A Proofs

A.1 Proof of Lemma 3.4

In parametric minimization problems, such as encountered in the definitions of ϕ-prox and ϕ-envelope, a sufficient condition for the continuity of the arg min map is given by uniform level boundedness [22, Definition 1.16 and Theorem 1.17] of the map $h : (z, v) \mapsto f(z) + \frac{1}{\lambda}\phi(v - z)$.

Definition A.1 (uniform level boundedness). *We say a function $h : \mathbb{R}^m \times \mathbb{R}^m \to \overline{\mathbb{R}}$ with values $h(z, v)$ is level-bounded in z locally uniformly in v if for each $\bar{v} \in \mathbb{R}^m$ and $\alpha \in \mathbb{R}$ there is a neighborhood V of $\bar{v}$ along with a bounded set $X \subset \mathbb{R}^m$ such that*

$$\{z : h(z, v) \leq \alpha\} \subset X$$

for all $v \in V$.

In the next lemma we establish the uniform level boundedness of the map $h : (z, v, \xi) \mapsto f(z) + \frac{1}{\lambda}\phi(v - z) - \langle \xi, z\rangle$ from ϕ-prox-boundedness so that [22, Theorem 1.17] can be invoked to assert the continuity of ϕ-prox and ϕ-envelope. The Lemma is stated in a more general form including an additional linear term $\langle \xi, z\rangle$, which is needed later on in the proof of Theorem 4.3, see Section A.2.

Lemma A.2. *Let $f : \mathbb{R}^m \to \overline{\mathbb{R}}$ be proper lsc and ϕ-prox-bounded with threshold $\lambda_f > 0$. Then for any $\lambda \in (0, \lambda_f)$, the function $h : \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m \to \overline{\mathbb{R}}$, defined via*

$$h(z, \xi, v) := f(z) + \frac{1}{\lambda}\phi(v - z) - \langle \xi, z\rangle,$$

is level-bounded in z locally uniformly in (ξ, v).

Proof. We assume the contrary: More precisely let $\lambda \in (0, \lambda_f)$ and assume that h is not level-bounded in z locally uniformly in v. On the one hand, this means that there exist $\bar{v}, \bar{\xi} \in \mathbb{R}^m$, $\alpha \in \mathbb{R}$ and sequences $v^\nu \to \bar{v}$, $\xi^\nu \to \bar{\xi}$ and z^ν, $\|z^\nu\| \to \infty$ such that

$$f(z^\nu) + \frac{1}{\lambda}\phi(v^\nu - z^\nu) - \langle \xi^\nu, z^\nu\rangle \leq \alpha.$$

On the other hand, we know that

$$f(z^\nu) + \frac{1}{\lambda'}\phi(v^\nu - z^\nu) \geq \beta,$$

for some $\lambda' > \lambda$, with $\lambda' \in (0, \lambda_f)$ and ν sufficiently large. Summing the inequalities yields:

$$\left(\frac{1}{\lambda} - \frac{1}{\lambda'}\right)\phi(v^\nu - z^\nu) - \langle \xi^\nu, z^\nu\rangle \leq \alpha - \beta.$$

Due to the super-coercivity of ϕ, for $\nu \to \infty$ this yields $\infty \leq \alpha - \beta$, a contradiction. □

Now we are ready to prove Lemma 3.4 invoking [22, Theorem 1.17]:

Lemma. *Let $f : \mathbb{R}^m \to \overline{\mathbb{R}}$ be proper lsc and ϕ-prox-bounded with threshold $\lambda_f > 0$. Then for any $\lambda \in (0, \lambda_f)$, $P^\phi_\lambda f$ and $e^\phi_\lambda f$ have the following properties:*

(i) *$P^\phi_\lambda f(v) \neq \emptyset$ is compact for all $v \in \operatorname{dom} e^\phi_\lambda f = \operatorname{dom} f + \operatorname{dom} \phi$, whereas $P^\phi_\lambda f(v) = \emptyset$ for $v \notin \operatorname{dom} e^\phi_\lambda f$.*

(ii) *The ϕ-envelope $e^\phi_\lambda f$ is continuous relative to $\operatorname{dom} e^\phi_\lambda f$.*

