From perspective maps to epigraphical projections

Michael P. Friedlander
Department of Computer Science/Department of Mathematics, University of British Columbia
2366 Main Mall Vancouver, BC, V6T 1Z4, Canada
michael@friedlander.io, https://friedlander.io

Ariel Goodwin
Department of Mathematics and Statistics, McGill University
805 Sherbrooke St West, Montréal, Québec, H3A 0B9, Canada
ariel.goodwin@mail.mcgill.ca, https://github.com/arielgoodwin

Tim Hoheisel
Department of Mathematics and Statistics, McGill University
805 Sherbrooke St West, Montréal, Québec, H3A 0B9, Canada
tim.hoheisel@mcgill.ca, https://www.math.mcgill.ca/hoheisel/

Dedicated to James V. Burke, our collaborator and friend, on the occasion of his 65th birthday

The projection onto the epigraph or a level set of a closed proper convex function can be achieved by finding a root of a scalar equation that involves the proximal operator as a function of the proximal parameter. This paper develops the variational analysis of this scalar equation. The approach is based on a study of the variational-analytic properties of general convex optimization problems that are (partial) infimal projections of the the sum of the function in question and the perspective map of a convex kernel. When the kernel is the Euclidean norm squared, the solution map corresponds to the proximal map, and thus the variational properties derived for the general case apply to the proximal case. Properties of the value function and the corresponding solution map—including local Lipschitz continuity, directional differentiability, and semismoothness—are derived. An SC^1 optimization framework for computing epigraphical and level-set projections is thus established. Numerical experiments on 1-norm projection illustrate the effectiveness of the approach as compared with specialized algorithms.

Key words: Proximal map, Moreau envelope, subdifferential, Fenchel conjugate, perspective map, epigraph, infimal projection, infimal convolution, set-valued map, coderivative, graphical derivative, semismoothness*, SC^1 optimization
MSC2000 subject classification: 52A4, 65K10, 90C25, 90C46

---

1. Introduction The Moreau proximal map of a closed proper convex function f that maps a finite-dimensional Euclidean space $\mathbb{E}_f$ to $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ is given by the minimizing set

$$P_\lambda f(x) = \underset{u\in\mathbb{E}_f}{\operatorname{argmin}} \left\{ f(u) + (1/2\lambda)\|x-u\|^2 \right\} \quad (\lambda > 0).$$

The proximal map is a central operation of algorithms for nonsmooth optimization, including first-order methods such as proximal gradient and operator splitting [3, 35]. Geometrically, the proximal map corresponds to the Euclidean projection $P_{\operatorname{epi} f}$ onto the epigraph $\operatorname{epi} f$; see Fig. 1. Indeed, for all positive λ and $x_\lambda := P_\lambda f(x)$,

$$\big(x_\lambda,\, f(x_\lambda)\big) = P_{\operatorname{epi} f}(x,\, f(x_\lambda) - \lambda). \tag{1}$$

Thus, the projection of an arbitrary point $(x,\alpha) \in \mathbb{E}_f \times \mathbb{R} \notin \operatorname{epi} f$ corresponds to the proximal map of the base point x using the parameter λ that is the unique positive root of the function

$$0 < \lambda \mapsto f(x_\lambda) - \lambda - \alpha. \tag{2}$$

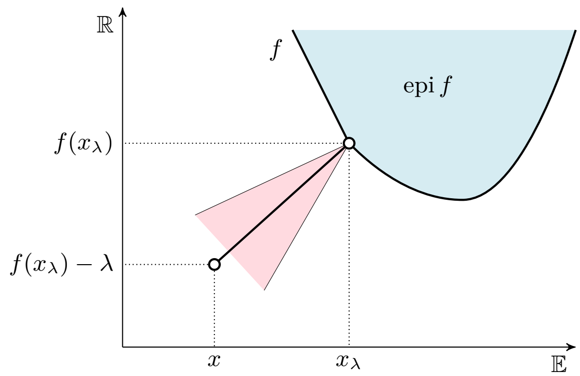


FIGURE 1. The proximal map $x_\lambda := P_\lambda f(x)$ corresponds to the projection of the pair $(x,\, f(x_\lambda)-\lambda))$ onto the epigraph of f; see (1).

This connection between epigraphical projection and the proximal map—described by Beck [4], Bauschke and Combettes [3, Section 29.5], Chierchia et al. [11, Proposition 1], and Meng et al. [31, 32]—is a defining feature of a class of *epigraphical first-order methods* for structured convex optimization over $\mathbb{E}_f$ that operate through a sequence of projections onto the epigraphs of the underlying functions. In effect, these methods operate on an equivalent optimization problem over $\mathbb{E}_f \times \mathbb{R}$ [11, 43, 44, 45].

This paper develops a general analysis that provides, among other things, the variational properties of the maps

$$(x,\, \lambda) \mapsto x_\lambda := P_\lambda f(x) \qquad \text{and} \qquad (x,\, \lambda) \mapsto f(x_\lambda),$$

defined on $\mathbb{E}_f \times \mathbb{R}$. This analysis and its supporting calculus allows us to determine the sensitivity of the epigraphical projection with respect to the simultaneous variation of the base point x and the scaling parameter λ. Although the resulting mathematical statements are key for our deeper understanding of epigraphical first-order methods, the overall analysis applies much more generally.

The approach we take is based on the variational analysis of the optimal value function

$$p_{L,\omega,f} : (x,\, \lambda) \in \mathbb{E}_x \times \mathbb{R} \mapsto \inf_{u \in \mathbb{E}_f} f(u) + \omega^\pi(L(u,x),\, \lambda) \tag{3}$$

and its corresponding solution map. Here, L is a linear map, and the perspective transform ω^π of a closed proper convex function ω is defined by $\operatorname{epi}\omega^\pi = \operatorname{cl}\mathbb{R}_+(\operatorname{epi}\omega \times \{1\})$. When the linear map L is defined as $(u,\, x) \mapsto x - u$, the value function (3) is the infimal convolution of the functions f and $\omega^\pi(\cdot,\, \lambda)$. For this reason, we refer to this value function as the *generalized convolution* of these two functions.

The convex calculus we establish in Section 3 for the analysis of the generalized convolution (3) provides a key tool for understanding several important cases. These include the variational properties of infimal convolution (Section 3.3); parametric constrained optimization (Section 3.4); the Moreau envelope of a convex function and the corresponding proximal map (Section 4); and epigraphical and level-set projections, including an SC^1 *optimization* [20, 36] method for numerically evaluating these projections (Section 7).

1.1. Contributions and related work The perspective map used in generalized convolution (3) first appears in Rockafellar [38, Corollary 13.5.1], without a particular name attached to it. More recently, Combettes [13], Combettes and Müller [14, 15], and Aravkin et al. [1], describe in detail the properties and applications of this map. Our systematic study of parametric optimization problems with perspective maps, outlined in Section 3, appears to be new.

1.1.1. Infimal convolution Section 3.3 establishes the variational properties of infimal convolution, which occurs when the map L is $(u,x) \mapsto x-u$. These results complement the functional smoothing framework described by Beck and Teboulle [5, Section 4.1] and Burke and Hoheisel [8, 9], wherein a smooth approximation to a function f is constructed through the infimal convolution with the perspective map of a smooth and strongly convex regularizer ω. Bougeard et al. [7] and Strömberg [42] provide early contributions to this topic. Theorem 3 describes the Lipschitzian properties of the corresponding optimal solution map—as a function of (x,λ). Corollary 3 establishes sufficient conditions for this solution map to be *semismooth** [23]. These conditions hold, for instance, when f is piecewise linear-quadratic. This analysis complements the study of the proximal case by Meng et al. [31, 32] and Milzarek [33].

1.1.2. Parametric constrained optimization A general form of parametric constrained optimization occurs when we specialize the convolution kernel ω in (3) to be the indicator function to a closed convex set. Section 3.4 focuses the variational analysis of the generalized convolution operation to obtain formulas for the sensitivity of the optimal value of parametric optimization problems with relaxed linear constraints. This analysis includes perturbations to the relaxation parameter and to the right-hand side.

1.1.3. Moreau envelope and proximal map In Section 4 we further focus our analysis of infimal convolution on the *proximal case*, which occurs when $\omega = \frac{1}{2}\|\cdot\|_2^2$. Here we develop the variational properties of the Moreau envelope and the associated proximal map as a function of the base point x and the proximal parameter λ, simultaneously. We also establish conditions under which the proximal map is semismooth*. Special attention is given to the limiting properties as $\lambda \downarrow 0$ (Propositions 4 and 5) and to continuity and smoothness properties of the proximal map (Corollaries 7 and 8 and Proposition 6). Milzarek's dissertation [33] includes a related analysis that generalizes the proximal parameter λ to a positive-definite matrix, but makes no statements regarding the limiting case where λ (or its matrix counterpart) vanishes, as we do in our general analysis. See also Attouch's seminal monograph [2].

1.1.4. Proximal value map In Section 5 we describe the main continuity properties of the *proximal value function*

$$0 < \lambda \mapsto f(P_\lambda f(\bar{x})), \tag{4}$$

where $\bar{x} \in \mathbb{E}$ is held fixed. Corollary 11 establishes its Lipschitzian properties and Corollary 12 characterizes it as the derivative of the map $\lambda \mapsto \lambda e_\lambda f(\bar{x})$ on $\mathbb{R}_{++}$. Proposition 7 describes sufficient conditions under which the proximal value function is semismooth.

1.1.5. Post compositions, and epigraphical and level-set projection We use our analysis of the proximal value function (4) to establish, via Proposition 8, novel variational formulas for the Moreau envelope and proximal map of *post-compositions*, i.e., functions of the form $g \circ \psi$, where the scalar function g is increasing and convex, and ψ is closed proper convex. As a consequence, Corollary 14 provides a refined version of the epigraphical projection conditions in (1), including analogous results for the projection onto the level set of f (Corollary 13). This analysis does not require the function to be finite-valued, and extends existing results [3, 4]. Importantly, Corollary 14 shows that the root of the aligning equation (2) coincides with the unique minimizer of a strongly convex scalar optimization problem. It follows from Proposition 7 that the objective for this problem is continuously differentiable with a locally Lipschitz derivative. We use this latter property to derive a novel SC^1 optimization method to find the root of the function (2) and its analog in the level-set case. Numerical experiments in Section 7.2 show that for projection onto the 1-norm unit ball, the resulting SC^1 method is competitive with two specialized state-of-the-art methods: CONDAT [16] and IBIS [30].

1.2. Notation Let $\Gamma_0(\mathbb{E})$ denote the set of functions $f:\mathbb{E}\to\overline{\mathbb{R}}$ that are proper closed convex, i.e., the epigraph $\operatorname{epi} f=\{(x,\alpha)\in\mathbb{E}\times\mathbb{R}\mid f(x)\leq\alpha\}$ contains no vertical lines and is closed convex. Its *level sets* are given by $\operatorname{lev}_\alpha f:=\{x\in\mathbb{E}\mid f(x)\leq\alpha\}$. The Fenchel conjugate of any function $f:\mathbb{E}\to\overline{\mathbb{R}}$ is $f^*(y)=\sup_{x\in\mathbb{E}}\{\langle y,x\rangle-f(x)\}$. The Jacobian of a differentiable map $F:\mathbb{R}^n\to\mathbb{R}^m$ at $x\in\mathbb{R}^n$ is denoted by $F'(x)$. We denote the Euclidean projection of $\bar{x}$ onto C by $P_C(\bar{x})$. Throughout, fractions such as $(1/(2\lambda))$ are abbreviated as $(1/2\lambda)$.

