Electronic Companion

EC.1. Preliminaries of Difference-of-Moreau-Envelopes Smoothing

EC.1.1. Moreau Envelope

We summarize some known properties of Moreau envelope in the next proposition.

PROPOSITION EC.1. *Suppose Assumption 1 holds and* $0 < \mu < 1/m_\phi$ *in* (10). *Then the following claims hold.*

1. $x_{\mu\phi}$ *is Lipschitz continuous with modulus* $\frac{1}{1-\mu m_\phi}$.
2. $M_{\mu\phi}$ *is differentiable with gradient* $\nabla M_{\mu\phi}(z) = \mu^{-1}(z - x_{\mu\phi}(z))$.
3. $\nabla M_{\mu\phi}$ *is Lipschitz continuous with modulus* $\frac{2-\mu m_\phi}{\mu-\mu^2 m_\phi}$.

Proof. The first two claims are well-known, see, e.g., (Zhang and Luo 2020b, Lemma 3.5) and (Rockafellar and Wets 2009, Proposition 13.37), combining which proves the last one. □

Proposition EC.1 suggests that $M_{\mu\phi}$ forms a smooth approximation of the possibly nonconvex nonsmooth function ϕ. Similarly, the Moreau envelope and proximal mapping of g are given by $M_{\mu g}$ and $x_{\mu g}$, respectively. Since g is convex, it is known that $x_{\mu g}$ is 1-Lipschitz and $M_{\mu g}$ is differentiable, whose gradient $\nabla M_{\mu g}(z) = (z - x_{\mu g}(z))/\mu$ is $1/\mu$-Lipschitz (Beck 2017, Theorem 6.60).

EC.1.2. Lipschitz Differentiability of F_μ

PROPOSITION EC.2. *Suppose Assumption 1 holds and* $0 < \mu < 1/m_\phi$ *in* (10). F_μ *is differentiable, and* $\nabla F_\mu(z) = \mu^{-1}(x_{\mu g}(z) - x_{\mu\phi}(z))$ *is Lipschitz continuous with modulus* $L_{F_\mu} = \frac{2-\mu m_\phi}{\mu-\mu^2 m_\phi}$.

Proof. By the Lipschitz differentiability of the Moreau envelope shown in Proposition EC.1 and the definition of F_μ, we know that F_μ is differentiable, and $\nabla F_\mu(z) = \nabla M_{\mu\phi}(z) - \nabla M_{\mu g}(z) = \mu^{-1}(x_{\mu g}(z) - x_{\mu\phi}(z))$. Since $x_{\mu g}$ and $x_{\mu\phi}$ are Lipschitz with modulus 1 and $\frac{1}{1-\mu m_\phi}$, respectively, we obtain the claimed $L_{F_\mu} = \frac{1}{\mu}(1 + \frac{1}{1-\mu m_\phi})$. □

If ϕ is convex, then the Lipschitz constant of ∇F_μ can be improved to $2/\mu$ (Hiriart-Urruty 1991).

EC.1.3. Correspondence of Stationary Points and Global Minima of F and F_μ

In addition to being smooth, the approximation F_μ captures both the local and global structure of the original function F. In particular, some properties of F_μ established in Hiriart-Urruty (1991) are summarized in the next proposition.

PROPOSITION EC.3 (**Hiriart-Urruty (1991)**). *Suppose Assumptions 1 and 2 hold and* $0 < \mu < 1/m_\phi$ *in* (10). *Then the following claims hold.*

1. *The set of global minimizers of* F_μ, $\arg\min F_\mu$, *is nonempty, and* $F^* = \min_{z\in\mathbb{R}^n} F_\mu(z)$.

2. *(Correspondence of Stationary Point) If z is a stationary point of F_μ, i.e., $\nabla F_\mu(z)=0$, then $x_{\mu\phi}(z)=x_{\mu g}(z)$, and $x_{\mu\phi}(z)$ is a stationary point of F in the sense of (8) with $F(x_{\mu\phi}(z))=F_\mu(z)$; conversely, if $x\in\mathbb{R}^n$ is a stationary point of F, then there exists $z\in\mathbb{R}^n$ such that $\nabla F_\mu(z)=0$, $z=x_\phi(x)=x_g(x)$, and $F_\mu(z)=F(x)$.*

3. *(Correspondence of Global Minima) If $z\in\arg\min F_\mu$, then $x_{\mu\phi}(z)\in\arg\min F$; conversely, if $x\in\arg\min F$, then there exists $z\in\arg\min F_\mu$ such that $x=x_{\mu\phi}(z)=x_{\mu g}(z)$.*

EC.2. Proofs in Section 3

EC.2.1. Proof of Proposition 1

Proof. By Proposition EC.3 and Lemma 1, we know $x_{\mu\phi}(\bar z)=x_{\mu g}(\bar z)$, and for all z such that $\|z-\bar z\|\le r$, we have $F(x_{\mu g}(\bar z))=F(x_{\mu\phi}(\bar z))=F_\mu(\bar z)\le F_\mu(z)\le F(x_{\mu g}(z))$. It suffices to show that, for all x sufficiently close to $x_{\mu g}(\bar z)$, there exists some $z\in\mathbb{R}^n$ such that $\|z-\bar z\|\le r$ and $x=x_{\mu g}(z)$. In particular, take $\tilde\nabla g(x)\in\partial g(x)$ and let $z=x+\mu\tilde\nabla g(x)$. By construction, we have $0=\tilde\nabla g(x)+\mu^{-1}(x-z)\in\partial g(x)+\mu^{-1}(x-z)$, and therefore $x=x_{\mu g}(z)$. It follows that

$$\begin{aligned}\|z-\bar z\| &= \|x+\mu\tilde\nabla g(x)-x_{\mu g}(\bar z)-\mu\tilde\nabla g(x_{\mu g}(\bar z))\|\\ &\le \|x-x_{\mu g}(\bar z)\|+\mu\|\tilde\nabla g(x)-\tilde\nabla g(x_{\mu g}(\bar z))\|\le\begin{cases}(1+\mu L_g)\|x-x_{\mu g}(\bar z)\|\le r\\ r-2\mu M_{\partial g}+2\mu M_{\partial g}=r,\end{cases}\end{aligned}$$