(iii) *For any sequence $v^\nu \to \bar{v}$ contained in $\operatorname{dom} e^\phi_\lambda f$ and $z^\nu \in P^\phi_\lambda f(v^\nu)$ we have $\{z^\nu\}_{\nu\in\mathbb{N}}$ is bounded and all its cluster points $\bar{z}$ lie in $P^\phi_\lambda f(\bar{v})$.*

Proof. Obviously it holds for the domain that $\operatorname{dom} e^\phi_\lambda f = \operatorname{dom} f + \operatorname{dom} \phi$. In view of Lemma A.2 (with $\xi = 0$) we assert that $h : (z, v) \mapsto f(z) + \frac{1}{\lambda}\phi(v - z)$ is level-bounded in z locally uniformly in v. Then we may invoke [22, Theorem 1.17] to assert that $P^\phi_\lambda f(v) \neq \emptyset$ is compact for any $v \in \operatorname{dom} e^\phi_\lambda f$ whereas $P^\phi_\lambda f(v) = \emptyset$ for $v \notin \operatorname{dom} e^\phi_\lambda f$

and in addition for any $\bar{v} \in \operatorname{dom} e_\lambda^\phi f$ and any sequence $z^\nu \in P_\lambda^\phi f(v^\nu)$ with $v^\nu \to \bar{v}$ contained in $\operatorname{dom} e_\lambda^\phi f$, that $\{z^\nu\}_{\nu\in\mathbb{N}}$ is bounded. Furthermore, as ϕ is continuous relative to its domain, we know for some $\bar{z} \in P_\lambda^\phi f(\bar{v})$ that $h(\bar{z}, \cdot)$ is continuous relative to $\bar{z} + \operatorname{dom}\phi$ containing $\bar{v}$. Through [22, Theorem 1.17] all cluster points of the sequence $z^\nu \in P_\lambda^\phi f(v^\nu)$ lie in $P_\lambda^\phi f(\bar{v})$ and $e_\lambda^\phi f(v^\nu) \to e_\lambda^\phi f(\bar{v})$ and therefore $e_\lambda^\phi f$ is continuous at $\bar{v}$ relative to $\operatorname{dom} e_\lambda^\phi f$. Since this holds for all $\bar{v} \in \operatorname{dom} e_\lambda^\phi f$, $e_\lambda^\phi f$ is continuous relative to $\operatorname{dom} e_\lambda^\phi f$. ☐

A.2 Proof of Theorem 4.3

In order to prove the desired statement we need the following intermediate result. For the sake of notational convenience recall the notion of Bregman distances as a short-hand notation for $\phi(w') - \phi(w) - \langle \nabla\phi(w), w' - w\rangle$:

Definition A.3 (Bregman distance)**.** *The ϕ-induced Bregman distance $B_\phi : \mathbb{R}^m \times \mathbb{R}^m \to \overline{\mathbb{R}}$ is defined by*

$$B_\phi(w', w) = \begin{cases} \phi(w') - \phi(w) - \langle \nabla\phi(w), w' - w\rangle & \text{if } w \in \operatorname{int}(\operatorname{dom}\phi) \\ +\infty & \text{otherwise.} \end{cases} \tag{19}$$

Lemma A.4. *Let $f : \mathbb{R}^m \to \overline{\mathbb{R}}$ proper lsc and ϕ-prox-bounded with threshold λ_f. In addition assume that f is finite and prox-regular at $\bar{z}$ for $\bar{y} \in \partial f(\bar{z})$ such that the subgradient inequality (11) is satisfied by constants $r > 0$ and $\epsilon > 0$ and let $\bar{v} \in \bar{z} + \operatorname{dom}\phi$. Then the following inequality holds for all $z \in \mathbb{R}^m$ and $r_1 \geq \max\{r, \lambda_f^{-1}\}$ sufficiently large:*

$$f(z) \geq f(\bar{z}) + \langle \bar{y}, z - \bar{z}\rangle - r_1 B_{\phi(\bar{v}-\cdot)}(z, \bar{z}). \tag{20}$$