For a set $C\subset\mathbb{E}$, its *indicator function* is $\delta_C:\mathbb{E}\to\overline{\mathbb{R}}$ given by $\delta_C(x):=0$ if $x\in C$ and $\delta_C(x)=+\infty$ otherwise. The subdifferential of δ_C is the *normal cone* of C, i.e., $N_C(\bar{x}):=\partial\delta_C(\bar{x}):=\{v\in\mathbb{E}\mid\langle v,x-\bar{x}\rangle\leq 0\ (x\in C)\}$, which is empty if $\bar{x}\notin C$. The relative interior of C is the set $\operatorname{ri} C$ [38, Section 6], and the *horizon cone* is $C^\infty:=\{v\in\mathbb{E}\mid\exists\{\lambda_k\}\downarrow 0,\ \{x_k\in C\}:\lambda_k x_k\to v\}$. The *horizon function* of $f\in\Gamma_0(\mathbb{E})$ is the closed proper convex and positively homogeneous function $f^\infty:\mathbb{E}\to\overline{\mathbb{R}}$ defined via $\operatorname{epi} f^\infty=(\operatorname{epi} f)^\infty$.

Let $f_k:\mathbb{E}\to\overline{\mathbb{R}}$. Then we say that the sequence $\{f_k\}$ *epi-converges* to a function $f:\mathbb{E}\to\overline{\mathbb{R}}$ if

$$\forall x\in\mathbb{E}:\begin{cases}\forall\{x_k\}\to x:\liminf_{k\to\infty} f_k(x_k)\geq f(x),\\ \exists\{x_k\}\to x:\limsup_{k\to\infty} f_k(x_k)\leq f(x),\end{cases}$$

and we write $f_k\xrightarrow{e} f$. The sequence $\{f_k\}$ is said to converge *continuously* to f if

$$\lim_{k\to\infty} f_k(x_k)=f(x)\quad\forall x\in\mathbb{E}\text{ and }\{x_k\}\to x,$$

and we write $f_k\xrightarrow{c} f$. Furthermore, $\{f_k\}$ is said to converge *pointwise* to f if

$$\lim_{k\to\infty} f_k(x)=f(x)\quad\forall x\in\mathbb{E},$$

and we write $f_k\xrightarrow{p} f$. We extend these notions to families of functions $\{f_\lambda\}_{\{\lambda\downarrow 0\}}$ via

$$f_\lambda\xrightarrow{\xi} f\quad:\Longleftrightarrow\quad\forall\{\lambda_k\}\downarrow 0:\ f_{\lambda_k}\xrightarrow{\xi} f\quad(\xi\in\{p,e,c\}).$$

2. Properties of the perspective map The perspective map ω^π that appears in the generalized infimal convolution (3) provides a mechanism for controlling, through the parameter λ, the degree to which the functions f and ω are combined. Beck and Teboulle [5] and Burke and Hoheisel [8] promoted this technique for generating smooth approximations to nonsmooth functions.

We work with the following definition of the perspective map of ω, which appears in Rockafellar [38, Corollary 13.5.1]:

$$\omega^\pi:(z,\lambda)\in\mathbb{E}_\omega\times\mathbb{R}\mapsto\begin{cases}\lambda\omega\left(z/\lambda\right) & \text{if }\lambda>0,\\ \omega^\infty(z) & \text{if }\lambda=0,\\ +\infty & \text{if }\lambda<0.\end{cases}\tag{5}$$

For positive values of the parameter λ, the perspective map corresponds to *epi-multiplication*:

$$(\lambda\star\omega)(x):=\lambda\omega\left(x/\lambda\right).$$

The following result confirms the consistency of the perspective map (5) as the parameter λ decreases towards zero.

Lemma 1 **(Variational convergence of epi-multiplication)**. *Let $\phi\in\Gamma_0(\mathbb{E})$. Then as $\lambda\downarrow 0$, $(\lambda\star\phi)(x)\to\phi^\infty(x)$ for all $x\in\operatorname{dom}\phi$, and $\lambda\star\phi\xrightarrow{e}\phi^\infty$.*

Proof. The pointwise convergence of $(\lambda \star \phi)$ over $\operatorname{dom}\phi$ follows from [38, Corollary 8.5.2]. To prove epi-convergence, observe that, for all $\lambda > 0$ and $x \in \mathbb{E}$, $(\lambda \star \phi)(x) = \phi^\pi(x, \lambda)$. Hence,

$$\liminf_{\substack{x\to\bar{x}\\ \lambda\downarrow 0}} (\lambda \star \phi)(x) = \liminf_{\substack{x\to\bar{x}\\ \lambda\downarrow 0}} \phi^\pi(x,\lambda) \ge \phi^\pi(\bar{x},0) = \phi^\infty(\bar{x}) \quad \forall \bar{x}\in\mathbb{E},$$

where the inequality follows because ω^π is a support function [38, Corollary 13.5.1] and thus closed [25, Proposition 2.1.2].

Fix any sequence $\{\lambda_k\}\downarrow 0$ and take $\bar{x}\in\operatorname{dom}\phi$. Then $(\lambda_k\star\phi)(\bar{x})\to\phi^\infty(\bar{x})$. Hence, in particular, with $x_k := \bar{x}$ $(k\in\mathbb{N})$,

$$\limsup_{k\to\infty} (\lambda_k\star\phi)(x_k) \le \phi^\infty(\bar{x}) \tag{6}$$

for all $\bar{x}\in\operatorname{dom}\phi$. Now let $\bar{x}\notin\operatorname{dom}\phi$, take $\hat{x}\in\operatorname{dom}\phi$ and define $x_k := \lambda_k\hat{x} + (1-\lambda_k)\bar{x}\to\bar{x}$. Then

$$\phi^\infty(\bar{x}) = \sup_{t>0}\frac{\phi(\hat{x}+t\bar{x})-\phi(\hat{x})}{t} \ge \frac{\phi\left(\hat{x}+\left(\frac{1}{\lambda_k}-1\right)\bar{x}\right)-\phi(\hat{x})}{\frac{1}{\lambda_k}-1} = \lambda_k\cdot\frac{\phi\left(\hat{x}+\left(\frac{1}{\lambda_k}-1\right)\bar{x}\right)-\phi(\hat{x})}{1-\lambda_k}$$

for all $k\in\mathbb{N}$ sufficiently large. Hence for such $k\in\mathbb{N}$,

$$(\lambda_k\star\phi)(x_k) = \lambda_k\phi\left(\frac{\lambda_k\hat{x}+(1-\lambda_k)\bar{x}}{\lambda_k}\right) \le (1-\lambda_k)\phi^\infty(\bar{x}) + \lambda_k\phi(\hat{x}).$$

Take the limit superior to obtain (6) here. This establishes epi-convergence. □

The following result summarizes key properties of the perspective map. It also provides a support-function representation, which means that it can be written as the support function $\sigma_{\mathcal{D}}(y) \equiv \delta^*_{\mathcal{D}}(y) = \sup_{x\in\mathcal{D}}\langle x,\, y\rangle$ for some set $\mathcal{D}$.

Proposition 1 **(Properties of perspective map)**. *For $\omega\in\Gamma_0(\mathbb{E}_\omega)$, the following hold:*

(a) $\omega^\pi(z,\lambda) = \sigma_{\operatorname{epi}\omega^*}(z,-\lambda)$, *hence* $\omega^\pi\in\Gamma_0(\mathbb{E}_\omega\times\mathbb{R})$ *is sublinear with* $\operatorname{dom}\omega^\pi = \mathbb{R}_+(\operatorname{dom}\omega\times\{1\})$*;*

(b) $(\omega^\pi)^*(y,\beta) = \delta_{\operatorname{epi}\omega^*}(y,-\beta)$*;*

(c) *for all* $(z,\lambda)\in\operatorname{dom}\omega^\pi$,

$$\partial\omega^\pi(z,\lambda) = \begin{cases}\{(y,-\beta)\mid y\in\partial\omega(z/\lambda),\ \beta=\omega^*(y)\} & \text{if } \lambda>0,\\ \{(y,-\beta)\mid y\in\partial\omega^\infty(z),\ (y,\beta)\in\operatorname{epi}\omega^*\} & \text{if } \lambda=0.\end{cases} \tag{7}$$

Proof. For Parts (a) and (b) see [38, Corollary 13.5.1]. Part (c) follows from [13, Proposition 2.3] or [1, Lemma 3.8]. □

The expression for the subdifferential (7), evaluated at the origin, reduces to $\partial\omega^\pi(0,0) = \{(y,-\beta)\in\operatorname{epi}\omega^*\}$, which is just the epigraph of ω^* under the reflection $(z,\lambda)\mapsto(z,-\lambda)$. This follows because the subdifferential formula $\partial\omega^\infty(0) = \partial\sigma_{\operatorname{dom}\omega^*}(0) = \operatorname{dom}\omega^*$; cf. [39, Corollary 8.25]. Combettes [13, Corollary 2.5] provides a simplified characterization of Proposition 1 under the additional assumption that ω is supercoercive [3, Definition 11.11].

3. Partial infimal projection with perspective maps

Our main objective in this section is to deduce the variational properties of the generalized infimal convolution $p_{L,\omega,f}$ defined by (3). Throughout this section, we make the assumptions that L is a linear map from $\mathbb{E}_f\times\mathbb{E}_x$ to $\mathbb{E}_\omega$ for Euclidean spaces $\mathbb{E}_i$, $i\in\{f,x,\omega\}$, that $f\in\Gamma_0(\mathbb{E}_f)$ and $\omega\in\Gamma_0(\mathbb{E}_\omega)$, and that $\operatorname{range}L\subseteq\mathbb{R}_+\operatorname{dom}\omega$. Under these standing assumptions, it follows from Theorem 1 below that $p_{L,\omega,f}$ is convex.

3.1. Infimal projection We lead with a general result on infimal projections.

THEOREM 1 **(Conjugate and subdifferentials of infimal projection).** *For a function* $\psi \in \Gamma_0(\mathbb{E}_1 \times \mathbb{E}_2)$*, the infimal projection*

$$p : x \in \mathbb{E}_1 \mapsto \inf_u \psi(x,u) \tag{8}$$

is convex and

(a) $p^* = \psi^*(\cdot,0)$*, which is closed and convex;*
(b) $\partial p(x) = \{v \mid (v,0) \in \partial\psi(x,\bar{u})\}$ *for all* $\bar{u} \in \operatorname{argmin} \psi(x,\cdot)$*;*
(c) $p^* \in \Gamma_0(\mathbb{E}_1)$ *if and only if* $\operatorname{dom} \psi^*(\cdot,0) \neq \emptyset$*;*
(d) $p \in \Gamma_0(\mathbb{E}_1)$ *if* $\operatorname{dom} \psi^*(\cdot,0) \neq \emptyset$*, and hence the infimum in (8) is attained when finite.*

Proof. For convexity of p and Parts (a,b,d,e), see, e.g., [26, Theorem 3.101]. Part (c) follows from Part (b) via Rockafellar [38, Theorem 23.5]. □

3.2. Generalized infimal convolution The following auxiliary result is used in this section to derive conjugate and a subdifferential formulas for the value function $p_{L,\omega,f}$.