where the two cases above correspond to the two claims respectively. $\square$

EC.2.2. Proof of Proposition 2

Proof. 1. We prove that $\mathrm{lev}_\alpha F_\mu$ is bounded. Without loss of generality, we consider $\alpha\ge F^*$, otherwise $\mathrm{lev}_\alpha F_\mu=\emptyset$ by Proposition EC.3. Firstly notice that

$$\mathrm{lev}_\alpha F_\mu\subseteq\{z: M_{\mu\phi}(z)-g(z)\le\alpha\}\subseteq\left\{z:\exists x \text{ s.t. } F(x)+\frac{1}{2\mu}\|x-z\|^2-L\|x-z\|\le\alpha+M\right\},\quad\text{(EC.1)}$$

where the first inclusion is due to $M_{\mu g}\le g$, and the second inclusion is due to our assumption on g. Using the fact that $L\|x-z\|\le\frac{tL^2}{2}+\frac{\|x-z\|^2}{2t}$ for any $t>0$ and taking $t=2\mu$, we have

$$F(x)+\frac{1}{2\mu}\|x-z\|^2-L\|x-z\|\ge F(x)+\frac{1}{4\mu}\|x-z\|^2-\mu L^2.\quad\text{(EC.2)}$$

Now (EC.1), (EC.2), and the fact that $F(x)+\frac{1}{4\mu}\|x-z\|^2\ge\max\{F(x),F^*+\frac{1}{4\mu}\|x-z\|^2\}$ imply that

$$\mathrm{lev}_\alpha F_\mu\subseteq\left\{z:\|z\|\le\sqrt{(4\mu)(\alpha+M+\mu L^2-F^*)}+\max_{x:F(x)\le\alpha+M+\mu L^2}\|x\|\right\}.\quad\text{(EC.3)}$$

Since $\mathrm{lev}_{\alpha+M+\mu L^2}F$ is compact, $\max_{x:F(x)\le\alpha+M+\mu L^2}\|x\|$ is finite, and hence $\mathrm{lev}_\alpha F_\mu$ is bounded.

2. Suppose $F = \phi - g \geq \alpha\|\cdot\| + r$ for some $a \in (0, +\infty)$ and $r \in R$. For a convex function f, we use f^* to denote its convex conjugate, i.e., $f^*(z) = \sup_x \langle z, x\rangle - f(x)$. By definition, we have

$$\mu F_\mu = \mu(M_{\mu\phi} - M_{\mu g}) = \left(\mu g + \frac{1}{2}\|\cdot\|^2\right)^* - \left(\mu\phi + \frac{1}{2}\|\cdot\|^2\right)^*$$
$$\geq \left(\mu g + \frac{1}{2}\|\cdot\|^2\right)^* - \left(\mu g + \mu\alpha\|\cdot\| + \mu r + \frac{1}{2}\|\cdot\|^2\right)^*, \tag{EC.4}$$

where the inequality uses the fact that $f_1 \geq f_2$ implies $f_1^* \leq f_2^*$ for any functions f_1 and f_2.

We first consider some properties of the first term in (EC.4). For simplicity, denote $p = \left(\mu g + \frac{1}{2}\|\cdot\|^2\right)^*$. Since $\mu g + \frac{1}{2}\|\cdot\|^2$ is strongly convex with modulus 1, its conjugate p is convex and has Lipschitz gradient with modulus 1 (Beck 2017, Theorem 5.26). Moreover, we claim that p is coercive: notice that for any $\bar{\alpha} \in (0, +\infty)$, we have

$$p(z) = \max_x \left\{\langle z, x\rangle - \mu g(x) - \frac{1}{2}\|x\|^2\right\} \geq \max_{x:\|x\|\leq\bar{\alpha}} \langle z, x\rangle - \max_{x:\|x\|\leq\alpha} \left\{\mu g(x) + \frac{1}{2}\|x\|^2\right\} = \bar{\alpha}\|z\| + \bar{r},$$

where $\bar{r} = -\max_{x:\|x\|\leq\alpha} \left\{\mu g(x) + \frac{1}{2}\|x\|^2\right\}$ is finite, since $\mu g + \frac{1}{2}\|\cdot\|^2$ achieves a finite maximum over the compact set $\{x : \|x\| \leq \alpha\}$.

Next we rewrite the second term in (EC.4): for any $z \in \mathbb{R}^n$,

$$\left(\mu g + \mu\alpha\|\cdot\| + \mu r + \frac{1}{2}\|\cdot\|^2\right)^*(z) = \min_x \left\{p(x) + (\alpha\mu\|\cdot\| + \mu r)^*(z - x)\right\}$$
$$= \min_{w:\|w\|\leq\alpha\mu} p(z - w) - \mu r, \tag{EC.5}$$

where the first equality is due to (Beck 2017, Theorem 4.17), and the second equality uses the following facts: $\|\cdot\|^* = \delta_{\{x:\|x\|\leq 1\}}$ (Beck 2017, Section 4.4.2), and $(\alpha\mu\|\cdot\| + \mu r)^*(w) = (\alpha\mu)\|\cdot\|^*(\frac{w}{\alpha\mu}) - \mu r$ (Beck 2017, Theorem 4.13, 4.14). Combining (EC.4) and (EC.5), we have

$$\mu F_\mu(z) \geq p(z) - \min_{w:\|w\|\leq\alpha\mu} p(z - w) + \mu r = \max_{w:\|w\|\leq\alpha\mu} p(z) - p(z - w) + \mu r$$
$$\geq \max_{w:\|w\|\leq\alpha\mu} \langle\nabla p(z), w\rangle - \frac{1}{2}\|w\|^2 + \mu r \geq \alpha\mu\|\nabla p(z)\| - \frac{1}{2}\alpha^2\mu^2 + \mu r, \tag{EC.6}$$

