))\mathrm{d}t \leq \frac{\mathcal{E}_\mu(0)}{\alpha(\lambda-1)}, \quad \int_0^T t\|A\dot{X}_\mu(t)\|_2^2\mathrm{d}t \leq \mathcal{E}_\mu(0). \tag{52}$$

By applying approximation scheme ($\mathcal{AS}$) and plugging in (46), we have

$$\int_0^T (F_\mu(X_\mu(t)) - F_\mu(x_\mu^*))\mathrm{d}t \leq \frac{\lambda\sigma_1^2}{2\alpha(\lambda-1)}\|x_0 - x^*\|_2^2, \quad \int_0^T t\|A\dot{X}_\mu(t)\|_2^2\mathrm{d}t \leq \frac{\lambda\sigma_1^2}{2}\|x_0 - x^*\|_2^2. \tag{53}$$

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

$$\int_0^\infty (F(X(t)) - F(x^*))\mathrm{d}t \leq \frac{\sigma_1^2}{2\alpha}\|x_0 - x^*\|_2^2, \quad \int_0^\infty t\|\dot{X}(t)\|_2^2\mathrm{d}t \leq \frac{\sigma_1^2}{2\sigma_d^2}\|x_0 - x^*\|_2^2. \tag{54}$$

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

$$\partial f(x_{k+1}) + \rho A^T(Ax_{k+1} - \hat{z}_k + \hat{u}_k) \ni 0, \tag{55a}$$

$$\frac{1}{\rho}\partial g(z_{k+1}) - (\alpha A x_{k+1} + (1-\alpha)\hat{z}_k - z_{k+1} + \hat{u}_k) \ni 0, \tag{55b}$$

$$u_{k+1} - (\alpha A x_{k+1} + (1-\alpha)\hat{z}_k - z_{k+1} + \hat{u}_k) = 0. \tag{55c}$$

Adding up (55b) and (55c) eliminates the common term $-(\alpha Ax_{k+1}+(1-\alpha)\hat{z}_k-z_{k+1}+\hat{u}_k)$ and reduces to a simple u-update:

$$u_{k+1}\in\frac{1}{\rho}\partial g(z_{k+1}). \tag{56}$$

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 $\hat{u}$ into (55a) leads to:

$$\partial f(x_{k+1})+A^T\partial g(z_k)+\rho A^T(Ax_{k+1}-\hat{z}_k)+\rho\gamma_{k+1}A^\top(u_{k+1}-u_k)\ni 0, \tag{57}$$

where again from (55c),

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

In addition, from equation (12d) and equation (12e), we find that $u_{k+1}-\hat{u}_k=u_{k+1}-(1+\gamma_{k+1})u_k+\gamma_{k+1}u_{k-1}$ and $z_{k+1}-\hat{z}_k=z_{k+1}-(1+\gamma_{k+1})z_k+\gamma_{k+1}z_{k-1}$. For $u_{k+1}-\hat{u}_k$, we add the term $u_k-u_k+u_{k-1}-u_{k-1}$ to the right hand side, the resulting equation is a combination of the second order difference and first order difference of the sequence $\{u_k\}$:

$$u_{k+1}-\hat{u}_k=(u_{k+1}-2u_k+u_{k-1})+(1-\gamma_{k+1})(u_k-u_{k-1}). \tag{59}$$

Similarly, the equation holds that:

$$z_{k+1}-\hat{z}_k=(z_{k+1}-2z_k+z_{k-1})+(1-\gamma_{k+1})(z_k-z_{k-1}). \tag{60}$$

We note that $1-\gamma_k=1-\frac{k}{k+r}=\frac{r}{\rho^{1/2}t+r}$. Taking the limit $\rho\to\infty$, under infinitesimal step sizes, using relationships (58), (59), (60) and the fact that $\dot{U}(t)=0,\ddot{U}(t)=0$, equation (57) becomes:

$$\partial f(X(t))+A^T\partial g(Z(t))+\frac{1}{\alpha}A^T(\frac{r}{t}\dot{Z}(t)+\ddot{Z}(t))\ni 0. \tag{61}$$