LEMMA 2 **(Domain and conjugate of linear-perspective composition).** *The function*

$$\eta : (u,x,\lambda) \in \mathbb{E}_f \times \mathbb{E}_x \times \mathbb{R} \mapsto \omega^\pi(L(u,x),\lambda)$$

is closed proper convex, i.e., $\eta \in \Gamma_0(\mathbb{E}_f \times \mathbb{E}_x \times \mathbb{R})$*. The nonempty domain and its (possibly empty) relative interior are given by*

$$\begin{aligned} \operatorname{dom}\eta \ &= \{(u,x,\lambda) \mid \lambda \geq 0,\ L(u,x) \in \lambda \cdot \operatorname{dom}\omega\}, \\ \operatorname{ri}(\operatorname{dom}\eta) &= \{(u,x,\lambda) \mid \lambda > 0,\ L(u,x) \in \lambda \cdot \operatorname{ri}(\operatorname{dom}\omega)\}. \end{aligned}$$

If $\operatorname{ri}(\operatorname{dom}\eta)$ *is nonempty, then* η^* *is the indicator to the set*

$$C = \{(w,z,\mu) \mid \exists y \mid (y,-\mu) \in \operatorname{epi}\omega^*,\ L^*(y) = (w,z)\}. \tag{9}$$

Proof. Proposition 1(a) asserts that $\eta \in \Gamma_0(\mathbb{E}_f \times \mathbb{E}_x \times \mathbb{R})$, and also yields the expression for its domain. Now assume that $\operatorname{ri}(\operatorname{dom}\eta)$ is nonempty, and that there exists an element (u,x) such that $L(u,x) \in \lambda \cdot \operatorname{ri}(\operatorname{dom}\omega)$ for some $\lambda > 0$. Define the linear map $\tilde{L} : (u,x,\lambda) \mapsto (L(u,x),\lambda)$. Then,

$$\begin{aligned} \emptyset &\neq \{(u,x,t) \mid t > 0,\ L(u,x) \in t \cdot \operatorname{ri}(\operatorname{dom}\omega)\} \\ &= \{(u,x,\lambda) \mid \exists t > 0 :\ L(u,x) \in t \cdot \operatorname{ri}(\operatorname{dom}\omega),\ \lambda = t\} \\ &= \tilde{L}^{-1}\mathbb{R}_{++}(\operatorname{dom}\omega \times \{1\}) \\ &\overset{\text{(i)}}{=} \tilde{L}^{-1}\operatorname{ri}(\mathbb{R}_+(\operatorname{dom}\omega \times \{1\})) \\ &\overset{\text{(ii)}}{=} \operatorname{ri}(\tilde{L}^{-1}\mathbb{R}_+(\operatorname{dom}\omega \times \{1\})) \\ &= \operatorname{ri}(\tilde{L}^{-1}\operatorname{dom}\omega^\pi) = \operatorname{ri}(\operatorname{dom}\eta), \end{aligned}$$

where (i) uses [38, Corollary 6.8.1] and (ii) uses [38, Theorem 6.7] and the fact that $L^{-1}\operatorname{ri}(\mathbb{R}_+(\operatorname{dom}\omega \times \{1\})) \neq \emptyset$.

To derive the formula for η^*, observe that by our reasoning above $\tilde{L}^{-1}\operatorname{ri}(\operatorname{dom}\omega^\pi) = \operatorname{ri}(\operatorname{dom}\eta) \neq \emptyset$. Hence, by [38, Theorem 16.3] and Proposition 1(b),

$$\begin{aligned} \eta^*(w,z,\mu) &= (\omega^\pi \circ \tilde{L})^*(w,z,\mu) \\ &= \inf_{(u,\alpha)} \left\{ (\omega^\pi)^*(u,\alpha) \,\middle|\, \tilde{L}^*(u,\alpha) = (w,z,\mu) \right\} \\ &= \inf_u \{(\omega^\pi)^*(u,\mu) \mid L^*(u) = (w,z)\} \\ &= \inf_u \{\delta_{\operatorname{epi}\omega^*}(u,-\mu) \mid L^*(u) = (w,z)\} \\ &= \delta_C(w,z,\mu), \end{aligned}$$

which establishes (9) □

We can now deduce the subdifferential and conjugate of the generalized convolution (3).

Theorem 2 **(Conjugate and subdifferential of the generalized convolution)**. *Under the assumptions of Lemma 2, suppose in addition that*

$$\exists (u,x)\in \mathrm{ri}\,(\mathrm{dom}\, f)\times \mathbb{E}_x:\ L(u,x)\in \mathbb{R}_{++}\mathrm{ri}\,(\mathrm{dom}\,\omega). \tag{10}$$

Then the following hold for the convex function $p_{L,\omega,f}$ defined in (3).

(a) $p^*_{L,\omega,f}(y,\mu)=\inf_w\{f^*(w)\mid \exists a:(a,-\mu)\in \mathrm{epi}\,\omega^*,\ L^*(a)=(-w,y)\}$ *and the infimum is attained when finite.*

(b) *For all $(x,\lambda)\in \mathrm{dom}\, p_{L,\omega,f}$ and all $\bar u\in \mathrm{argmin}_{u\in\mathbb{E}_f}\{f(u)+\omega^\pi(L(u,x),\lambda)\}$,*

$$\partial p_{L,\omega,f}(x,\lambda)=\begin{cases}\{(v,-\omega^*(y))\mid y\in\partial\omega\,(L(\bar u,x)/\lambda),\ (0,v)\in\mathcal{D}(\bar u,y)\} & \text{if } \lambda>0,\\ \{(v,-\beta)\mid y\in\partial\omega^\infty(L(\bar u,x)),\ (0,v)\in\mathcal{D}(\bar u,y),\ (y,\beta)\in \mathrm{epi}\,\omega^*\} & \text{if } \lambda=0,\end{cases}$$

where $\mathcal{D}(u,y):=\partial f(u)\times\{0\}+L^*(y)$.

(c) $p^*_{L,\omega,f}\in\Gamma_0(\mathbb{E}_x\times\mathbb{R})$ *if and only if there exist $w\in\mathrm{dom}\, f^*, a\in\mathrm{dom}\,\omega^*, (y,\mu)\in\mathbb{E}_x\times\mathbb{R}$ such that $(a,-\mu)\in\mathrm{epi}\,\omega^*$ and $L^*(a)=(-w,y)$. In this case, $p_{L,\omega,f}\in\Gamma_0(\mathbb{E}_x\times\mathbb{R})$ and the infimum is attained when finite.*

Proof. Set $p=p_{L,\omega,f}$. Part (a). Observe that $p(x,\lambda)=\inf_u\psi(u,x,\lambda)$ for $\psi=\phi+\eta$ with $\phi(u,x,\lambda)=f(u)$ (and η as in Lemma 2). We hence compute

$$\begin{aligned} p^*(y,\mu) &= \psi^*(0,y,\mu)\\ &= (\phi+\eta)^*(0,y,\mu)\\ &= \inf_{(w,z,\delta)}\phi^*(w,z,\delta)+\eta^*(-w,y-z,\mu-\delta)\\ &= \inf_w f^*(w)+\delta_C(-w,y,\mu)\\ &= \inf_w\{f^*(w)\mid\exists a:(a,-\mu)\in\mathrm{epi}\,\omega^*,\ L^*(a)=(-w,y)\}.\end{aligned}$$

Here the first identity uses Theorem 1. The second is clear from our definitions above. The third relies on [38, Theorem 16.4] and the fact that assumption (10) is, in view of Lemma 2(b) and the fact that $\mathrm{ri}\,(\mathrm{dom}\,\phi)=\mathrm{ri}\,(\mathrm{dom}\, f)\times\mathbb{E}_x\times\mathbb{R}$, equivalent to the condition $\mathrm{ri}\,(\mathrm{dom}\,\eta)\cap\mathrm{ri}\,(\mathrm{dom}\,\phi)\neq\emptyset$. The fifth uses the fact that $\phi^*(v,y,\mu)=f^*(v)+\delta_{\{0\}}(y,\mu)$ and Lemma 2 b). The last identity is simply the definition of the set C in said proposition.

Part (b). By (10) we can apply [38, Theorems 23.8-23.9] to find

$$\begin{aligned}\partial\psi(u,x,\lambda) &= \partial f(u)\times\{0\}\times\{0\}+\tilde L^*\partial\omega^\pi(\tilde L(u,x,\lambda))\\ &= \partial f(u)\times\{0\}\times\{0\}+(L^*\times\mathrm{id})\partial\omega^\pi(L(u,x),\lambda).\end{aligned}$$

Apply Proposition 1(c) and combine with Theorem 1 to obtain the desired result.

Part(c) follows from Theorem 1(d). □

3.3. Infimal convolution We now consider the value function

$$p_{\omega,f}:(x,\lambda)\in\mathbb{E}\times\mathbb{R}\mapsto\inf_{u\in\mathbb{E}} f(u)+\omega^\pi(x-u,\lambda), \tag{11}$$

which corresponds to the standard infimal convolution between f and ω^π. This is a special case of (3) where $L(u,x)=x-u$ and $\mathbb{E}_i=\mathbb{E}$ with $i=f,x,w$. The following result specializes Theorem 1.

Corollary 1 **(Conjugate and subdifferential of infimal convolution)**. *For the function $p_{\omega,f}$ given by* (11)*, assume that $f,\omega\in\Gamma_0(\mathbb{E})$ and*

$$\exists (u,x)\in \operatorname{ri}(\operatorname{dom} f)\times\mathbb{E}:\; x-u\in\mathbb{R}_{++}\operatorname{ri}(\operatorname{dom}\omega). \tag{12}$$

Then the following hold.
(a) $p^*_{\omega,f}(y,\mu)=f^*(y)+\delta_{\operatorname{epi}\omega^*}(y,-\mu)$.
(b) *For all $(x,\lambda)\in\operatorname{dom} p_{\omega,f}$ and all $\bar u\in\operatorname{argmin}_{u\in\mathbb{E}}\{f(u)+\omega^\pi(x-u,\lambda)\}$ we have*

$$\partial p_{\omega,f}(x,\lambda)=\begin{cases}\left\{(y,-\beta)\;\middle|\; y\in\partial f(\bar u)\cap\partial\omega\left(\frac{x-\bar u}{\lambda}\right),\;\beta=\omega^*(y)\right\} & \text{if } \lambda>0,\\ \left\{(y,-\beta)\mid y\in\partial f(\bar u)\cap\partial w^\infty(x-\bar u),\;(y,\beta)\in\operatorname{epi}\omega^*\right\} & \text{if } \lambda=0.\end{cases}$$

(c) $p^*_{\omega,f}\in\Gamma_0(\mathbb{E})$ *if and only if* $\operatorname{dom} p^*_{\omega,f}=(\operatorname{dom} f^*\times\mathbb{E})\cap\operatorname{epi}\omega^*\neq\emptyset$*. In this case, $p_{\omega,f}\in\Gamma_0(\mathbb{E})$ also, and the infimum is attained when finite.*

Proof. Use Theorem 2(a)–(c) and observe that $L^*(a)=(-a,a)$. □

3.3.1. Infimal convolution solution map Thus far, our analysis has focused exclusively on the variational properties of the optimal value function (3) and its specializations. We now turn our attention to the optimal solution map

$$P_{\omega,f}:(x,\lambda)\in\mathbb{E}_x\times\mathbb{R}\mapsto \operatorname*{argmin}_{u\in\mathbb{E}_f} f(u)+\omega^\pi(x-u,\lambda) \tag{13}$$

for the infimal convolution defined by (11). In this section we describe the variational-analytic properties of the solution map, including (Lipschitz) continuity and (directional) smoothness. To this end, we introduce required technical machinery from variational analysis [34, 39].