where the second inequality is due to the Lipschitz differentiability of p, and the last inequality holds with $w = \alpha\mu\frac{\nabla p(z)}{\|\nabla p(z)\|}$ when $\|\nabla p(z)\| > 0$, or any w with $\|w\| = \alpha\mu$ when $\|\nabla p(z)\| = 0$. Notice that (EC.6) further suggests that

$$\liminf_{\|z\|\to\infty} F_\mu(z) \geq \liminf_{\|z\|\to\infty} \alpha\|\nabla p(z)\| - \frac{1}{2}\alpha^2\mu + r \geq \alpha \liminf_{\|z\|\to\infty} \frac{p(z) - p(0)}{\|z\|} - \frac{1}{2}\alpha^2\mu + r = +\infty, \tag{EC.7}$$

where the second inequality is due to the convexity of p: $\|\nabla p(z)\|\|z\| \geq \nabla p(z)^\top z \geq p(z) - p(0)$, and the last equality is due to p being coercive (see an equivalent characterization in (Rockafellar and Wets 2009, Definition 3.25)). Therefore, (EC.7) implies that F_μ is level-bounded.

3. Since dom ϕ is compact, there exists $R>0$ such that dom $\phi \subseteq \{x : \|x\| \le R\}$. Notice that

$$
\begin{aligned}
F_\mu(z) \ge & \min_{x:\|x\|\le R}\left\{\phi(x)+\frac{1}{2\mu}\|x\|^2-\frac{1}{\mu}\langle x,z\rangle\right\}+\max_x\left\{\frac{1}{\mu}\langle x,z\rangle-g(x)-\frac{1}{2\mu}\|x\|^2\right\} \\
\ge & \hat{\phi}^*-\frac{R}{\mu}\|z\|+\max_x\left\{\frac{1}{\mu}\langle x,z\rangle-g(x)-\frac{1}{2\mu}\|x\|^2\right\},
\end{aligned}
$$

where in the last inequality, $\hat{\phi}^*=\min_x\{\phi(x)+\frac{1}{2\mu}\|x\|^2\}$ is well-defined by the strong convexity of $\phi+\frac{1}{2\mu}\|\cdot\|^2$. Pick any $\alpha\in(0,+\infty)$. By a similar argument used in the previous part, we have

$$
\max_x\left\{\frac{1}{\mu}\langle x,z\rangle-g(x)-\frac{1}{2\mu}\|x\|^2\right\}\ge\left(\frac{R}{\mu}+\alpha\right)\|z\|-\max_{x:\|x\|\le R+\mu\alpha}\left\{g(x)+\frac{1}{2\mu}\|x\|^2\right\}.
$$

Combining the above two inequalities, we have $F_\mu(z)\ge\alpha\|z\|+r$, where $r=\hat{\phi}^*-\max_{x:\|x\|\le R+\mu\alpha}\left\{g(x)+\frac{1}{2\mu}\|x\|^2\right\}$ is finite. Therefore, we conclude that F_μ is coercive, and hence also level-bounded.

4. We show that $\mathrm{lev}_\alpha F_\mu$ is bounded. Let $z\in\mathrm{lev}_\alpha F_\mu$. By Lemma 1 and the assumption that F is level-bounded, we know that $x_{\mu\phi}(z)\in\mathrm{lev}_\alpha F$ and hence is bounded. The definition of $x_{\mu\phi}$ gives $z\in\partial(\mu\phi+\frac{1}{2}\|\cdot\|^2)(x_{\mu\phi}(z))$. Since $x_{\mu\phi}(z)$ is bounded, and $\mu\phi+\frac{1}{2}\|\cdot\|^2$ is a (strongly) convex function whose domain is $\mathbb{R}^n$, we conclude that z is bounded (Rockafellar 1970, Theorem 24.7). □

EC.3. Proofs in Section 4

EC.3.1. Proof of Theorem 1

Proof. Notice that since ∇F_μ is L_{F_μ}-Lipschitz and $\alpha\le 1/L_{F_\mu}$, we have

$$
F_\mu(z^k)-F_\mu(z^{k+1})\ge\left(\frac{1}{\alpha}-\frac{L_{F_\mu}}{2}\right)\|z^{k+1}-z^k\|^2\ge\frac{1}{2\alpha}\|z^{k+1}-z^k\|^2=\frac{\alpha}{2\mu^2}\|x_{\mu g}(z^k)-x_{\mu\phi}(z^k)\|^2.
$$

Summing the above inequality over $k=0,\cdots,K-1$ for some positive integer $K-1$, we have

$$
\sum_{k=0}^{K-1}\|x_{\mu g}(z^k)-x_{\mu\phi}(z^k)\|^2\le\frac{2\mu^2}{\alpha}(F_\mu(z^0)-F_\mu(z^K))\le\frac{2\mu^2}{\alpha}(F_\mu(z^0)-F^*). \tag{EC.8}
$$

Let $\bar{k}=\arg\min_{k=0,\cdots,K-1}\|x_{\mu g}(z^k)-x_{\mu\phi}(z^k)\|^2$, then from (EC.8) it holds

$$
\|x_{\mu g}(z^{\bar{k}})-x_{\mu\phi}(z^{\bar{k}})\|\le\left(\frac{2\mu^2(F_\mu(z^0)-F^*)}{\alpha K}\right)^{1/2}. \tag{EC.9}
$$

For any $k\in\mathbb{Z}_+$, due to the optimality of the proximal mapping $x_{\mu\phi}(z^k)$ and $x_{\mu g}(z^k)$, we have

$$
\xi^k=\mu^{-1}(z^k-x_{\mu\phi}(z^k))-\mu^{-1}(z^k-x_{\mu g}(z^k))\in\partial\phi(x_{\mu\phi}(z^k))-\partial g(x_{\mu g}(z^k)). \tag{EC.10}
$$