Proof. By prox-regularity of f we know there exist $r > 0$ and $\epsilon > 0$ such that the subgradient inequality

$$f(z) \geq f(\bar{z}) + \langle \bar{y}, z - \bar{z}\rangle - \frac{r}{2}\|z - \bar{z}\|^2, \tag{21}$$

holds for $\|z - \bar{z}\| < \epsilon$. By Assumption (A3) and [4, Proposition 2.10] we have that

$$B_{\phi(\bar{v}-\cdot)}(z, \bar{z}) = \phi(\bar{v} - z) - \phi(\bar{v} - \bar{z}) + \langle \nabla\phi(\bar{v} - \bar{z}), z - \bar{z}\rangle \geq \frac{\mu}{2}\|z - \bar{z}\|^2, \tag{22}$$

for some $\mu > 0$. Summing (21) and (22) yields (20), which holds for any z with $\|z - \bar{z}\| < \epsilon$ and $r_1 \geq \frac{r}{\mu}$. To show the assertion we prove that (20) also holds for any z with $\|z - \bar{z}\| \geq \epsilon$ for $r_1 \geq \max\{\frac{r}{\mu}, \lambda_f^{-1}\}$ sufficiently large. By ϕ-prox-boundedness it holds for $\lambda \in (0, \lambda_f)$ and $\bar{v} \in \operatorname{dom} e_\lambda^\phi f$ that $+\infty > e_\lambda^\phi f(\bar{v}) > -\infty$. We have

$$f(z) \geq e_\lambda^\phi f(\bar{v}) - \frac{1}{\lambda}\phi(\bar{v} - z), \tag{23}$$

showing, that the desired inequality (20) is implied by

$$e_\lambda^\phi f(\bar{v}) - \frac{1}{\lambda}\phi(\bar{v} - z) \geq f(\bar{z}) + \langle \bar{y}, z - \bar{z}\rangle - r_1 B_{\phi(\bar{v}-\cdot)}(z, \bar{z}),$$

which is equivalent to

$$(r_1 - \lambda^{-1}) B_{\phi(\bar{v}-\cdot)}(z, \bar{z}) \geq f(\bar{z}) - e_\lambda^\phi f(\bar{v}) + \langle \bar{y}, z - \bar{z}\rangle + \frac{1}{\lambda}\phi(\bar{v} - \bar{z}) - \frac{1}{\lambda}\left\langle \nabla\phi(\bar{v} - \bar{z}), z - \bar{z}\right\rangle,$$

and (using Cauchy-Schwarz) implied by

$$(r_1 - \lambda^{-1})\frac{B_{\phi(\bar{v}-\cdot)}(z, \bar{z})}{\|z - \bar{z}\|} \geq \frac{f(\bar{z}) - e_\lambda^\phi f(\bar{v}) + \frac{1}{\lambda}\phi(\bar{v} - \bar{z})}{\|z - \bar{z}\|} + \|\bar{y} - \lambda^{-1}\nabla\phi(\bar{v} - \bar{z})\|. \tag{24}$$

Due to the super-coercivity of ϕ (24) holds for z with $\|z - \bar{z}\| \geq \gamma$ for some $\gamma > 0$ sufficiently large. To make (24) also hold for z with $\epsilon \leq \|z - \bar{z}\| < \gamma$ we can choose $r_1 > \max\{r, \lambda_f^{-1}\}$ sufficiently large as $B_{\phi(\bar{v}-\cdot)}(z, \bar{z})$ is bounded away from zero, due to the strict convexity of $\phi(\bar{v} - \cdot)$. ☐

Throughout the proof we work with graphical localizations of set-valued mappings F, which are constructed graphically by intersecting the graph of F with some neighborhood of some reference point $(\bar{z}, \bar{y}) \in \operatorname{gph} F$:

Definition A.5 (graphical localization). *For $F : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ and a pair $(\bar z, \bar y) \in \operatorname{gph} F$, a graphical localization of F at $\bar z$ for $\bar y$ is a set-valued mapping T such that*

$$\operatorname{gph} T = (U \times V) \cap \operatorname{gph} F$$

for some neighborhoods U of $\bar z$ and V of $\bar y$, so that

$$T(z) := \begin{cases} F(z) \cap V & \text{if } z \in U, \\ \emptyset & \text{otherwise.} \end{cases} \tag{25}$$

In the proof of Theorem 4.3, we shall invoke the following generalized implicit function theorem specialized from [11, Theorem 2B.7], originally due to Robinson [20] for analyzing the solutions of parametric variational inequalities.