Let $S:\mathbb{E}_1\rightrightarrows\mathbb{E}_2$ be a set-valued map between spaces $\mathbb{E}_1$ and $\mathbb{E}_2$. The domain and graph of S, respectively, are the sets $\operatorname{dom} S:=\{x\mid S(x)\neq\emptyset\}$ and $\operatorname{gph} S:=\{(x,u)\in\mathbb{E}_1\times\mathbb{E}_2\mid u\in S(x)\}$. The *outer limit* of S at $\bar x$ is

$$\operatorname*{Lim\,sup}_{x\to\bar x} S(x):=\{y\in\mathbb{E}_x\mid \exists\{x_k\}\to\bar x,\{y_k\in S(x_k)\}\to y\}.$$

Now let $A\subset\mathbb{E}$. The *tangent cone* of A at $\bar x\in A$ is $T_A(\bar x):=\operatorname{Lim\,sup}_{t\downarrow 0}(A-\bar x)/t$. The *regular normal cone* of A at $\bar x\in A$ is the polar of the tangent cone, i.e., $\hat N_A(\bar x):=\{v\mid\langle v,y\rangle\le 0\;\forall y\in T_A(\bar x)\}$. The *limiting normal cone* of A at $\bar x\in A$ is $N_A(\bar x):=\operatorname{Lim\,sup}_{x\to\bar x}\hat N_A(x)$. The *coderivative* of S at $(\bar x,\bar y)\in\operatorname{gph} S$ is the map $D^*S(\bar x\mid\bar y):\mathbb{E}_2\rightrightarrows\mathbb{E}_1$ defined via

$$v\in D^*S(\bar x\mid\bar y)(y)\quad\Longleftrightarrow\quad (v,-y)\in N_{\operatorname{gph} S}(\bar x,\bar y).$$

The *graphical derivative* of S at $(\bar x,\bar y)$ is the map $DS(\bar x\mid\bar y):\mathbb{E}_f\rightrightarrows\mathbb{E}_x$ given by

$$v\in DS(\bar x\mid\bar y)(u)\quad\Longleftrightarrow\quad (u,v)\in T_{\operatorname{gph} S}(\bar x,\bar y),$$

or, equivalently, $DS(\bar x\mid\bar y)(u)=DS(\bar x|\bar y)(u)=\operatorname*{Lim\,sup}_{t\downarrow 0,\; u'\to u}\frac{S(\bar x+tu')-\bar y}{t}$ [39, Eq. 8(14)]. The *strict graphical derivative* of S at $(\bar x,\bar y)$ is $D_*S(\bar x\mid\bar y):\mathbb{E}_f\rightrightarrows\mathbb{E}_x$ given by

$$D_*S(\bar x\mid\bar y)(w)=\left\{z\;\middle|\;\exists\left\{\begin{array}{l}\{t_k\}\downarrow 0,\;\{w_k\}\to w,\;\{z_k\}\to z,\\ \{(x_k,y_k)\in\operatorname{gph} S\}\to(\bar x,\bar y)\end{array}\right\}: z_k\in\frac{S(x_k+t_kw_k)-y_k}{t_k}\right\}.$$

We adopt the convention to set $D^*S(\bar{x}) := D^*S(\bar{x} \mid \bar{u})$ if $S(\bar{x})$ is a singleton, and proceed analogously for the graphical derivatives.

The above generalized derivatives possess the following definiteness properties when applied to a maximally monotone operator $T : \mathbb{E} \rightrightarrows \mathbb{E}$, which (by definition) satisfies the inequality

$$\langle v - w,\, x - y\rangle \geq 0 \quad \forall (v,w) \in T(x) \times T(y),$$

and there is no enlargement of $\operatorname{gph} T$ without destroying this inequality. Our conclusion relies on *Minty parameterization*.

LEMMA 3. *Let $T : \mathbb{E} \rightrightarrows \mathbb{E}$ be maximally monotone and let $(\bar{y}, \bar{u}) \in \operatorname{gph} T$. Then the pair $(w, z) \in \mathbb{E} \times \mathbb{E}$ satisfies $\langle w,\, z\rangle \geq 0$ if one of the following conditions hold:*
(a) $w \in D^*T(\bar{y} \mid \bar{u})(z)$;
(b) $z \in D_*T(\bar{y} \mid \bar{u})(w)$;
(c) $z \in DT(\bar{y} \mid \bar{u})(w)$.

Proof. Part (a). See [34, Theorem 5.6].

Part (b). For $z \in D_*T(\bar{y} \mid \bar{u})(w)$ there exist $\{z_k\} \to z, \{t_k \downarrow 0\}, \{(y_k, u_k) \in \operatorname{gph} T\} \to (\bar{y}, \bar{u})$, and $\{w_k\} \to w$ such that

$$t_k z_k \in T(y_k + t_k w_k) - u_k \quad \forall k \in \mathbb{N}. \tag{14}$$

Now let $\lambda > 0$ and set $J_{\lambda T} := (\lambda T + \mathrm{id}\,)^{-1}$. By Minty parameterization [3, Remark 23.23], there exists $\{x_k\}$ such that $(y_k, u_k) = (J_{\lambda T}(x_k),\ (x_k - J_\lambda(x_k)/\lambda)$ for all $k \in \mathbb{N}$. Combining this with (14) yields $x_k + t_k(\lambda z_k + w_k) \in (\lambda T + \mathrm{id}\,)(y_k + t_k w_k)$. Thus, as $y_k = J_{\lambda T}(x_k)$, we have $t_k w_k = J_{\lambda T}(x_k + t_k(\lambda z_k + w_k) - J_{\lambda T}(x_k)$ $(k \in \mathbb{N})$. Because $J_{\lambda T}$ is firmly nonexpansive [3, Proposition 23.8] and hence 1-Lipschitz, it follows that $\|w_k\| \leq \|\lambda z_k + w_k\|$ $(k \in \mathbb{N})$, hence $\|w\| \leq \|\lambda z + w\|$. We infer that $-(\lambda/2)\|z\|^2 \leq \langle z,\, w\rangle$. Since $\lambda > 0$ was arbitrary, letting $\lambda \downarrow 0$ gives the desired inequality.

Part (c). Follows from Part (b) and the fact that $DS(\bar{x} \mid \bar{u})(w) \subset D_*S(\bar{x} \mid \bar{u})(w)$ for all $w \in \mathbb{E}_f$. □

We record another auxiliary result. Here we call $S : \mathbb{E}_f \rightrightarrows \mathbb{E}_x$ *proto-differentiable* at $(\bar{x}, \bar{u}) \in \operatorname{gph} S$ if for any $\bar{z} \in DS(\bar{x} \mid \bar{u})(\bar{w})$ and any $\{t_k\} \downarrow 0$ there exist $\{w_k\} \to \bar{w}$ and $\{z_k\} \to \bar{z}$ such that $z_k \in (S(\bar{x} + t_k w_k) - \bar{u})/t_k$ for all $k \in \mathbb{N}$.

LEMMA 4. *Let $S : \mathbb{E}_1 \rightrightarrows \mathbb{E}_2$ be given by $S = F + T$, where F is smooth and T is proto-differentiable at $(\bar{x}, \bar{u} - F(\bar{x}))$. Then S is proto-differentiable at $(\bar{x}, \bar{u})$.*

Proof. Let $z \in DS(\bar{x} \mid \bar{u})(w)$ and $\{t_k\} \downarrow 0$. Then $z - F'(\bar{x})w \in DT(\bar{x} \mid \bar{u} - F(\bar{x}))(w)$, cf. [39, Exercise 10.43]. By assumption on T, there exist $\tilde{z}_k \to z - F'(\bar{x})w$ and $w_k \to w$ such that $\tilde{z}_k \in [T(\bar{x} + t_k w_k) - (\bar{u} - F(\bar{x}))]/t_k$, i.e., $\tilde{z}_k + [F(\bar{x} + t_k w_k) - F(\bar{x})]/t_k \in [S(\bar{x} + t_k w_k) - \bar{u}]/t_k$ for all $k \in \mathbb{N}$. Therefore, $z_k := \tilde{z}_k + [F(\bar{x} + t_k w_k) - F(\bar{x})]/t_k \to z$ and $z_k \in [S(\bar{x} + t_k w_k) - \bar{u}]/t_k$ for all $k \in \mathbb{N}$ which shows the proto-differentiability of S at $(\bar{x}, \bar{u})$. □

The next and main result in this subsection is based on the implicit mapping framework described by Rockafellar and Wets [39, Theorem 9.56] together with Lemma 3.

THEOREM 3 **(Variational properties of the solution map)**. *Let $f \in \Gamma_0(\mathbb{E})$ and let $\omega : \mathbb{E} \to \mathbb{R}$ be strictly convex, level-bounded and twice continuously differentiable. Let $\bar{x} \in \mathbb{E}$ and $\bar{\lambda} > 0$, set $\bar{y} := P_{\omega,f}(\bar{x}, \bar{\lambda})$ and $\bar{V} := \nabla^2\omega\left(\frac{\bar{x} - \bar{y}}{\bar{\lambda}}\right)$. Then for the solution map $P_{\omega,f}$ from (13) the following hold:*
(a) *We have* $\operatorname{dom} P_{\omega,f} \subset \mathbb{E} \times \mathbb{R}_+$ *and $P_{\omega,f}(\cdot, \lambda)$ is single-valued for all $\lambda > 0$.*
(b) *If $\bar{V}$ is positive definite, then $P_{\omega,f}$ is locally Lipschitz at $(\bar{x}, \bar{\lambda})$.*
(c) *If $\bar{V}$ is positive definite and ∂f is* proto-differentiable *at $\left(\bar{y}, \nabla\omega\left(\frac{\bar{x} - \bar{y}}{\bar{\lambda}}\right)\right)$, then $P_{\omega,f}$ is is directionally differentiable*[1] *at $(\bar{x}, \bar{\lambda})$. Concretely, for all $(d, \Delta) \in \mathbb{E} \times \mathbb{R}$, we have*

$$P'_{\omega,f}((\bar{x}, \bar{\lambda}); (d, \Delta)) = \left[\bar{\lambda} D(\partial f)\left(\bar{y} \;\middle|\; \nabla\omega\left(\frac{\bar{x} - \bar{y}}{\bar{\lambda}}\right)\right) + \bar{V}\right]^{-1} \left(\bar{V} d - \frac{\Delta}{\bar{\lambda}} \bar{V}(\bar{x} - \bar{y})\right).$$

[1] In fact, *semidifferentiable* at $(\bar{x}, \bar{\lambda})$ in the sense of [39, p. 332].