In view of (EC.9), we have (12) proved due to the claimed upper bound K in (13). Since F_μ is level-bounded and $\{F_\mu(z^k)\}_{k\in\mathbb{N}}$ is monotonically non-increasing, we know the sequence $\{z^k\}_{k\in\mathbb{N}}$ is bounded and therefore has at least one limit point z^*. Let $\{z^{k_j}\}_{j\in\mathbb{N}}$ denote the subsequence convergent to z^*. Since $x_{\mu\phi}$ and $x_{\mu g}$ are continuous, (EC.8) implies $x_{\mu\phi}(z^*) = x_{\mu g}(z^*)$. Since g is continuous, we have $\lim_{j\to\infty} g(x_{\mu g}(z^{k_j})) = g(x_{\mu g}(z^*))$; in addition,

$$\begin{aligned}\phi(x_{\mu\phi}(z^*)) &\le \liminf_{j\to\infty} \phi(x_{\mu\phi}(z^{k_j})) \le \limsup_{j\to\infty} \phi(x_{\mu\phi}(z^{k_j})) \\ &\le \lim_{j\to\infty} \left[\phi(x_{\mu\phi}(z^*)) + \frac{1}{2\mu}\|x_{\mu\phi}(z^*) - z^{k_j}\|^2 - \frac{1}{2\mu}\|x_{\mu\phi}(z^{k_j}) - z^{k_j}\|^2\right] = \phi(x_{\mu\phi}(z^*)),\end{aligned}$$

where the first inequality is due to the lower-semicontinuity of ϕ and the last inequality is due to the optimality of $x_{\mu\phi}(z^{k_j})$ in each Moreau envelope evaluation, and therefore we also have $\lim_{j\to\infty} \phi(x_{\mu\phi}(z^{k_j})) = \phi(x_{\mu\phi}(z^*))$. Taking limit on (EC.10) along the subsequence gives (8). □

EC.3.2. Proof of Lemma 2

Proof We first show that the sequence is bounded from below:

$$\mathcal{F}(x^k, z^k) \ge \min_x f(x) + h(x) + \frac{1}{2\mu}\|x - z^k\|^2 - M_{\mu g}(z^k) = F_\mu(z^k) \ge F^*,$$

where the last inequality is due to Proposition EC.2. Next we show the descent in x:

$$\begin{aligned}\mathcal{F}(x^{k+1}, z^k) \le& f(x^k) + \langle \nabla f(x^k), x^{k+1} - x^k\rangle + \frac{L_f}{2}\|x^{k+1} - x^k\|^2 + h(x^{k+1}) + \frac{1}{2\mu}\|x^{k+1} - z^k\|^2 - M_{\mu g}(z^k) \\ \le& f(x^k) + h(x^k) + \frac{1}{2\mu}\|x^k - z^k\|^2 - M_{\mu g}(z^k) + \left(\frac{L_f}{2} - \frac{1}{2\mu}\right)\|x^{k+1} - x^k\|^2 \\ =& \mathcal{F}(x^k, z^k) - \left(\frac{\mu^{-1} - L_f}{2}\right)\|x^{k+1} - x^k\|^2, \end{aligned} \tag{EC.11}$$

where the first inequality is due to the Lipschitz differentiability of f and the second inequality is due to x^{k+1} being the minimizer of some μ^{-1}-strongly convex function. The descent with respect to z is given as:

$$\begin{aligned}&\mathcal{F}(x^{k+1}, z^k) - \mathcal{F}(x^{k+1}, z^{k+1}) \\ =& \frac{1}{\mu}\left(\frac{1}{\beta} - \frac{1}{2}\right)\|z^{k+1} - z^k\|^2 + M_{\mu g}(z^{k+1}) - M_{\mu g}(z^k) - \left\langle \frac{1}{\mu}(z^k - x_{\mu g}(z^k)), z^{k+1} - z^k \right\rangle \\ \ge& \frac{1}{\mu}\left(\frac{1}{\beta} - \frac{1}{2}\right)\|z^{k+1} - z^k\|^2, \end{aligned} \tag{EC.12}$$

where we replace x^{k+1} by $x_{\mu g}(z^k) + \frac{1}{\beta}(z^{k+1} - z^k)$ to get the equality, and the inequality is due to $M_{\mu g}$ being convex and $\nabla M_{\mu g}(z^k) = \mu^{-1}(z^k - x_{\mu g}(z^k))$. Combining (EC.11) and (EC.12) gives (15).
□

EC.3.3. Proof of Theorem 2

Proof. By Lemma 2, we have

$$\mathcal{F}(x^k, z^k) - \mathcal{F}(x^{k+1}, z^{k+1}) \ge \min\{c_1, c_2\}\left(\|x^{k+1} - x^k\|^2 + (\|z^{k+1} - z^k\|^2\right),$$

summing which from $k=0$ to some positive integer $K-1$ gives

$$\sum_{k=0}^{K-1}\left(\|x^{k+1}-x^k\|^2+(\|z^{k+1}-z^k\|^2\right)\le\frac{\mathcal{F}(x^0,z^0)-\mathcal{F}(x^K,z^K)}{\min\{c_1,c_2\}}\le\frac{\mathcal{F}(x^0,z^0)-F^*}{\min\{c_1,c_2\}}. \tag{EC.13}$$

Let $\bar{k}=\underset{k=0,\cdots,K-1}{\arg\min}\ \|x^{k+1}-x^k\|^2+\|z^{k+1}-z^k\|^2$, then (EC.13) implies

$$\max\{\|x^{\bar{k}+1}-x^{\bar{k}}\|,\|z^{\bar{k}+1}-z^{\bar{k}}\|\}\le\left(\frac{\mathcal{F}(x^0,z^0)-F^*}{\min\{c_1,c_2\}K}\right)^{1/2}. \tag{EC.14}$$

Due to the optimality condition of x^{k+1} and $x_{\mu g}(z^k)$, we have

$$\xi^{k+1}\in\nabla f(x^{k+1})+\partial h(x^{k+1})-\partial g(x_{\mu g}(z^k))=\partial\phi(x^{k+1})-\partial g(x_{\mu g}(z^k)),$$

which proves (17a). Now in view of (EC.14), we have

$$\max\{\|\xi^{\bar{k}+1}\|,\|x_{\mu g}(z^{\bar{k}})-x^{\bar{k}+1}\|\}\le\left(L_f+\frac{\mu^{-1}+1}{\beta}\right)\left(\frac{\mathcal{F}(x^0,z^0)-F^*}{\min\{c_1,c_2\}K}\right)^{1/2},$$

which proves (17b) and (18).