Theorem A.6 (generalized implicit function theorem). *Consider a function $G : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ and a set-valued map $T : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ with $(\bar v, \bar z) \in \operatorname{int} \operatorname{dom} G$ and $0 \in G(\bar v, \bar z) + T(\bar z)$, and suppose that*

$$\widehat{\operatorname{lip}}_v(G; (\bar v, \bar z)) := \limsup_{\substack{v, v' \to \bar v \\ z \to \bar z \\ v \neq v'}} \frac{\|G(v, z) - G(v', z)\|}{\|v - v'\|} \leq \gamma < \infty.$$

Let $H : \mathbb{R}^m \to \mathbb{R}^m$ be a strict estimator of G w.r.t. z uniformly in v at $(\bar v, \bar z)$ with constant μ, i.e.,

$$\widehat{\operatorname{lip}}_z(e; (\bar v, \bar z)) \leq \mu < \infty \quad \text{for } e(v, z) = G(v, z) - H(z).$$

Suppose that $(H + T)^{-1}$ has a Lipschitz continuous single-valued localization around 0 for $\bar z$ with modulus κ and $\kappa\mu < 1$. Then the solution mapping

$$S : v \in \mathbb{R}^m \mapsto \{z \in \mathbb{R}^m : 0 \in G(v, z) + T(z)\}$$

has a Lipschitz continuous single-valued localization around $\bar v$ for $\bar z$ with modulus $\frac{\kappa\gamma}{1 - \kappa\mu}$.

We are now ready to prove the desired statement Theorem 4.3:

Theorem. *Let $f : \mathbb{R}^m \to \overline{\mathbb{R}}$ proper lsc and ϕ-prox-bounded with threshold λ_f. Let $\bar v \in \bar z + \operatorname{dom} \phi$. Then for any $\lambda \in (0, \lambda_f)$ sufficiently small and f finite and prox-regular at $\bar z$ for $\bar y \in \partial f(\bar z)$ with*

$$\bar y = \frac{1}{\lambda} \nabla \phi(\bar v - \bar z)$$

the following statements hold true:

(i) *$P^\phi_\lambda f$ is a singled-valued, Lipschitz map near $\bar v$ such that $\bar z = P^\phi_\lambda f(\bar v)$ and*

$$P^\phi_\lambda f(v) = (I + \nabla \phi^* \circ \lambda T)^{-1}(v), \tag{26}$$

where T is the f-attentive ϵ-localization of ∂f near $(\bar z, \bar y)$, i.e. the set-valued mapping $T : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ defined by $T(z) := \{y \in \partial f(z) : \|y - \bar y\| < \epsilon\}$ if $\|z - \bar z\| < \epsilon$ and $f(z) < f(\bar z) + \epsilon$, and $T(z) := \emptyset$ otherwise.

(ii) *$e^\phi_\lambda f$ is Lipschitz differentiable around $\bar v$ with*

$$\nabla e^\phi_\lambda f(v) = \frac{1}{\lambda} \nabla \phi(v - z). \tag{27}$$

Proof. (i) Due to the prox-regularity at $\bar z$ for $\bar y$ with constants $\epsilon > 0$ and $r > 0$ and the ϕ-prox-boundedness of f with threshold λ_f we can invoke Lemma A.4 to assert that for some $\lambda \in (0, \min\{\lambda_f, r^{-1}\})$ sufficiently small the following inequality holds *globally* for all $z \neq \bar z$:

$$f(z) + \frac{1}{\lambda}\phi(\bar v - z) > f(\bar z) + \frac{1}{\lambda}\phi(\bar v - \bar z) + \langle \bar y, z - \bar z \rangle - \frac{1}{\lambda} \langle \nabla \phi(\bar v - \bar z), z - \bar z \rangle . \tag{28}$$

By assumption $\bar{y} = \frac{1}{\lambda}\nabla\phi(\bar{v} - \bar{z})$ showing that

$$f(z') + \frac{1}{\lambda}\phi(\bar{v} - z') > f(\bar{z}) + \frac{1}{\lambda}\phi(\bar{v} - \bar{z}),$$

for any $z' \neq \bar{z}$. Therefore we can assert $P^\phi_\lambda f(\bar{v}) = \bar{z}$.