Proof. Set $P := P_{\omega,f}$. Part (a). For $\lambda > 0$ and $x \in \mathbb{E}$, the function $u \mapsto f(y) + \omega^\pi(x-y,\lambda)$ is lsc, proper, strictly convex and level-bounded, and therefore attains a unique minimum.

Part (b). Without loss of generality, let $\mathbb{E} = \mathbb{R}^n$, and observe that, for $\lambda > 0$, we have $P(x,\lambda) = \{y \mid 0 \in S(x,\lambda,y)\}$, where $S(x,\lambda,u) := \partial f(y) - \nabla\omega\left(\frac{x-u}{\lambda}\right)$ $(\lambda > 0)$. Use [39, Exercise 10.43] to deduce

$$D^*S(\bar{x},\bar{\lambda},\bar{y}\mid 0)(y) = \left[-\frac{1}{\bar{\lambda}}\bar{V}y, \frac{(\bar{x}-\bar{y})^T}{\bar{\lambda}^2}\bar{V}y, \frac{1}{\bar{\lambda}}\bar{V}y\right] + \{0\}\times\{0\}\times D^*(\partial f)\left(\bar{y}\;\middle|\;\nabla\omega\left(\frac{\bar{x}-\bar{y}}{\bar{\lambda}}\right)\right)(y).$$

Hence, $(r,\gamma,0) \in D^*S(\bar{x},\bar{\lambda},\bar{y}\mid 0)(y)$ if and only if

$$r = -\frac{1}{\bar{\lambda}}\bar{V}y, \quad \gamma = \frac{(\bar{x}-\bar{y})^T}{\bar{\lambda}^2}\bar{V}y, \quad -\frac{1}{\bar{\lambda}}\bar{V}y \in D^*(\partial f)\left(\bar{y}\middle|\nabla\omega\left(\frac{\bar{x}-\bar{y}}{\bar{\lambda}}\right)\right)(y).$$

Invoke Lemma 3(a) and use $\bar{V} \succ 0$ to deduce $y = 0$, hence $r = 0$ and $\gamma = 0$. Therefore, by [39, Theorem 9.56 (a)], we see that P has the *Aubin property* at $(\bar{x},\bar{\lambda})$ for $\bar{y}$, and since P is single-valued, it is locally Lipschitz at $(\bar{x},\bar{\lambda})$.

Part (c). With the definitions from Part (b), recall that the implication

$$(r,\gamma,0) \in D^*S(\bar{x},\bar{\lambda},\bar{y}\mid 0)(y) \quad \Rightarrow \quad (r,\gamma) = 0,\ y = 0$$

was proved. Now let

$$0 \in D_*S(\bar{x},\bar{\lambda},\bar{y}\mid 0)\begin{pmatrix}0\\0\\w\end{pmatrix} = \frac{1}{\bar{\lambda}}\bar{V}w + D_*(\partial f)\left(\bar{y}\mid\nabla\omega\left(\frac{\bar{x}-\bar{y}}{\bar{\lambda}}\right)\right)(w),$$

see [39, Exercise 10.43], i.e.,

$$-\frac{1}{\bar{\lambda}}\bar{V}w \in D_*(\partial f)\left(\bar{y}\mid\nabla\omega\left(\frac{\bar{x}-\bar{y}}{\bar{\lambda}}\right)\right)(w).$$

By Lemma 3(b), we find that $w = 0$. Since ∂f is assumed to be proto-differentiable at $\left(\bar{y},\nabla\omega\left(\frac{\bar{x}-\bar{y}}{\bar{\lambda}}\right)\right)$, Lemma 4 yields that S is proto-differentiable at $((\bar{x},\bar{\lambda},\bar{y}),0)$. We can now apply [39, Theorem 9.56(c)] to obtain the desired result. □

Remark 1 (Proto-differentiability of ∂f from full amenability). Let $f \in \Gamma_0(\mathbb{E})$ and $\bar{x} \in \operatorname{dom} f$. By [39, Corollary 13.41], there exists a neighborhood V of $\bar{x}$ such that ∂f is proto-differentiable at $x \in V \cap \operatorname{dom} f$ for any $v \in \partial f(x)$ if f is *fully amenable* at $\bar{x}$ in the sense that (on a neighborhood of $\bar{x}$) $f = g \circ F$ with $g \in \Gamma_0(\mathbb{E}_x)$ piecewise linear-quadratic and $F \in C^2(\mathbb{E}_f,\mathbb{E}_x)$ such that

$$\ker F'(\bar{x})^* \cap N_{\mathrm{cl}\,(\operatorname{dom} g)}(F(\bar{x})) = \{0\}.$$

This comprises the following special cases:
- $f(x) = \max_{i=1}^m f_i(x)$ with $f_i \in \Gamma_0(\mathbb{E}) \cap C^2$;
- f is (convex and) piecewise linear quadratic;
- f is (convex and) twice continuously differentiable.

Since a strongly convex function is both strictly convex and level-bounded (in fact supercoercive) and has positive definite Hessian everywhere, and since we have $D(\partial f) = \nabla^2 f$ wherever f is twice continuously differentiable, we immediately obtain the following result which, of course, can also be derived directly from the implicit function theorem.

Corollary 2 **(Differentiability of the solution map)**. *Let $(\bar{x},\bar{\lambda}) \in \mathbb{E}\times\mathbb{R}_{++}$ such that $f \in \Gamma_0(\mathbb{E})$ is twice continuously differentiable around $P_{\omega,f}(\bar{x},\bar{\lambda})$, and let $\omega \in \Gamma_0(\mathbb{E})$ be strongly convex and twice continuously differentiable. Then $P_{\omega,f}$ from (13) is continuously differentiable around $(\bar{x},\bar{\lambda})$. Concretely, for all (x,λ) sufficiently close to $(\bar{x},\bar{\lambda})$ and for all $(d,\Delta) \in \mathbb{E}\times\mathbb{R}$, we have*

$$P'_{\omega,f}(x,\lambda)(d,\Delta) = \left(\lambda\nabla^2 f(y) + V\right)^{-1}\left[Vd - \Delta\cdot V\left(\frac{x-y}{\lambda}\right)\right],$$

where $y := P_{\omega,f}(x,\lambda)$ and $V := \nabla^2\omega\left(\frac{x-y}{\lambda}\right)$.

3.3.2. Semismoothness* We now refine our study of smoothness properties of the solution map $P_{\omega,f}$. We base our analysis on the notion of *semismoothness** recently established by Gfrerer and Outrata [23], which, in turn, relies on the notion of the *directional normal cone* introduced by Ginchev and Mordukohovich [24] and further advanced by Gfrerer et al. [6, 21, 22].

For $\bar{x} \in A \subset \mathbb{E}$, the directional normal cone in the direction $\bar{u} \in \mathbb{E}$ is given by

$$N(\bar{x};\bar{u}) = \operatorname*{Lim\,sup}_{u\to\bar{u},\, t\downarrow 0} \hat{N}_A(\bar{x}+tu).$$

Note that $N(\bar{x};\bar{u}) = \emptyset$ if $\bar{u} \notin T_A(\bar{x})$ and that $N(\bar{x};\bar{u}) \subset N_A(\bar{x})$ for all $u \in \mathbb{E}$. Given a set-valued map $S : \mathbb{E}_f \rightrightarrows \mathbb{E}_x$, based on the directional normal cone, we define the *directional coderivative* [21] $D^*S((\bar{x},\bar{u});(u,v)) : \mathbb{E}_x \rightrightarrows \mathbb{E}_f$ of S at $(\bar{x},\bar{y}) \in \operatorname{gph} S$ in the direction (u,v) via

$$\operatorname{gph} D^*S((\bar{x},\bar{u});(u,v))(v^*) = \left\{ u^* \in \mathbb{E}_f \mid (u^*,-v^*) \in N_{\operatorname{gph} S}((\bar{x},\bar{y});(u,v)) \right\}.$$

As $N(\bar{x};\bar{u}) = \emptyset$ if $\bar{u} \notin T_A(\bar{x})$, we also have

$$\operatorname{dom} D^*S((\bar{x},\bar{u});(u,v)) = \emptyset \quad \forall (u,v) \notin DS(\bar{x} \mid \bar{u}). \tag{15}$$

Definition 1 (Semismothness*). The set $A \subset \mathbb{E}$ is *semismooth** at $\bar{x} \subset A$ if

$$\langle x^*, u \rangle = 0 \quad \forall u \in \mathbb{E},\ x^* \in N_A(\bar{x};u).$$

The map $S : \mathbb{E}_1 \rightrightarrows \mathbb{E}_2$ is *semismooth** at $(\bar{x},\bar{y}) \in \operatorname{gph} S$ if $\operatorname{gph} S$ is semismooth* at $(\bar{x},\bar{y})$, i.e.,

$$\langle u, u^* \rangle = \langle v, v^* \rangle \quad \forall (u,v) \in \mathbb{E}_1 \times \mathbb{E}_2,\ (v^*,u^*) \in \operatorname{gph} D^*S((\bar{x},\bar{u});(u,v)).$$

The notion of *metric (sub)regularity* is used only in the next two results, and hence we refer the reader to the abundant literature for a definition, e.g., [18].

Proposition 2 **(Metric regularity and semismoothness*)**. *Let $F : \mathbb{E}_1 \to \mathbb{E}_2$ be continuously differentiable at $\bar{x}$, let $Q \subset \mathbb{E}_2$ be semismooth* (as a set) at $F(\bar{x})$ and let $S : \mathbb{E}_1 \rightrightarrows \mathbb{E}_2$, $S(x) := F(x) - Q$ be metrically subregular at $(\bar{x},0)$. Then $F^{-1}(Q)$ is semismooth* at $\bar{x}$ (as a set).*

Proof. By [6, Theorem 3.1], for any $h \in \mathbb{E}_1$,

$$N_{F^{-1}(Q)}(\bar{x};h) \subset F'(\bar{x})^* N_Q(F(\bar{x});F'(\bar{x})h), \tag{16}$$

see also [6, Remark 2.1]. Since Q is semismooth* at $F(\bar{x})$,

$$\langle v, z \rangle = 0 \quad \forall z \in \mathbb{E}_2,\ v \in N_Q(F(\bar{x});z).$$

Therefore

$$\langle v, F'(\bar{x})h \rangle = 0 \quad \forall h \in \mathbb{E}_1,\ v \in N_Q(F(\bar{x});F'(\bar{x})h),$$

and hence

$$\langle u, h \rangle = 0 \quad \forall h \in \mathbb{E}_1,\ u \in F'(\bar{x})^* N_Q(F(\bar{x});F'(\bar{x})h).$$

By (16) this implies that

$$\langle u, h \rangle = 0 \quad \forall h \in \mathbb{E}_1,\ u \in N_{F^{-1}(Q)}(\bar{x};h),$$

i.e., $F^{-1}(Q)$ is semismooth* at $\bar{x}$. □

Corollary 3 **(Semismoothness* of the infimal convolution solution map)**. *Let $f \in \Gamma_0(\mathbb{E})$, let $(\bar{x},\bar{\lambda}) \in \mathbb{E} \times \mathbb{R}_{++}$ and let ω be strongly convex and twice continuously differentiable. Then the map $P_{\omega,f}$ from (13) is semismooth* at $((\bar{x},\bar{\lambda}), P_{\omega,f}(\bar{x},\bar{\lambda}))$ if ∂f is semismooth* at $\big(P_{\omega,f}(\bar{x},\bar{\lambda}), \nabla\omega(\frac{1}{\bar{\lambda}}[\bar{x} - P_{\omega,f}(\bar{x},\bar{\lambda})])\big)$.*