Next we show that if F_μ is level bounded, then $\{(x^k,z^k)\}_{k\in\mathbb{N}}$ stays bounded. Since F_μ is continuous and level-bounded, and $\mathcal{F}(x^0,z^0)\ge\mathcal{F}(x^k,z^k)\ge F_\mu(z^k)$, we know that z^k stays in some compact level set of F_μ. Since the mapping $x_{\mu g}$ is continuous, $x_{\mu g}(z^k)$ is also bounded. Consequently, $x^{k+1}=x_{\mu g}(z^k)+\frac{1}{\beta}(z^{k+1}-z^k)$ stays bounded for all $k\in\mathbb{N}$. Therefore, the sequence $\{(x^k,z^k)\}_{k\in\mathbb{N}}$ has a limit point, denoted as (x^*,z^*). Let $\{(x^{k_j},z^{k_j})\}_{j\in\mathbb{N}}$ be a subsequence converging to (x^*,z^*). Since $\|x^{k_j}-x^{k_j-1}\|\to 0$ and $\|z^{k_j}-z^{k_j-1}\|\to 0$, taking limit on $x^{k+1}=x_{\mu g}(z^k)+\frac{1}{\beta}(z^{k+1}-z^k)$ along the subsequence gives $x^*=x_{\mu g}(z^*)$. Finally the asymptotical convergence follows a similar argument as in the proof of Theorem 1. □

EC.4. Proofs in Section 5

EC.4.1. Proof of Lemma 3

Proof. We first verify condition (20) under the first two conditions. Since g is Lipschitz, we have $-g(y)\ge -g(x)-L_g\|x-y\|\ge -g(x)-\frac{1}{2}\|x-y\|^2-\frac{L_g^2}{2}$. For $0<\mu\le 1$, it follows that

$$v(\mu,\rho)\ge\inf_{x\in\mathbb{R}^n}\left\{f(x)-g(x)+\frac{\rho}{2}\|Ax-b\|^2\right\}-\frac{L_g^2}{2}. \tag{EC.15}$$

The first case implies that (EC.15) is finite for any $\rho\ge 0$. For the second case, notice that we can choose $\rho>0$ big enough so that $\nabla^2 f+\rho A^\top A\succ 0$. Since $-g$ dominates an affine function, the objective in the right-hand side of (EC.15) is level-bounded, and hence $v(\mu,\rho)>-\infty$.

Next we verify condition (20) for the third case. Denote $\nabla^2 f=F$ and $\nabla^2 g=G$. The Hessian of the objective in (x,y) in the right-hand side of (20) is positive-definite if $\mu < \lambda_{\max}(G)^{-1}$ and its Schur complement

$$\begin{aligned} S(\mu,\rho) :=& F + \rho A^\top A + \frac{1}{\mu} I_n - \left(-\frac{1}{\mu} I_n\right)\left(\frac{1}{\mu} I_n - G\right)^{-1}\left(-\frac{1}{\mu} I_n\right) \\ =& F + \rho A^\top A + \frac{1}{\mu} I_n - \frac{1}{\mu^2}\left[\mu I_n + \mu^2 G (I_n - \mu G)^{-1}\right] = F + \rho A^\top A - G(I_n - \mu G)^{-1} \end{aligned}$$

is positive-definite, where we use the Woodbury matrix identity in the second equality, and $I_n \in \mathbb{R}^{n\times n}$ denotes the identity matrix. Since $F \succ 0$ over the null space of A by assumption, we can choose $\rho > 0$ large enough so that $F + \rho A^\top A \succ 0$. Since $\lambda_{\max}(G(I-\mu G)^{-1}) \to \lambda_{\max}(G)$ as $\mu \to 0$, we know $S(\mu,\rho) \succ 0$ if the smallest eigenvalue of F over the null space of A is strictly greater than $\lambda_{\max}(G)$. This completes the proof. □

EC.4.2. Proof of Lemma 4

Proof. Similar to the derivation in (EC.11)-(EC.12), the descent of ψ in x and z are given as

$$\begin{aligned} \psi(x^k, z^k, \lambda^k) - \psi(x^{k+1}, z^k, \lambda^k) &\geq \left(\frac{\mu^{-1} - L_f}{2}\right) \|x^{k+1} - x^k\|^2, \\ \psi(x^{k+1}, z^k, \lambda^k) - \psi(x^{k+1}, z^{k+1}, \lambda^k) &\geq \frac{1}{\mu}\left(\frac{1}{\beta} - \frac{1}{2}\right) \|z^{k+1} - z^k\|^2. \end{aligned}$$

In addition,

$$\psi(x^{k+1}, z^{k+1}, \lambda^k) - \psi(x^{k+1}, z^{k+1}, \lambda^{k+1}) = \langle \lambda^k - \lambda^{k+1}, Ax^{k+1} - b\rangle = -\frac{1}{\rho}\|\lambda^{k+1} - \lambda^k\|^2.$$

Adding the above three expressions completes the proof. □

EC.4.3. Proof of Lemma 5

Proof. For $k \in \mathbb{N}$, the update of x^{k+1} gives $\nabla f(x^k) + A^\top \lambda^{k+1} + \mu^{-1}(x^{k+1} - z^k) = 0$, which implies that for $k \in \mathbb{N}$, $A^\top(\lambda^{k+1} - \lambda^k) = \mu^{-1}(x^k - x^{k+1}) + (\nabla f(x^{k-1}) - \nabla f(x^k)) + \mu^{-1}(z^k - z^{k-1})$. Since $\lambda^{k+1} - \lambda^k = \rho(Ax^{k+1} - b)$ belongs to the column space of A, we have

$$\sigma^+_{\min}(A)\|\lambda^{k+1} - \lambda^k\| \leq \|A^\top(\lambda^{k+1} - \lambda^k)\| \leq \mu^{-1}\|x^{k+1} - x^k\| + L_f\|x^k - x^{k-1}\| + \mu^{-1}\|z^k - z^{k-1}\|,$$

where the first inequality is due to the min-max theorem of eigenvalues of a real symmetric matrix. Dividing both sides by $\sigma^+_{\min}(A)$ gives the desired inequality. □

EC.4.4. Proof of Lemma 6

Proof. 1. By Lemma 5, squaring both sides gives

$$\|\lambda^{k+1} - \lambda^k\|^2 \leq c_3\|x^{k+1} - x^k\|^2 + c_4\|x^k - x^{k-1}\|^2 + c_3\|z^k - z^{k-1}\|^2.$$

Then (28) follows from Lemma 4 and constants defined in (27).