Due to the ϕ-prox-boundedness of f and the coercivity of ϕ we can invoke Lemma A.2 to assert that $h(z, v) := f(z) + \frac{1}{\lambda}\phi(v - z)$ is level-bounded in z locally uniformly in v. By Lemma 3.4 it follows that $P^\phi_\lambda f(v) \neq \emptyset$ for any $v \in \operatorname{dom} e^\phi_\lambda f$. Furthermore, we assert, that for any sequence $z^\nu \in P^\phi_\lambda f(v^\nu)$, $v^\nu \to \bar{v}$ we have $\{z^\nu\}$ is bounded and all its cluster points lie in $P^\phi_\lambda f(\bar{v}) = \bar{z}$, meaning $z^\nu \to \bar{z}$ and $e^\phi_\lambda f(v^\nu) \to e^\phi_\lambda(\bar{v})$. In addition, we have $f(z^\nu) \to f(\bar{z})$ as $e^\phi_\lambda f(v^\nu) = f(z^\nu) + \frac{1}{\lambda}\phi(v^\nu - z^\nu) \to e^\phi_\lambda f(\bar{v}) = f(\bar{z}) + \frac{1}{\lambda}\phi(\bar{v} - \bar{z})$. Overall this shows, that for any v, sufficiently near $\bar{v}$ we have $z \in P^\phi_\lambda f(v)$, $\|z - \bar{z}\| < \epsilon$, $|f(z) - f(\bar{z})| < \epsilon$ and $\|\frac{1}{\lambda}\nabla\phi(v - z) - \bar{y}\| < \epsilon$, due to the continuity of $\nabla\phi$ on a neighborhood of $\bar{v} - \bar{z}$. From applying Fermat's rule [22, Theorem 10.1] to $P^\phi_\lambda f(v)$ we obtain

$$\frac{1}{\lambda}\nabla\phi(v - z) \in \partial f(z),$$

and we can assert that $\frac{1}{\lambda}\nabla\phi(v - z) \in T(z)$ via the arguments above. This shows that

$$\emptyset \neq P^\phi_\lambda f(v) \subset \left(T - \frac{1}{\lambda}\nabla\phi(v - \cdot)\right)^{-1}(0), \tag{29}$$

where ∂f is replaced by T. It is straightforward to verify, using the fact $(\nabla\phi)^{-1} = \nabla\phi^*$ from Lemma 3.2, that

$$\left(T - \frac{1}{\lambda}\nabla\phi(v - \cdot)\right)^{-1}(0) = (I + \nabla\phi^* \circ \lambda T)^{-1}(v).$$

We proceed showing that $(I + \nabla\phi^* \circ \lambda T)^{-1}$ is a single-valued Lipschitz map in a neighborhood of $\bar{v}$. Via the inclusion above this means that $P^\phi_\lambda f = (I + \nabla\phi^* \circ \lambda T)^{-1}$ on this neighborhood.

To this end, consider the function $h : \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m \to \overline{\mathbb{R}}$ defined via $h(v, \xi, z) := f(z) + \frac{1}{\lambda}\phi(v - z) - \langle \xi, z\rangle$. In view of Lemma A.2 we assert that $h(v, \xi, z)$ is level-bounded in z locally uniformly in (v, ξ). Through [22, Theorem 1.17] we know that for any sequence $\xi^\nu \to 0$ with $\inf_{z\in\mathbb{R}^m} h(\bar{v}, \xi^\nu, z) < \infty$ there is $z^\nu \in \arg\min_{z\in\mathbb{R}^m} h(\bar{v}, \xi^\nu, z) \neq \emptyset$ with $z^\nu \to \bar{z} = \arg\min_{z\in\mathbb{R}^m} h(\bar{v}, 0, z) = P^\phi_\lambda f(\bar{v})$ and $\inf_{z\in\mathbb{R}^m} h(\bar{v}, \xi^\nu, z) \to \inf_{z\in\mathbb{R}^m} h(\bar{v}, 0, z) = e^\phi_\lambda f(\bar{v})$.