Proof. Without loss of generality, assume $\mathbb{E}=\mathbb{R}^n$. Let $F:\mathbb{R}^n\times\mathbb{R}_{++}\times\mathbb{R}^n\to\mathbb{R}^{2n}$, $F(x,\lambda,z):=(z,\nabla\omega([x-z]/\lambda)$. Then for all $x,z\in\mathbb{R}^n$ and $\lambda>0$, setting $V:=\nabla^2\omega([x-z]/\lambda)\succ 0$ we have

$$F'(x,\lambda,z)=\begin{pmatrix} 0 & 0 & I\\ \frac{1}{\lambda}V & -\frac{1}{\lambda^2}V(x-z) & -\frac{1}{\lambda}V\end{pmatrix}.$$

Hence, $\ker F'(x,\lambda,z)^*=\{0\}$ for all $x,z\in\mathbb{R}^n$, $\lambda>0$. Thus, $(x,\lambda,z)\mapsto F(x,\lambda,z)-\operatorname{gph}\partial f$ is metrically regular. As $\operatorname{gph}P_{\omega,f}=F^{-1}(\operatorname{gph}\partial f)$, if ∂f is semismooth* at $\left(P_{\omega,f}(\bar x,\bar x),\nabla\omega\left(\frac{\bar x-P_{\omega,f}(\bar x,\bar\lambda)}{\bar\lambda}\right)\right)=F(\bar x,\bar\lambda,P_{\omega,f}(\bar x,\bar\lambda))$, by Proposition 2, $P_{\omega,f}$ is semismooth* at $((\bar x,\bar\lambda),P_{\omega,f}(\bar x,\bar\lambda))$. ☐

Corollary 3 provides a sufficient criterion for establishing semismoothness* of the solution map P on the interior of its domain. It will be a topic of future research to exploit this on a broad scale, but we can immediately state the following result for a function $f\in\Gamma_0(\mathbb{E})$ which is either twice continuously differentiable or *piecewise linear-quadratic* (PLQ) in the sense of Rockafellar and Wets [39, Definition 10.20].

Proposition 3 **(Semismoothness* of the subdifferential)**. *For $f\in\Gamma_0(\mathbb{E})$, the subgradient ∂f is semismooth* at $(\bar x,\bar y)\in\operatorname{gph}\partial f$ under one of the following conditions:*

(a) *f is twice continuously differentiable at $\bar x$;*

(b) *f is piecewise linear-quadratic (in which case ∂f is semismooth* on $\mathbb{E}$).*

Proof. Assume condition (a) holds. If f is twice continuously differentiable, then $D(\partial f)(\bar x\mid\bar y)=\nabla^2 f(\bar x)=D^*(\partial f)(\bar x\mid\bar y)$, see [39, Example 8.43]. Now let $(u,v)\in T_{\operatorname{gph}\partial f}(\bar x,\bar y)$, i.e., $v\in D(\partial f)(\bar x\mid\bar y)(u)=\{\nabla^2 f(\bar x)u\}$, and let $(x^*,y^*)\in N_{\operatorname{gph}\partial f}((\bar x,\bar y);(u,v))\subset N_{\operatorname{gph}\partial f}(\bar x,\bar y)$, hence $x^*\in D^*(\partial f)(\bar x\mid\bar y)(-y^*)=\{-\nabla^2 f(\bar x)y^*\}$. Thus, we have $\langle(x^*,y^*),(u,v)\rangle=\langle y^*,\nabla^2 f(\bar x)u\rangle-\langle\nabla^2 f(\bar x)y^*,u\rangle=0$.

Now assume condition (b) holds. It follows from [39, Proposition 12.30] that $\operatorname{gph}\partial f$ is a finite union of polyhedra. Then [23, Proposition 3.4/3.5] yields that $\operatorname{gph}\partial f$ is semismooth*, which gives the desired statement. ☐

3.4. Constrained optimization

We now consider an application of Theorem 1 to derive the variational properties of the optimal value of the constrained optimization problem

$$v:(x,\lambda)\in\mathbb{E}_x\times\mathbb{R}\mapsto\inf_{u\in\mathbb{E}_f}\{f(u)\mid L(u,x)\in\lambda S\},\tag{17}$$

where $S\subset\mathbb{E}_\omega$ is a closed convex set. This function can be viewed as a special case of (3), where $\omega=\delta_S$ for some closed convex set $S\subset\mathbb{E}_\omega$. To see this, it is sufficient to note that

$$\delta_S^\pi(z,t)=\begin{cases}\delta_{\lambda S}(z) & \text{if } \lambda>0,\\ \delta_{S^\infty}(z) & \text{if } \lambda=0,\\ +\infty & \text{otherwise,}\end{cases}$$

and thus $L(u,x)\in\lambda S$ if and only if $\delta_S^\pi(L(u,x),\lambda)$ vanishes. Let $S^\circ:=\{v\mid\langle v,s\rangle\le 1\,\forall x\in S\}$ be the polar to the set S.

The following result is an immediate consequence of the general study in Theorem 2.

Corollary 4 **(Conjugate and subdifferential of the constrained value function)**. *Let v be given by (17) with $S\subset\mathbb{E}_\omega$ closed and convex, and assume that*

$$\exists u\in\operatorname{ri}\operatorname{dom}f,\ x\in\mathbb{E}_x:\ L(u,x)\in\mathbb{R}_{++}(\operatorname{ri}S).$$

Then the following hold.

(a) *We have*

$$v^*(y,\mu) = \inf_w \{ f^*(w) \mid \exists a \in -\mu S^\circ \,:\, L^*(a) = (-w, y) \}.$$

If S is a cone then $v^*(y,\mu) = \inf_w \{ f^*(w) + \delta_{\mathbb{R}_-}(\mu) \mid (-w,y) \in L^*(S^\circ) \}$.

(b) *For any $(x,\lambda) \in \operatorname{dom} v$ and $\bar u \in \operatorname{argmin}_u \{ f(u) \mid L(u,x) \in \lambda S \}$,*

$$\partial v(x,\lambda) = \begin{cases} \{ (v, -\sigma_S(y)) \mid y \in N_S(L(\bar u, x)/\lambda),\ (0,v) \in \mathcal{D}(\bar u, y) \} & \text{if } \lambda > 0, \\ \{ (v, -\beta) \mid \exists y \in N_{S^\infty}(L(\bar u, y)) \cap (\beta S^\circ) \,:\, (0,v) \in \mathcal{D}(\bar u, y) \} & \text{if } \lambda = 0, \end{cases}$$

where $\mathcal{D}(u,y) := \partial f(u) \times \{0\} + L^*(y)$. If S is bounded (hence compact), then

$$\partial v(x,\lambda) = \begin{cases} \{ (v, -\sigma_S(y)) \mid y \in N_S\left(L(\bar u, x)/\lambda\right),\ (0,v) \in \mathcal{D}(\bar u, y) \} & \text{if } \lambda > 0, \\ \{ (v, -\beta) \mid \exists y \in \beta S^\circ \,:\, (0,v) \in \mathcal{D}(\bar u, y) \} & \text{if } \lambda = 0. \end{cases}$$

(c) *We have $v^* \in \Gamma_0(\mathbb{E}_x \times \mathbb{R})$ if and only if there exist $y \in \mathbb{E}_x$, $w \in \operatorname{dom} f^*$, $\beta \in \mathbb{R}$ such that $(-w,y) \in -\beta L^*(S^\circ)$. In this case, also $v \in \Gamma_0(\mathbb{E}_x \times \mathbb{R})$ and the infimum is attained when finite.*

Proof. Part (a) follows from Theorem 2(a) with $w^* = \sigma_S$. If S is a cone then $w^* = \delta_{S^\circ}$. Part (b) follows from Theorem 2(b), observing that $\omega^\infty = \delta_{S^\infty}$ and that $S^\infty = \{0\}$ if S is bounded, in which case $N_{S^\infty} = \mathbb{E}_\omega$. Part (c) follows from (a) and Theorem 2(c). □

3.4.1. Relaxed linear constraints As an immediate specialization of Corollary 4 we obtain a result on the value function

$$v : (b,\lambda) \in \mathbb{R}^m \times \mathbb{R} \mapsto \inf_{x \in \mathbb{R}^n} \{ f(x) \mid \|Ax - b\| \le \lambda \}, \tag{18}$$

where $f \in \Gamma_0(\mathbb{R}^n)$, $A \in \mathbb{R}^{m \times n}$ is a matrix, and $\|\cdot\|$ is any norm in $\mathbb{R}^n$. Denote the associated dual norm by $\|\cdot\|^\circ$, and the corresponding unit-norm ball by $\mathbb{B}$.

Corollary 5 **(Relaxed linear constraints value function)**. *If there exists a pair $(x,\lambda) \in \operatorname{dom} f \times \mathbb{R}_{++}$ such that $\|Ax - b\| < \lambda$, then the following hold.*

(a) *(conjugate) $v^*(y,\mu) = f^*(A^T y) + \delta_{\mu \mathbb{B}^\circ}(y)$, which is closed proper convex if and only if there exists β and $\|y\|^\circ \le \beta$ such that $A^T y \in \operatorname{dom} f^*$. In this case, v is closed proper convex and the infimum is attained when finite.*

(b) *(subdifferential) For any $(b,\lambda) \in \operatorname{dom} v$ and $\bar x$ that achieves the infimum in (18) (and hence $\|A\bar x - b\| \le \lambda$),*

$$\partial v(b,\lambda) = \begin{cases} \{ (y, -\|y\|^\circ) \mid y \in N_{\mathbb{B}}\left([A\bar x - b]/\lambda\right),\ -A^T y \in \partial f(\bar x) \} & \text{if } \lambda > 0, \\ \{ (y, -\beta) \mid \|y\|^\circ \le \beta,\ -A^T y \in \partial f(\bar x) \} & \text{if } \lambda = 0. \end{cases}$$

(c) *(primal existence) For $\lambda > 0$ and any $b \in \mathbb{R}^m$, if*

$$f^\infty(y) > 0 \quad \forall y \in \ker A \setminus \{0\}, \tag{19}$$

then $\operatorname{argmin}_x \{ f(x) + \delta^\pi_{\mathbb{B}}(Ax - b, \lambda) \} \ne \emptyset$. This holds, e.g., when f is level-bounded or $\operatorname{rank} A = n$.

Proof. Part (a). The expression for the conjugate v^* follows from Corollary 4(a) by observing that $L : (x,b) \mapsto Ax - b$ has adjoint $L^* : z \mapsto (A^T z, -z)$ and that $\sigma_{\mathbb{B}} = \|\cdot\|^\circ$. The remaining claims for Part (a) follow from Theorem 1.