2. Recall that

$$\begin{aligned}\Psi_k \ge \psi(x^k, z^k, \lambda^k) \ge & f(x^k) - g(z^k) + \frac{\rho}{2}\|Ax^k - b\|^2 + \frac{1}{2\mu}\|x^k - z^k\|^2 + \frac{1}{\rho}\langle \lambda^k, \lambda^k - \lambda^{k-1}\rangle \\ \ge & v(\mu,\rho) + \frac{1}{2\rho}(\|\lambda^k\|^2 - \|\lambda^{k-1}\|^2),\end{aligned}$$

where the second inequality is due to $M_{\mu g}(z) \le g(z)$ and $\lambda^k = \lambda^{k-1} + \rho(Ax^k - b)$, and the third inequality is due to (20). This further implies that

$$\sum_{k=1}^{K} (\Psi_k - v(\mu,\rho)) \ge \frac{1}{2\rho}(\|\lambda^K\|^2 - \|\lambda^0\|^2) \ge -\frac{1}{2\rho}\|\lambda^0\|^2 > -\infty,$$

for all positive integer K. Since Ψ_k is non-increasing, we must have $\Psi_k \ge v(\mu,\rho)$ for all $k \in \mathbb{Z}_+$; otherwise, there exists some $\delta > 0$ such that $\Psi_k - v(\mu,\rho) < -\delta$ for all large enough k, then the above summation would converge to $-\infty$ as $K \to \infty$.

3. For any positive integer K, summing (28) from 0 to $K-1$ gives

$$\begin{aligned}&\kappa_{\min} K \min_{k=0,\cdots,K-1} \left\{ \|x^{k+1} - x^k\|^2 + \|z^{k+1} - z^k\|^2 + \|x^k - x^{k-1}\|^2 + \|z^k - z^{k-1}\|^2 \right\} \\ \le & \sum_{k=0}^{K-1} \left(\kappa_1\|x^{k+1} - x^k\|^2 + \kappa_2\|z^{k+1} - z^k\|^2 + \kappa_3\|x^k - x^{k-1}\|^2 + \kappa_4\|z^k - z^{k-1}\|^2 \right) \\ \le & \sum_{k=0}^{K-1} (\Psi_k - \Psi_{k+1}) = \Psi_0 - \Psi_K \le \Psi_0 - v(\mu,\rho),\end{aligned} \tag{EC.16}$$

where the last inequality is due to $\Psi_K \ge v(\mu,\rho)$ for all $K \ge 1$. Now let $\bar{k}$ be the minimizer in (EC.16); it follows that

$$\begin{aligned}&\max\left\{ \|x^{\bar{k}+1} - x^{\bar{k}}\|^2, \|z^{\bar{k}+1} - z^{\bar{k}}\|^2, \|x^{\bar{k}} - x^{\bar{k}-1}\|^2, \|z^{\bar{k}} - z^{\bar{k}-1}\|^2 \right\} \\ \le & \left(\|x^{\bar{k}+1} - x^{\bar{k}}\|^2 + \|z^{\bar{k}+1} - z^{\bar{k}}\|^2 + \|x^{\bar{k}} - x^{\bar{k}-1}\|^2 + \|z^{\bar{k}} - z^{\bar{k}-1}\|^2 \right) \le \frac{\Psi_0 - v(\mu,\rho)}{\kappa_{\min} K}.\end{aligned}$$

This completes the proof. □

EC.4.5. Proof of Lemma 7

Proof. By (34) and the μ-strong convexity of the function in the following line, it holds for all $x \in \mathcal{H}$ that

$$\begin{aligned}&\langle \nabla f(x^k) - \xi_g^k, x - x^k\rangle + h(x) + \langle \lambda^k, Ax - b\rangle + \frac{\rho}{2}\|Ax - b\|^2 + \frac{1}{2\mu}\|x - z^k\|^2 \\ \ge & \langle \nabla f(x^k) - \xi_g^k, x^{k+1} - x^k\rangle + h(x^{k+1}) + \langle \lambda^k, Ax^{k+1} - b\rangle + \frac{\rho}{2}\|Ax^{k+1} - b\|^2 \\ &+ \frac{1}{2\mu}\|x^{k+1} - z^k\|^2 + \langle \zeta^{k+1}, x - x^{k+1}\rangle + \frac{1}{2\mu}\|x^{k+1} - x\|^2.\end{aligned} \tag{EC.17}$$

Using the Lipschitz condition of ∇f, the convexity of g, inequality (EC.17) with $x = x^k$, the fact that $\langle \zeta^{k+1}, x^{k+1} - x^k \rangle \leq \frac{1}{4\mu}\|x^{k+1} - x^k\|^2 + \mu\|\zeta^{k+1}\|^2$, and the choice that $\|\zeta^{k+1}\| \leq \epsilon_{k+1}$, the descent in x can be derived as follows:

$$P(x^{k+1}, z^k, \lambda^k) \leq P(x^k, z^k, \lambda^k) + \mu\epsilon_{k+1}^2 - \frac{\mu^{-1} - 2L_f}{4}\|x^{k+1} - x^k\|^2.$$