From applying Fermat's rule [22, Theorem 10.1] to the minimization problem above we know that $\xi^\nu + \frac{1}{\lambda}\nabla\phi(\bar{v} - z^\nu) \in \partial f(z^\nu)$ and for ν sufficiently large we have that $\|\xi^\nu + \frac{1}{\lambda}\nabla\phi(\bar{v} - z^\nu) - \bar{y}\| \leq \epsilon$ due to the continuity of $\nabla\phi$ on a neighborhood of $\bar{v} - \bar{z}$. In addition, we have $f(z^\nu) \to f(\bar{z})$ as $\inf_{z\in\mathbb{R}^m} h(\bar{v}, \xi^\nu, z) = f(z^\nu) + \frac{1}{\lambda}\phi(\bar{v} - z^\nu) - \langle \xi^\nu, z^\nu\rangle \to e^\phi_\lambda f(\bar{v}) = f(\bar{z}) + \frac{1}{\lambda}\phi(\bar{v} - \bar{z})$. Overall this means that for any ξ sufficiently near 0 we have:

$$\xi \in T(z) - \frac{1}{\lambda}\nabla\phi(\bar{v} - z),$$

for some z near $\bar{z}$, and ∂f is interchangeable with T.

Now pick any $(z^1, y^1), (z^2, y^2) \in \operatorname{gph} T$. Then it follows from (11) that

$$f(z^2) \geq f(z^1) + \left\langle y^1, z^2 - z^1\right\rangle - \frac{r}{2}\|z^2 - z^1\|^2, \tag{30}$$

$$f(z^1) \geq f(z^2) + \left\langle y^2, z^1 - z^2\right\rangle - \frac{r}{2}\|z^1 - z^2\|^2, \tag{31}$$

and furthermore, due to Assumption (A3) and [4, Proposition 2.10] we have that

$$B_{\phi(\bar{v}-\cdot)}(z^2, z^1) = \phi(\bar{v} - z^2) - \phi(\bar{v} - z^1) + \left\langle \nabla\phi(\bar{v} - z^1), z^2 - z^1\right\rangle \geq \frac{\mu}{2}\|z^2 - z^1\|^2.$$

for some $\mu > 0$. Then we have

$$\phi(\bar{v} - z^2) \geq \phi(\bar{v} - z^1) - \left\langle \nabla\phi(\bar{v} - z^1), z^2 - z^1\right\rangle + \frac{\mu}{2}\|z^2 - z^1\|^2, \tag{32}$$

$$\phi(\bar{v} - z^1) \geq \phi(\bar{v} - z^2) - \left\langle \nabla\phi(\bar{v} - z^2), z^1 - z^2\right\rangle + \frac{\mu}{2}\|z^1 - z^2\|^2. \tag{33}$$

Summing the four inequalities yields for $\xi^1 := y^1 - \frac{1}{\lambda}\nabla\phi(\bar{v} - z^1)$ and $\xi^2 := y^2 - \frac{1}{\lambda}\nabla\phi(\bar{v} - z^2)$:

$$\left\langle z^1 - z^2, \xi^1 - \xi^2 \right\rangle \geq (\tfrac{\mu}{\lambda} - r)\|z^1 - z^2\|^2, \tag{34}$$

Consequently, the map $T - \frac{1}{\lambda}\nabla\phi(\bar{v} - \cdot)$ is $(\frac{\mu}{\lambda} - r)$-strongly monotone.