Part (b) follows from Corollary 4(b) with the foregoing observations. d) For $\lambda > 0$ and $b \in \mathbb{R}^m$, the effective objective function in (18) is $\phi(x) := f(x) + \delta_{\lambda\mathbb{B}}(Ax - b)$. With $\hat{x}$ such that $\|A\hat{x} - b\| \leq \lambda$, which exists by the hypothesis of the theorem, we have

$$(\delta_{\lambda\mathbb{B}} \circ (A(\cdot) - b))^\infty(x) = \sup_{\tau>0} \delta_{\lambda\mathbb{B}}(A\hat{x} - b + \tau Ax) = \delta_{\ker A}(x),$$

where the second identity uses the property that $\lambda\mathbb{B}$ is bounded. With [39, Exercise 3.29] we hence find that $\phi^\infty = f^\infty + \delta_{\ker A}$, which shows, using [39, Theorem 3.26], that ϕ is level-bounded if (19) holds. □

4. Moreau envelope and proximal map

In this section we outline existing and new results regarding the variational properties of the Moreau envelope and the proximal map of a closed proper convex function.

4.1. The Moreau envelope

The *Moreau envelope* of $f \in \Gamma_0(\mathbb{E})$ is defined by

$$e_\lambda f(x) := \min_{u\in\mathbb{E}} \left\{ f(u) + (1/2\lambda)\|x - u\|^2 \right\} \quad \forall x \in \mathbb{E},\ \lambda > 0,$$

which has a Lipschitz gradient given by $\nabla e_\lambda f(x) = \frac{1}{\lambda}(x - P_\lambda f(x))$.

The following result summarizes limiting properties of the Moreau envelope as $\lambda \downarrow 0$.

Proposition 4 **(Convergence of the Moreau envelope)**. *For $f \in \Gamma_0(\mathbb{E})$, the following hold as $\lambda \downarrow 0$:*

(a) $e_\lambda f \xrightarrow{e} f$ *and* $e_\lambda f \xrightarrow{p} f$ *(in fact $e_\lambda f(x) \uparrow f(x)$ for all $x \in \mathbb{E}$);*
(b) $\lambda f \xrightarrow{e} \delta_{\mathrm{cl}\,(\mathrm{dom}\, f)}$*;*
(c) $\lambda e_\lambda f(x) \to \frac{1}{2} d^2_{\mathrm{cl}\,(\mathrm{dom}\, f)}(\bar{x})$ *as* $x \to \bar{x}$*;*
(d) $\lambda \partial f$ *converges to* $N_{\mathrm{cl}\,(\mathrm{dom}\, f)}$ graphically *in the sense of [39, Definition 5.32];*
(e) *for* $x \in \mathrm{dom}\, \partial f$ *we have* $\nabla e_\lambda f(x) \to \mathrm{argmin}_{g\in\partial f(x)} \|g\|$.

Proof. Part (a). See, e.g., [39, Theorem 1.25, Proposition 7.4].

Part (b). By Lemma 1(b) and [38, Theorem 13.3], $\lambda \star f^* \xrightarrow{e} (f^*)^\infty = \sigma_{\mathrm{dom}\, f}$. Wijsman's theorem [39, Theorem 11.34] then yields $\lambda f = (\lambda \star f^*)^* \xrightarrow{e} \delta_{\mathrm{cl}\,(\mathrm{dom}\, f)}$.

Part (c). By Part (b), $\lambda f \xrightarrow{e} \delta_{\mathrm{cl}\,(\mathrm{dom}\, f)}$. Hence, by [39, Theorem 7.37],

$$\lambda e_\lambda f = e_1(\lambda f) \xrightarrow{c} e_1 \delta_{\mathrm{cl}\,(\mathrm{dom}\, f)} = \tfrac{1}{2} d^2_{\mathrm{cl}\,(\mathrm{dom}\, f)}.$$

Part (d). Follows from Part (b) and Attouch [39, Theorem 12.35].

Part (e). See [2, Remark 3.32]. □

Note that Proposition 4(e) implies that there exists $K > 0$ such that

$$\forall \bar{x} \in \mathrm{dom}\, \partial f\ \exists K > 0\ \forall \lambda > 0 : \|P_\lambda f(\bar{x}) - \bar{x}\| \leq K\lambda. \tag{20}$$

Proposition 4(a) suggests the following extension of the Moreau envelope at $\lambda = 0$:

$$p_f : (x, \lambda) \in \mathbb{E} \times \mathbb{R} \mapsto \begin{cases} e_\lambda f(x) & \text{if } \lambda > 0, \\ f(x) & \text{if } \lambda = 0, \\ +\infty & \text{if } \lambda < 0. \end{cases}$$

This is exactly the value function $p_{\omega,f}$ from (11) with $\omega = \frac{1}{2}\|\cdot\|^2$. Hence, we may rely on our general study on infimal convolution from Section 3.3 to understand the properties of this extension of the Moreau envelope.

Corollary 6 **(Conjugate and subdifferential of the Moreau envelope)**. *Let $f\in\Gamma_0(\mathbb{E})$. Then $p_f\in\Gamma_0(\mathbb{E}\times\mathbb{R})$ and*

(a) $p_f^*(y,\mu)=f^*(y)+\delta_{\mathrm{epi}\,\frac{1}{2}\|\cdot\|^2}(y,-\mu)$ *and* $p_f^*\in\Gamma_0(\mathbb{E}\times\mathbb{R})$*;*

(b) *for all* $(x,\lambda)\in\operatorname{dom}p_f$*,*

$$\partial p_f(x,\lambda)=\begin{cases}\left(\frac{1}{\lambda}[x-P_\lambda f(x)],\ -\frac{1}{2}\|\frac{1}{\lambda}[x-P_\lambda f(x)]\|^2\right) & \text{if } \lambda>0,\\ \left\{(v,\beta)\ \middle|\ -v\in\partial f(x),\ \frac{1}{2}\|v\|^2\le\beta\right\} & \text{if } \lambda=0.\end{cases}$$

Proof. We are in the situation of Corollary 1 with $\omega=\frac{1}{2}\|\cdot\|^2$. In particular, the qualification condition (12) is trivially satisfied. □

4.2. Properties of the proximal map We now turn our attention to the proximal map. It is straightforward to show $P_\lambda f(x)\to x$ as $\lambda\downarrow 0$ for any $x\in\operatorname{dom}f$. The following proposition, which generalizes this statement, can be derived from monotone operator theory [39, Theorem 12.37]. The proof that we provide here instead relies on epigraphical convergence.

Proposition 5 **(Convergence of the proximal map)**. *Let $f\in\Gamma_0(\mathbb{E})$ and $\bar{x}\in\mathbb{E}$. Then* $\lim_{\substack{\lambda\downarrow 0,\\ x\to\bar{x}}}P_\lambda f(x)=P_{\mathrm{cl}\,(\mathrm{dom}\,f)}(\bar{x})$.

Proof. Let $\{\lambda_k\}\downarrow 0$, $\{x_k\}\to\bar{x}$, and $\phi_k(u):=\lambda_k f(u)+\frac{1}{2}\|u-x_k\|^2$. Use Proposition 4(b) to deduce $\lambda_k f\xrightarrow{e}\delta_{\mathrm{cl}\,(\mathrm{dom}\,f)}$. Then because $\frac{1}{2}\|(\cdot)-x_k\|^2\xrightarrow{c}\frac{1}{2}\|(\cdot)-\bar{x}\|^2$, we obtain $\phi_k\xrightarrow{e}\phi:=\delta_{\mathrm{cl}\,(\mathrm{dom}\,f)}+\frac{1}{2}\|(\cdot)-\bar{x}\|^2$; see [39, Theorem 7.46 b)]. Now observe that $P_{\lambda_k}f(x_k)=\operatorname{argmin}\phi_k$ and $P_{\mathrm{cl}\,(\mathrm{dom}\,f)}(\bar{x})=\operatorname{argmin}\phi$. Since all functions ϕ_k are convex and ϕ is level-bounded (in fact, strongly convex), the sequence $\{\phi_k\}$ is, by [39, Exercise 7.32 c)], eventually level-bounded (in the sense of [39, p. 266]). Therefore, we can apply [39, Theorem 7.33], with $\varepsilon_k=0$ $(k\in\mathbb{N})$, to deduce $P_{\lambda_k}f(x_k)\to P_{\mathrm{cl}\,(\mathrm{dom}\,f)}(\bar{x})$. □

We record the following auxiliary result.

Lemma 5. *Let $f\in\Gamma_0(\mathbb{E})$ and fix positive scalars λ and μ. Then for all $x\in\mathbb{E}$,*

$$\begin{aligned}\frac{1}{2\mu}&\left(\|P_\mu f(x)-x\|^2-\|P_\lambda f(x)-x\|^2+\|P_\mu f(x)-P_\lambda f(x)\|^2\right)\\ &\le f(P_\lambda f(x))-f(P_\mu f(x))\\ &\le\frac{1}{2\lambda}\left(\|P_\mu f(x)-x\|^2-\|P_\lambda f(x)-x\|^2-\|P_\mu f(x)-P_\lambda f(x)\|^2\right),\end{aligned}\tag{21}$$

and

$$\|P_\lambda f(x)-P_\mu f(x)\|^2\le\frac{\mu-\lambda}{\lambda+\mu}\left(\|P_\mu f(x)-x\|^2-\|P_\lambda f(x)-x\|^2\right).\tag{22}$$

Proof. Set $P(\tau):=P_\tau f(\bar{x})$ for all $\tau>0$. To obtain the bounds in (21), use [39, Eq. 7(34)] to infer

$$f(x)+\frac{1}{2\tau}\|x-\bar{x}\|^2-f(P(\tau))-\frac{1}{2\tau}\|P(\tau)-\bar{x}\|^2\ge\frac{1}{2\tau}\|x-P(\tau)\|^2\quad\forall\tau>0,\ \forall x\in\mathbb{E}.$$

For $\tau=\lambda$ and $x=P(\mu)$, we hence obtain

$$f(P(\mu))+\frac{1}{2\lambda}\|P(\mu)-\bar{x}\|^2-f(P(\lambda))-\frac{1}{2\lambda}\|P(\lambda)-\bar{x}\|^2\ge\frac{1}{2\lambda}\|P(\mu)-P(\lambda)\|^2.$$

Analogously, for $\tau=\mu$ and $x=P(\lambda)$, we find that

$$f(P(\lambda))+\frac{1}{2\mu}\|P(\lambda)-\bar{x}\|^2-f(P(\mu))-\frac{1}{2\mu}\|P(\mu)-\bar{x}\|^2\ge\frac{1}{2\mu}\|P(\lambda)-P(\mu)\|^2.$$

Combining the last two inequalities now yields (21).

Next, use (21) to obtain

$$\frac{1}{\mu}\left(\|P(\mu)-x\|^2-\|P(\lambda)-x\|^2+\|P(\lambda)-P(\mu)\|^2\right) \\ \le \frac{1}{\lambda}\left(\|P(\mu)-x\|^2-\|P(\lambda)-x\|^2-\|P(\lambda)-P(\mu)\|^2\right),$$

or, equivalently

$$\left(\frac{1}{\lambda}+\frac{1}{\mu}\right)\|P(\lambda)-P(\mu)\|^2 \le \left(\frac{1}{\lambda}-\frac{1}{\mu}\right)\left(\|P(\mu)-x\|^2-\|P(\lambda)-x\|^2\right),$$

which is equivalent to the desired inequality (22) □

4.3. Proximal map extension Proposition 5 suggests the following extension of the proximal map of $f\in\Gamma_0(\mathbb{E})$:

$$P_f:\mathbb{E}\times\mathbb{R}\rightrightarrows\mathbb{E},\quad P_f(x,\lambda):=\begin{cases}P_\lambda f(x) & \text{if } \lambda>0,\\ P_{\mathrm{cl}\,(\mathrm{dom}\, f)}(x) & \text{if } \lambda=0,\\ \emptyset & \text{if } \lambda<0.\end{cases}$$

The next result clarifies continuity properties of the proximal map extension P_f.