The descent with respect to variable z can be derived as follows:

$$P(x^{k+1}, z^k, \lambda^k) - P(x^{k+1}, z^{k+1}, \lambda^k) = \frac{1}{2\mu}\left(\frac{2}{\beta} - 1\right)\left\|z^{k+1} - z^k\right\|^2 \geq \frac{1}{2\beta\mu}\left\|z^{k+1} - z^k\right\|^2,$$

where we replace x^{k+1} by $z^k + \frac{1}{\beta}(z^{k+1} - z^k)$ to get the equality, and the inequality is due to $\beta \leq 1$. Finally, similar to Lemma 4, the change in λ is

$$P(x^{k+1}, z^{k+1}, \lambda^k) - P(x^{k+1}, z^{k+1}, \lambda^{k+1}) = -\frac{1}{\rho}\|\lambda^{k+1} - \lambda^k\|^2.$$

Combining the above three expressions completes the proof. □

EC.4.6. Proof of Lemma 8

Proof. The optimality condition (EC.17) with $k = 0$ and $x = \bar{x}$ (by Assumption 4, $A\bar{x} = b$) gives

$$\begin{aligned}
&\langle \nabla f(x^0) - \xi_g^0, x^1 - x^0 \rangle + h(x^1) + \langle \lambda^0, Ax^1 - b \rangle + \frac{\rho}{2}\|Ax^1 - b\|^2 + \frac{1}{2\mu}\|x^1 - z^0\|^2 \\
\leq &\langle \nabla f(x^0) - \xi_g^0, \bar{x} - x^0 \rangle + h(\bar{x}) + \frac{1}{2\mu}\|\bar{x} - z^0\|^2 + \langle \zeta^1, x^1 - \bar{x} \rangle.
\end{aligned} \tag{EC.18}$$

Since $\|\zeta^1\| \leq \epsilon_1 \leq 1$, $\mu < L_f^{-1}$, and $x^1, \bar{x} \in \mathcal{H}$, we have

$$\langle \zeta^1, x^1 - \bar{x} \rangle \leq \frac{\mu}{2}\|\zeta^1\|^2 + \frac{1}{2\mu}\|x^1 - \bar{x}\|^2 \leq \frac{L_f^{-1}}{2} + \frac{1}{2\mu}D_{\mathcal{H}}^2. \tag{EC.19}$$

The above two inequalities together with the L_h-Lipschitz continuity of h and $\bar{x}, z^0 \in \mathcal{H}$ give that

$$\rho\|Ax^1 - b\|^2 \leq 2(M_{\nabla f} + M_{\partial g} + L_h)D_{\mathcal{H}} + 2\mu^{-1}D_{\mathcal{H}}^2 + L_f^{-1} + 2\|\lambda^0\| \max_{x \in \mathcal{H}} \|Ax - b\|. \tag{EC.20}$$

Notice that due to ∇f being Lipschitz and g being convex, $P(x^1, z^0, \lambda^0)$ is bounded from above by

$$\begin{aligned}
&f(x^0) - g(x^0) + \frac{L_f}{2}\|x^1 - x^0\|^2 + \langle \nabla f(x^0) - \xi_g^0, x^1 - x^0 \rangle + h(x^1) + \langle \lambda^0, Ax^1 - b \rangle + \frac{\rho}{2}\|Ax^1 - b\|^2 + \frac{1}{2\mu}\|x^1 - z^0\|^2 \\
\leq &f(x^0) - g(x^0) + \frac{L_f}{2}\|x^1 - x^0\|^2 + \langle \nabla f(x^0) - \xi_g^0, \bar{x} - x^0 \rangle + h(\bar{x}) + \frac{1}{2\mu}\|\bar{x} - z^0\|^2 + \langle \zeta^1, x^1 - \bar{x} \rangle \\
\leq &\max_{x \in \mathcal{H}}\{f(x) + h(x) - g(x)\} + (L_h + M_{\nabla f} + M_{\partial g})D_{\mathcal{H}} + \frac{L_f + 2\mu^{-1}}{2}D_{\mathcal{H}}^2 + \frac{L_f^{-1}}{2},
\end{aligned} \tag{EC.21}$$

where the first inequality is due to (EC.18), and the second inequality is due to the compactness of $\mathcal{H}$, the L_h-Lipschitz continuity of h, and (EC.19). By Lemma 7, (EC.20), and (EC.21), we have

$$P(x^1, z^1, \lambda^1) = P(x^1, z^1, \lambda^0) + \rho\|Ax^1 - b\|^2 \leq P(x^1, z^0, \lambda^0) + \rho\|Ax^1 - b\|^2 \leq \overline{P}.$$

This completes the proof. □

EC.4.7. Proof of Lemma 9

Proof. By step 4 in Algorithm 3, there exists $\xi_h^{k+1} \in \partial h(x^{k+1})$ such that

$$\zeta^{k+1} = \nabla f(x^k) - \xi_g^k + \xi_h^{k+1} + A^\top \lambda^{k+1} + \frac{1}{\mu}(x^{k+1} - z^k). \tag{EC.22}$$

Since $\lambda^0, b \in \mathrm{Im}(A)$ and $\lambda^{k+1} = \lambda^k + \rho(Ax^{k+1} - b)$, $\lambda^{k+1} \in \mathrm{Im}(A)$ for all $k \in \mathbb{N}$. Since $z^0 \in \mathcal{H}$, $x^k \in \mathcal{H}$, and $z^{k+1} = (1-\beta)z^k + \beta x^{k+1}$, $z^k \in \mathcal{H}$ for all $k \in \mathbb{N}$ as well. Consequently,

$$\|\lambda^{k+1}\| \le \frac{1}{\sigma_{\min}^+(A)} \left(M_{\nabla f} + M_{\partial g} + \frac{D_{\mathcal{H}}}{\mu} + 1 \right) + \frac{\|\xi_h^{k+1}\|}{\sigma_{\min}^+(A)}. \tag{EC.23}$$