Define $H(z) := -\frac{1}{\lambda}\nabla\phi(\bar{v} - z)$, $G(v, z) := -\frac{1}{\lambda}\nabla\phi(v - z)$ and $e(v, z) := G(v, z) - H(z)$. Then the above argument implies that $\xi \mapsto (T + H)^{-1}(\xi)$ is a single-valued, $(\frac{\mu}{\lambda} - r)^{-1}$-Lipschitz map in a neighbourhood of 0 such that $0 \in T(\bar{z}) + G(\bar{v}, \bar{z})$. Due to Assumption (A3) and [4, Proposition 2.10] $\nabla\phi$ is Lipschitz on a neighbourhood of $\bar{v} - \bar{z} \in \operatorname{int}\operatorname{dom}\phi$ and we may conclude that $\widehat{\operatorname{lip}}_v(G; (\bar{v}, \bar{z}))$ and $\widehat{\operatorname{lip}}_z(e; (\bar{v}, \bar{z}))$ are finite. This implies that H is a strict estimator of G w.r.t. z uniformly in v at $(\bar{v}, \bar{z}) \in \operatorname{int}\operatorname{dom} G$. Invoking Theorem A.6, we assert that $v \mapsto \{z \in \mathbb{R}^m : 0 \in G(v, z) + T(z)\}$ has a single-valued, Lipschitz localization around $\bar{v}$ for $\bar{z}$. Since $\bar{z} = P_\lambda^\phi f(\bar{v})$ is single-valued at $\bar{v}$ and for any v near $\bar{v}$ and $z \in P_\lambda^\phi f(v)$ it holds that z is near $\bar{z}$, we may conclude that $P_\lambda^\phi f(v) = \{z \in \mathbb{R}^m : 0 \in G(v, z) + T(z)\}$ is single-valued and Lipschitz on a neighborhood of $\bar{v}$.

(ii) Let (z, v), (z', v') be sufficiently close to $(\bar{z}, \bar{v})$ with $z = P_\lambda^\phi f(v)$, $z' = P_\lambda^\phi f(v')$. Then, by Fermat's rule [22, Theorem 10.1] it holds $y \in \partial f(z)$ such that $y = \frac{1}{\lambda}\nabla\phi(v - z)$. Furthermore, by assumption the subgradient inequality (11) holds true at $(z, y) \in \operatorname{gph} T$. This means in particular that y is a proximal subgradient [22, Definition 8.45] of f at z. In view of [22, Proposition 8.46 (e)], we assert $y \in \hat{\partial} f(z)$. Thus (and using the differentiability of ϕ on $\operatorname{dom}\phi$) one can derive

$$\begin{aligned} e_\lambda^\phi f(v') - e_\lambda^\phi f(v) &= f(z') - f(z) + \frac{1}{\lambda}\phi(v' - z') - \frac{1}{\lambda}\phi(v - z) \\ &\geq \langle y, z' - z\rangle + o(\|z' - z\|) + \frac{1}{\lambda}\langle \nabla\phi(v - z), (v' - v) - (z' - z)\rangle + o(\|(v' - v) - (z' - z)\|). \end{aligned} \tag{35}$$

Using the conclusion from (i) that $v \mapsto z$ is a Lipschitz map near $\bar{v}$ there is some α such that

$$\|z' - z\| \leq \alpha\|v' - v\|.$$

This shows that $o(\|(v' - v) - (z' - z)\|) + o(\|z' - z\|) = o(\|v' - v\|)$ and we get from (35) that

$$e_\lambda^\phi f(v') - e_\lambda^\phi f(v) - \frac{1}{\lambda}\langle \nabla\phi(v - z), v' - v\rangle \geq o(\|v' - v\|). \tag{36}$$

On the other hand, we have

$$e_\lambda^\phi f(v') = \inf_{z''} f(z'') + \frac{1}{\lambda}\phi(v' - z'') \leq f(z) + \frac{1}{\lambda}\phi(v' - z). \tag{37}$$

Due to the differentiability of ϕ, we have

$$\phi(v' - z) = \phi(v - z) + \langle \nabla\phi(v - z), v' - v\rangle + o(\|v' - v\|). \tag{38}$$

This yields

$$\begin{aligned} e_\lambda^\phi f(v') - e_\lambda^\phi f(v) &\leq f(z) + \frac{1}{\lambda}\phi(v' - z) - f(z) - \frac{1}{\lambda}\phi(v - z) \\ &= \frac{1}{\lambda}\langle \nabla\phi(v - z), v' - v\rangle + o(\|v' - v\|). \end{aligned} \tag{39}$$

Combining (36) and (39), we conclude that $e_\lambda^\phi f$ is differentiable at v and $\frac{1}{\lambda}\nabla\phi(v - z) = \nabla e_\lambda^\phi f(v)$. Since furthermore $\nabla\phi$ is Lipschitz around $\bar{v} - \bar{z}$ it holds that $\nabla e_\lambda^\phi f$ is the composition of two locally Lipschitz maps and therefore Lipschitz around $\bar{v}$. ☐

References

[1] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. *Mathematics of Operations Research*, 35:438–457, 2010.