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

$$Z(t)=AX(t),\qquad \dot{Z}(t)=A\dot{X}(t),\qquad \ddot{Z}(t)=A\ddot{X}(t).$$

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

$$0\in\partial F(X(t))+\left(\frac{1}{\alpha}A^\top A\right)(\ddot{X}(t)+\frac{r}{t}\dot{X}(t)),$$

□

C.4. Proof of Theorem 12

Proof of Theorem 12. Recall that x_μ^* is the minimizer of F_μ. For each $\mu>0$, consider the energy functional of Moreau-Yosida approximation defined as

$$\mathcal{E}_\mu(t)=t^2(F_\mu(X_\mu(t))-F_\mu(x_\mu^*))+\frac{1}{2\alpha}\left\|A\left(\lambda(X_\mu(t)-x_\mu^*)+t\dot{X}_\mu(t)\right)\right\|_2^2+\frac{\lambda(r-\lambda-1)}{2\alpha}\|A(X_\mu(t)-x_\mu^*)\|_2^2 \tag{62}$$

where λ is a constant chosen within $2\leq\lambda\leq r-1$. Because F_μ is a continuously differentiable function, we could write the time derivative of $\mathcal{E}_\mu(t)$ as

$$\begin{aligned}\dot{\mathcal{E}}_\mu&=2t(F_\mu(X_\mu(t))-F_\mu(x_\mu^*))+t^2\nabla F_\mu(X_\mu(t))^\top\dot{X}_\mu+\left(\lambda(X_\mu-x_\mu^*)+t\dot{X}_\mu\right)^\top(\frac{1}{\alpha}A^\top A)\left((\lambda+1)\dot{X}_\mu+t\ddot{X}_\mu\right)\\&\quad+\lambda(r-\lambda-1)(X_\mu-x_\mu^*)^\top(\frac{1}{\alpha}A^\top A)\dot{X}_\mu\end{aligned}$$

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

$$\dot{\mathcal{E}}_\mu = -\lambda t\left(F_\mu(x^*_\mu) - F_\mu(X_\mu(t)) - (x^*_\mu - X_\mu)^\top \nabla F_\mu(X_\mu(t))\right) - (\lambda-2)t(F_\mu(X_\mu(t)) - F_\mu(x^*_\mu)) - \frac{(r-1-\lambda)t}{\alpha}\|A\dot{X}_\mu\|_2^2 \le 0 \tag{63}$$

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

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

$$\mathcal{E}(t) = t^2(F(X(t)) - F(x^*)) + \frac{1}{2\alpha}\left\|A\left(\lambda(X(t)-x^*) + t\dot{X}(t)\right)\right\|_2^2 + \frac{\lambda(r-\lambda-1)}{2\alpha}\|A(X(t)-x^*)\|_2^2$$

At time t_0, there is an upper bound on $\mathcal{E}(t_0)$ as

$$\mathcal{E}(t_0) = t_0^2(F(x_0) - F(x^*)) + \frac{\lambda(r-1)}{2\alpha}\|A(x_0-x^*)\|_2^2 \le \frac{2\alpha + \lambda(r-1)\sigma_1^2}{2\alpha}\Delta_0^2 \tag{64}$$

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 $r \ge 3$ and $t \ge t_0$ that

$$\frac{\mathrm{d}}{\mathrm{d}t}(t\|X_\mu - x^*_\mu\|_2^2) = \|X_\mu - x^*_\mu\|_2^2 + 2t(X_\mu - x^*_\mu)^\top\dot{X}_\mu \le \frac{1}{2}\|2(X_\mu - x^*_\mu) + t\dot{X}_\mu\|_2^2 \le \frac{\alpha\mathcal{E}_\mu}{\sigma_d^2} \le \frac{\alpha\mathcal{E}_\mu(t_0)}{\sigma_d^2}$$

By integrating over (t_0, t), this gives us

$$t\|X_\mu - x^*_\mu\|_2^2 - t_0\|x_0 - x^*_\mu\|_2^2 \le \frac{\alpha(t-t_0)}{\sigma_d^2}\mathcal{E}_\mu(t_0)$$

By applying the approximation scheme ($\mathcal{AS}$) argument (details as in Appendix A) as $\mu \to 0$, we have for a.e. $t \ge t_0$ that

$$\|X - x^*\|_2^2 \le \frac{\alpha\mathcal{E}(t_0)}{\sigma_d^2} + \|x_0 - x^*\|_2^2$$

Combining with the upper bound of $\mathcal{E}(t_0)$ in (64), we derive for a.e. $t \ge t_0$ that

$$\|X(t) - x^*\|_2 \le C_1\Delta_0 \tag{65}$$

with factor $C_1 = \sqrt{\frac{\alpha + (r-1)\sigma_1^2 + \sigma_d^2}{\sigma_d^2}}$. Here we choose $\lambda = 2$ to minimize C_1.