By (Melo et al. 2020, Lemma 4.7), we can bound $\|\xi_h^{k+1}\|$ as follows:

$$\bar{d}\|\xi_h^{k+1}\| \le (\bar{d} + \|x^{k+1} - \bar{x}\|)L_h + \left\langle \xi_h^{k+1}, x^{k+1} - \bar{x} \right\rangle \le 2D_{\mathcal{H}}L_h + \left\langle \xi_h^{k+1}, x^{k+1} - \bar{x} \right\rangle. \tag{EC.24}$$

Using (EC.22), we can further bound the inner product term in (EC.24) by

$$\begin{aligned}
\langle \xi_h^{k+1}, x^{k+1} - \bar{x} \rangle &= \left\langle \zeta^{k+1} - \nabla f(x^k) + \xi_g^k - A^\top \lambda^{k+1} - \frac{1}{\mu}(x^{k+1} - z^k), x^{k+1} - \bar{x} \right\rangle \\
&\le \left(M_{\nabla f} + M_{\partial g} + \frac{D_{\mathcal{H}}}{\mu} + 1 \right) D_{\mathcal{H}} - \left\langle \lambda^{k+1}, Ax^{k+1} - b \right\rangle \\
&\le \left(M_{\nabla f} + M_{\partial g} + \frac{D_{\mathcal{H}}}{\mu} + 1 \right) D_{\mathcal{H}} + \frac{1}{\rho}\|\lambda^{k+1}\|\|\lambda^k\| - \frac{1}{\rho}\|\lambda^{k+1}\|^2,
\end{aligned} \tag{EC.25}$$

where we use the facts that $A\bar{x} = b$ and $\|\zeta^{k+1}\| \le \epsilon_{k+1} \le 1$ in the first inequality, and $Ax^{k+1} - b = \frac{1}{\rho}(\lambda^{k+1} - \lambda^k)$ to get the second inequality. Combining (EC.23), (EC.24) and (EC.25), we have

$$\frac{\|\lambda^{k+1}\|^2}{\rho\sigma_{\min}^+(A)} + \bar{d}\|\lambda^{k+1}\| \le \frac{\|\lambda^{k+1}\|\|\lambda^k\|}{\rho\sigma_{\min}^+(A)} + \frac{2D_{\mathcal{H}}}{\sigma_{\min}^+(A)} \left(M_{\nabla f} + M_{\partial g} + \frac{D_{\mathcal{H}}}{\mu} + L_h + 1 \right),$$

which further implies that, for all $k \in \mathbb{N}$,

$$\left(\frac{\|\lambda^{k+1}\|}{\rho\sigma_{\min}^+(A)} + \bar{d} \right) \|\lambda^{k+1}\| \le \frac{\|\lambda^{k+1}\|}{\rho\sigma_{\min}^+(A)} \|\lambda^k\| + \bar{d}\Lambda,$$

The claim then follows from an inductive argument: if $\|\lambda^k\| \le \Lambda$, then the above inequality implies that $\|\lambda^{k+1}\| \le \Lambda$ as well, and $\|\lambda^0\| \le \Lambda$ holds by the definition of Λ. □

EC.4.8. Proof of Lemma 10

Proof. 1. For all $k \in \mathbb{N}$,

$$\begin{aligned}
P(x^k, z^k, \lambda^k) &= f(x^k) + h(x^k) - g(x^k) + \langle \lambda, Ax^k - b \rangle + \frac{\rho}{2}\|Ax^k - b\|^2 + \frac{1}{2\mu}\|x^k - z^k\|^2 \\
&\ge \min_{x \in \mathbb{R}^n} \{f(x) + h(x) - g(x)\} - \Lambda \max_{x \in \mathcal{H}} \|Ax - b\| > -\infty,
\end{aligned}$$

where the inequality is due to the continuity of f, h, g, and $\|\cdot\|$ over compact domain $\mathcal{H}$.

2. Lemma 7 implies that, for all $k \in \mathbb{N}$,

$$\eta \left(\|x^{k+1} - x^k\|^2 + \|z^{k+1} - z^k\|^2 \right) \leq P(x^k, z^k, \lambda^k) - P(x^{k+1}, z^{k+1}, \lambda^{k+1}) + \frac{1}{\rho}\|\lambda^{k+1} - \lambda^k\|^2 + \mu\epsilon_{k+1}^2.$$

Summing the above inequality over $k = 1, \ldots, K$,

$$\begin{aligned} &\eta K \min_{k\in[K]} \left\{ \|x^{k+1} - x^k\|^2 + \|z^{k+1} - z^k\|^2 \right\} \leq \eta \sum_{k=1}^{K} \left\{ \|x^{k+1} - x^k\|^2 + \|z^{k+1} - z^k\|^2 \right\} \\ \leq & P(x^1, z^1, \lambda^1) - P(x^{K+1}, z^{K+1}, \lambda^{K+1}) + \frac{1}{\rho}\sum_{k=1}^{K} \|\lambda^{k+1} - \lambda^k\|^2 + \mu \sum_{k=1}^{K} \epsilon_{k+1}^2 \\ \leq & \overline{P} - \underline{P} + \mu E + \frac{4}{\rho}\sum_{k=1}^{K+1} \|\lambda^k\|^2 \leq \overline{P} - \underline{P} + \mu E + \frac{4(K+1)\Lambda^2}{\rho}, \end{aligned}$$

where we use the the fact that $P(x^{K+1}, z^{K+1}, \lambda^{K+1}) \geq \underline{P}$ for all integer $K \in \mathbb{Z}_+$ and $\|\lambda^{k+1} - \lambda^k\|^2 \leq 2\|\lambda^{k+1}\|^2 + 2\|\lambda^k\|^2$ in the third inequality, and Lemma 9 in the last inequality. Let $\bar{k}$ be the minimizer in the first line of the above chain of inequalities, then dividing both sides by any positive integer K gives

$$\max\left\{ \|x^{\bar{k}+1} - x^{\bar{k}}\|^2, \|z^{\bar{k}+1} - z^{\bar{k}}\|^2 \right\} \leq \frac{\overline{P} - \underline{P} + \mu E}{\eta K} + \frac{8\Lambda^2}{\eta\rho}.$$

This completes the proof. $\square$