[2] A. Auslender. *Optimisation: Méthodes Numériques.* Masson, 1976.

[3] H. H. Bauschke, J. M. Borwein, and P. L. Combettes. Bregman monotone optimization algorithms. *SIAM Journal on Control and Optimization*, 42(2):596–636, 2003.

[4] H. H. Bauschke and A. S. Lewis. Dykstras algorithm with Bregman projections: A convergence proof. *Optimization*, 48(4):409–427, 2000.

[5] D. P. Bertsekas and J. N. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods.* Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1989.

[6] J. Bolte, S. Sabach, and M. Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146(1-2):459–494, 2014.

[7] Y. Censor and S. A. Zenios. Proximal minimization algorithm with D-functions. *Journal of Optimization Theory and Applications*, 73(3):451–464, 1992.

[8] P. Chaudhari, A. Oberman, S. Osher, S. Soatto, and G. Carlier. Deep relaxation: partial differential equations for optimizing deep neural networks. *Research in the Mathematical Sciences*, 5(3):30, 2018.

[9] P. L. Combettes and N. N. Reyes. Moreau's decomposition in Banach spaces. *Mathematical Programming*, 139(1-2):103–114, 2013.

[10] Damek Davis and Dmitriy Drusvyatskiy. Stochastic model-based minimization of weakly convex functions. *SIAM Journal on Optimization*, 29(1):207–239, 2019.

[11] A. L. Dontchev and R. T. Rockafellar. *Implicit Functions and Solution Mappings: A View From Variational Analysis.* Springer, New York, 2009.

[12] E. Laude, T. Wu, and D. Cremers. A nonconvex proximal splitting algorithm under Moreau-Yosida regularization. In *Artificial Intelligence and Statistics (AISTATS)*, 2018.

[13] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[14] C. Lescarret. Applications "prox" dans un espace de Banach. *Comptes Rendus de l'Académie des Sciences de Paris Série A*, 265:676–678, 1967.

[15] A. S. Lewis, D. R. Luke, and J. Malick. Local linear convergence for alternating and averaged nonconvex projections. *Foundations of Computational Mathematics*, 9(4):485–513, 2009.

[16] J.-J. Moreau. Fonctions convexes duales et points proximaux dans un espace Hilbertien. *Comptes Rendus de l'Académie des Sciences de Paris Série A*, 255:2897–2899, 1962.

[17] J.-J. Moreau. Proximité et dualité dans un espace hilbertien. *Bulletin de la Société Mathématique de France*, 93(2):273–299, 1965.

[18] P. Ochs. Local convergence of the heavy-ball method and iPiano for non-convex optimization. *Journal of Optimization Theory and Applications*, 177:153–180, 2018.

[19] R. A. Poliquin and R. T. Rockafellar. Prox-regular functions in variational analysis. *Trans. Amer. Math. Soc.*, 348:1805–1838, 1996.

[20] S. M. Robinson. Strongly regular generalized equations. *Mathematics of Operations Research*, 5:43–62, 1980.

[21] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, New Jersey, 1970.

[22] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis.* Springer, New York, 1998.

[23] I. Sutskever, J. Martens, G. Dahl, and G. Hinton. On the importance of initialization and momentum in deep learning. In *International Conference on Machine Learning (ICML)*, pages 1139–1147, 2013.

[24] M. Teboulle. Entropic proximal mappings with applications to nonlinear programming. *Mathematics of Operations Research*, 17(3):670–690, 1992.

[25] P. Tseng. Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of Optimization Theory and Applications*, 109(3):475–494, 2001.

[26] D. Wexler. Prox-mappings associated with a pair of Legendre conjugate functions. *Revue française d'automatique informatique recherche opérationnelle*, 7(R2):39–65, 1973.

[27] S. Zhang, A. Choromanska, and Y. LeCun. Deep learning with elastic averaging SGD. In *Advances in Neural Information Processing Systems (NIPS)*, pages 685–693, 2015.