From (63), we know that $\mathcal{E}_\mu(t)$ is nonincreasing for $t \ge t_0$, for all $\mu > 0$. By applying ($\mathcal{AS}$) we find that $\mathcal{E}(t)$ is nonincreasing for a.e. $t \ge t_0$. Using the nonnegativity of all three terms in (62) and monotonicity of $\mathcal{E}(t)$ on a.e. $t \ge t_0$, we have for a.e. $t \ge t_0$ that

$$F(X(t)) - F(x^*) \le \frac{1}{t^2}\mathcal{E}(t) \le \frac{1}{t^2}\mathcal{E}(t_0) \le \frac{C_2}{t^2}\Delta_0^2$$

where factor $C_2 = 1 + (r-1)\sigma_1^2/\alpha$ is given by (64) with $\lambda = 2$, and

$$\|\lambda(X(t) - x^*) + t\dot{X}\|_2^2 \le \frac{2\alpha}{\sigma_d^2}\mathcal{E}(t) \le \frac{2\alpha}{\sigma_d^2}\mathcal{E}(t_0) \le \frac{2\alpha + \lambda(r-1)\sigma_1^2}{\sigma_d^2}\Delta_0^2$$

Therefore, by triangle inequality and (65),

$$\|\dot{X}(t)\|_2 \le \frac{1}{t}\|\lambda(X(t)-x^*) + t\dot{X}(t)\|_2 + \frac{1}{t}\lambda\|X(t) - x^*\|_2 \le \frac{C_3}{t}\Delta_0$$

with factor $C_3 = \sqrt{\frac{2\alpha + 2(r-1)\sigma_1^2}{\sigma_d^2}} + 2C_1$. Here we choose $\lambda = 2$ to minimize C_3.

From (63), we have

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

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

$$t(F_\mu(X_\mu(t)) - F_\mu(x_\mu^*)) \le -\frac{1}{r-3}\dot{\mathcal{E}}_\mu \quad \text{and} \quad t\|\dot{X}_\mu\|_2^2 \le -\frac{\alpha}{(r-3)\sigma_d^2}\dot{\mathcal{E}}_\mu$$

By integrating over (t_0, ∞), the inequalities above give

$$\int_{t_0}^{\infty} t(F_\mu(X_\mu(t)) - F_\mu(x_\mu^*))\mathrm{d}t \le \frac{\mathcal{E}_\mu(t_0)}{r-3} \quad \text{and} \quad \int_{t_0}^{\infty} t\|\dot{X}_\mu(t)\|_2^2\mathrm{d}t \le \frac{\alpha\mathcal{E}_\mu(t_0)}{(r-3)\sigma_d^2}$$

By applying ($\mathcal{AS}$) and plugging in (64), we have

$$\int_{t_0}^{\infty} t(F(X(t)) - F(x^*))\mathrm{d}t \le C_4\Delta_0^2 \quad \text{and} \quad \int_{t_0}^{\infty} t\|\dot{X}(t)\|_2^2\mathrm{d}t \le C_5\Delta_0^2$$

with factors $C_4 = \frac{2\alpha+(r-1)^2\sigma_1^2}{2(r-3)\alpha}$ and $C_5 = \frac{\alpha+(r-1)\sigma_1^2}{(r-3)\sigma_d^2}$. □

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 λ satisfying $\lambda - 2 \ge 0$ and $r - 1 - \lambda \ge 0$ simultaneously when $0 < r < 3$. Here we consider a new energy functional for the Moreau-Yosida approximation

$$\mathcal{E}_\mu(t) = t^2(F_\mu(X_\mu(t)) - F_\mu(x_\mu^*)) + \frac{1}{2\alpha}\left\|\frac{2r}{3}A(X_\mu(t) - x_\mu^*) + tA\dot{X}_\mu(t)\right\|_2^2 + \frac{r(3-r)}{9\alpha}\|A(X_\mu(t) - x_\mu^*)\|_2^2 \tag{66}$$

By taking its time derivative, we have

$$\begin{aligned}\dot{\mathcal{E}}_\mu =& 2t(F_\mu(X_\mu(t)) - F_\mu(x_\mu^*)) + t^2\nabla F_\mu(X_\mu(t))^\top\dot{X}_\mu + \left(\frac{2r}{3}(X_\mu - x_\mu^*) + t\dot{X}_\mu\right)^\top(\frac{1}{\alpha}A^\top A)\left((\frac{2r}{3}+1)\dot{X}_\mu + t\ddot{X}_\mu\right)\\ &+ \frac{2r(3-r)}{9}(X_\mu - x_\mu^*)^\top(\frac{1}{\alpha}A^\top A)\dot{X}_\mu\end{aligned}$$

By using the equality of $tA^\top A\ddot{X}_\mu$ and $-rA^\top A\dot{X}_\mu - \alpha t\nabla F_\mu(X_\mu)$ and applying the convexity of F_μ, we have

$$\dot{\mathcal{E}}_\mu \le \frac{2(3-r)}{3}t(F_\mu(X_\mu(t)) - F_\mu(x_\mu^*)) + \frac{4r(3-r)}{9\alpha}(X_\mu - x_\mu^*)^\top A^\top A\dot{X}_\mu + \frac{3-r}{3\alpha}t\|A\dot{X}_\mu\|_2^2$$

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

$$\dot{\mathcal{E}}_\mu - \frac{2(3-r)}{3t}\mathcal{E}_\mu \le -\frac{2r(3-r)(3+r)}{27\alpha t}\|A(X_\mu - x_\mu^*)\|_2^2 \le 0$$

This implies that, for $\mathcal{H}_\mu(t) := t^{-\frac{2(3-r)}{3}}\mathcal{E}_\mu(t)$, for all $t \ge t_0$,

$$\dot{\mathcal{H}}_\mu = t^{-\frac{2(3-r)}{3}} \cdot (\dot{\mathcal{E}}_\mu - \frac{2(3-r)}{3t}\mathcal{E}_\mu) \le 0$$

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

$$\mathcal{H}(t_0) \le t_0^{-\frac{2(3-r)}{3}} \cdot \left(1 + \frac{r(3+r)}{9\alpha}\sigma_1^2\right)\Delta_0^2$$

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

$$F(X(t)) - F(x^*) \le \frac{1}{t^{\frac{2r}{3}}}\mathcal{H}(t) \le \frac{1}{t^{\frac{2r}{3}}}\mathcal{H}(t_0) \le \frac{C_6 t_0^{-\frac{2(3-r)}{3}}\Delta_0^2}{t^{\frac{2r}{3}}}$$

with factor $C_6 = 1 + \frac{r(3+r)\sigma_1^2}{9\alpha}$.

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

$$\left\| \frac{2r}{3}(X(t) - x^*) + t\dot{X} \right\|_2^2 \leq \frac{2\alpha}{\sigma_d^2} t^{\frac{2(3-r)}{3}} \mathcal{H}(t) \leq \frac{2\alpha}{\sigma_d^2} t^{\frac{2(3-r)}{3}} \mathcal{H}(t_0) \leq \frac{2\alpha C_6 t_0^{-\frac{2(3-r)}{3}}}{\sigma_d^2} t^{\frac{2(3-r)}{3}} \Delta_0^2$$

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. $t \geq t_0$, $\|X(t) - x^*\|_2 \leq C_0 \Delta_0$. Then triangle inequality gives us, for a.e. $t \geq t_0$, that

$$\|\dot{X}\|_2 \leq \frac{1}{t} \left\| \frac{2r}{3}(X(t) - x^*) + t\dot{X} \right\|_2 + \frac{2r}{3t} \|X(t) - x^*\|_2 \leq \frac{C_7 t_0^{-\frac{3-r}{3}} \Delta_0}{t^{\frac{r}{3}}}$$

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