Corollary 7 **(Lipschitz continuity of the proximal map)**. *Let $f\in\Gamma_0(\mathbb{E})$. Then P_f is continuous on* $\mathrm{dom}\, P_f=\mathbb{E}\times\mathbb{R}_+$ *and is locally Lipschitz on* $\mathrm{int}\,(\mathrm{dom}\, P_f)$. *If $\bar{x}\in\mathrm{dom}\,\partial f$, then P_f is* upper Lipschitz *(or* calm*) at $(\bar{x},0)$, and the map $\mathbb{R}_+\ni\mu\mapsto P_f(\bar{x},\mu)$ is locally Lipschitz at 0, i.e., there exist positive scalars κ and ε such that*

$$\|P_f(\bar{x},0)-P_f(x,\lambda)\|\le\kappa\|(\bar{x}-x,\lambda)\|\quad\forall(x,\lambda)\in B_\varepsilon(\bar{x},0)\cap\mathrm{dom}\, P_f, \tag{23a}$$

$$\|P_f(\bar{x},\lambda)-P_f(\bar{x},\mu)\|\le\kappa|\mu-\lambda|\qquad\forall\lambda,\mu\in[0,\varepsilon]. \tag{23b}$$

Proof. The continuity to the boundary of the domain follows from Proposition 5. The local Lipschitz continuity on $\mathrm{int}\,(\mathrm{dom}\, P_f)$ follows from Theorem 3 with $\omega=\frac{1}{2}\|\cdot\|^2$.

Now assume that $\bar{x}\in\mathrm{dom}\,\partial f$, which implies $P_f(\bar{x},0)=\bar{x}\in\mathrm{dom}\, f$. Then for all $\lambda>0$,

$$\|P_f(x,\lambda)-P_f(\bar{x},0)\|\le\|P_\lambda f(x)-P_\lambda f(\bar{x})\|+\|\bar{x}-P_\lambda f(\bar{x})\|\le\|x-\bar{x}\|+K\lambda,$$

where $K>0$ is given via (20) and we use the property that $P_\lambda f$ is 1-Lipschitz [3]. Set $\kappa:=\max\{1,K\}$ to obtain (23a).

Let $P:=P_f(\bar{x},\cdot)$. By (23a), there exist positive scalars κ and ε such that $\|P(\tau)-\bar{x}\|\le\kappa\tau$ for all $\tau\in(0,\varepsilon]$. Hence for μ and λ in $(0,\varepsilon]$,

$$\begin{aligned}\|P(\mu)-P(\lambda)\|^2 &\le \tfrac{\mu-\lambda}{\mu+\lambda}\left(\|P(\mu)-\bar{x}\|^2-\|P(\lambda)-\bar{x}\|^2\right)\\ &= \tfrac{\mu-\lambda}{\mu+\lambda}\left(\|P(\mu)-\bar{x}\|-\|P(\lambda)-\bar{x}\|\right)\cdot\left(\|P(\mu)-\bar{x}\|+\|P(\lambda)-\bar{x}\|\right)\\ &\le \tfrac{\mu-\lambda}{\mu+\lambda}\kappa(\mu+\lambda)\left(\|P(\mu)-\bar{x}\|-\|P(\lambda)-\bar{x}\|\right)\\ &\le \kappa|\mu-\lambda|\cdot\|P(\mu)-P(\lambda)\|,\end{aligned}$$

where the first inequality follows from (22) of Lemma 5, and the last inequality uses the reverse triangle inequality. Use (23a) to obtain (23b). □

The following example shows that the assumption $\bar{x}\in\mathrm{dom}\,\partial f$ required for Eq. (23) is not redundant.

Example 1 (Upper Lipschitz continuity of proximal map). Consider the following two functions, both contained in $\Gamma_0(\mathbb{R})$:

$$f(x) = \begin{cases} -\log x & \text{if } x > 0, \\ +\infty & \text{otherwise,} \end{cases} \qquad g(x) = \begin{cases} -\sqrt{x} & \text{if } x \geq 0, \\ +\infty & \text{otherwise.} \end{cases}$$

The corresponding extended proximal maps are

$$P_f(x,\lambda) = \begin{cases} \frac{1}{2}(x + \sqrt{x^2+4\lambda}) & \text{if } \lambda > 0, \\ \max\{x, 0\} & \text{if } \lambda = 0, \end{cases} \qquad P_g(0,\lambda) = \left(\frac{\lambda}{2}\right)^{2/3} \quad \forall \lambda \geq 0;$$

cf. Beck [4, Lemma 6.5] for the expression for P_f. Observe that $\operatorname{dom} f$ does not include the origin, and $|P_f(0,0) - P_f(0,\lambda)| = \sqrt{\lambda}$ for all $\lambda > 0$, which is not upper Lipschitz at $(0,0)$. Next, observe that $\operatorname{dom} \partial g$ does not include the origin, and P_g is not upper Lipschitz at $(0,0)$. ⋄

The next result on directional differentiability of P_f follows from Theorem 3(c) with $\omega = \frac{1}{2}\|\cdot\|^2$.

Corollary 8 **(Directional differentiability of the proximal map)**. *Let $f \in \Gamma_0(\mathbb{E})$ and fix $(x,\lambda) \in \mathbb{E} \times \mathbb{R}_{++}$. If ∂f is proto-differentiable at $\left(P_f(x,\lambda),\, \frac{1}{\lambda}[x - P_f(x,\lambda)]\right)$, then P_f is directionally differentiable at (x,λ) with*

$$P_f'((x,\lambda);(d,\Delta)) = \left[\lambda D(\partial f)\left(P_f(x,\lambda) \;\middle|\; \tfrac{1}{\lambda}[x - P_f(x,\lambda)]\right) + I\right]^{-1} \left(d - \Delta \tfrac{1}{\lambda}[x - P_f(x,\lambda)]\right)$$

for all $(d,\Delta) \in \mathbb{E} \times \mathbb{R}$. In particular, for any $\lambda > 0$,

$$(P_\lambda f)'(x;\cdot) = \left[\lambda D(\partial f)\left(P_f(x,\lambda) \,\middle|\, \tfrac{1}{\lambda}[x - P_f(x,\lambda)]\right) + I\right]^{-1}(\cdot)$$

4.3.1. Semismoothness* of P_f We now establish semismoothness* of the extended proximal map P_f on $\mathbb{E} \times \mathbb{R}_{++}$. We lead with an auxiliary result.

Lemma 6. *The map $S : \mathbb{E}_1 \rightrightarrows \mathbb{E}_2$ is semismooth* at $(y, z - y)$ if and only is $S + \mathrm{id}$ is semismooth* at (y,z).*

Proof. The map S is semismooth* at $(y, z-y)$ if and only if

$$v \in DS(y \mid z-y)(u),\; u^* \in D^*S((y,z-y);(u,v))(v^*) \Rightarrow \langle u,\, u^*\rangle = \langle v,\, v^*\rangle$$

$$\iff \left\{ \begin{array}{l} v + u \in D(S+\mathrm{id})(y|z)(u), \\ u^* + v^* \in D^*(S+\mathrm{id})((y,z);(u,u+v))(v^*) \end{array} \right\} \Rightarrow \langle u,\, u^*+v^*\rangle = \langle u+v,\, v^*\rangle$$

$$\iff S + \mathrm{id} \text{ semismooth* at } (y,z).$$

Here the first equivalence is the definition of semismoothness* and (15). The second uses the sum rule for the graphical derivative [39, Exercise 10.43] and the directional coderivative [6, Corollary 5.3 (+ comment)], respectively, when one summand is smooth (here the identity map). The last equivalence is a variable change and the definition of semismoothness* and (15) again. □

Proposition 6 **(Semismoothness* of P_f)**. *For $f \in \Gamma_0(\mathbb{E})$,*

(a) *P_f is semismooth* at (x,λ) if ∂f semismooth* at $\left(P_f(x,\lambda),\, \frac{1}{\lambda}[x - P_f(x,\lambda)]\right)$;*

(b) *$P_\lambda f$ is semismooth* at x if and only if ∂f is semismooth* at $\left(P_\lambda f(x), \frac{1}{\lambda}[x - P_\lambda f(x)]\right)$.*

Proof. Part (a) follows from Corollary 3 with $\omega = \frac{1}{2}\|\cdot\|^2$. For Part (b), observe that $P_\lambda f = (\lambda \partial f + \mathrm{id})^{-1}$ is semismooth* at x if and only if $\lambda\partial f + \mathrm{id}$ is semismooth* at $(P_\lambda f(x),\, x)$ [23, p. 7]. By Lemma 6, this is the case if and only if $\lambda \partial f$ is semismooth* at $(P_\lambda f(x),\, x - P_\lambda f(x))$ which, in turn, holds if and only if ∂f is semismooth* at $\left(P_\lambda f(x),\, \frac{1}{\lambda}[x - P_\lambda f(x)]\right)$. □

Various papers study the semismoothness à la Qi and Sun [37] of P_f on $\mathbb{E}\times\mathbb{R}_{++}$. Most of these results, trace the semismoothness of the latter back to the semismoothness of the Euclidean projection onto epi f. The work by Meng et al. [31, 32] deserves explicit mention, and a good discussion of these results can be found in Milzarek's thesis [33]. Bearing our applications in Section 7 in mind, this is somewhat of a circular strategy, and hence we opened up a different path via our study in Section 3.3.1 on semismooth* properties of solution maps. For a map that is locally Lipschitz at a point, semismoothness* differs from traditional semismoothness only in directional differentiability as the following result by Gfrerer and Outrata [23, Corollary 3.8] shows.

Lemma 7 **(Semismooth vs. semismooth*)**. *Let $F: D\subset\mathbb{E}_1\to\mathbb{E}_2$ be locally Lipschitz at $x\in\operatorname{int} D$. Then the following are equivalent:*

(a) *F is semismooth at x;*
(b) *F is semismooth* and directionally differentiable at x.*

This lemma gives the following immediate consequence about semismoothness of P_f.

Corollary 9 **(Semismoothness of P_f)**. *Let $f\in\Gamma_0(\mathbb{E})$ and fix $(x,\lambda)\in\mathbb{E}\times\mathbb{R}_{++}$. If ∂f is proto-differentiable and semismooth* at $\big(P_f(x,\lambda),\,\frac{1}{\lambda}[x-P_f(x,\lambda)]\big)$, then P_f is semismooth at (x,λ). This holds, in particular, if f is PLQ or twice continuously differentiable at $P_f(x,\lambda)$, in which case P_f is continuously differentiable at (x,λ).*

Proof. For the first statement combine Corollary 8, Proposition 6, and Lemma 7. For the second, invoke Remark 1 and Proposition 3. □

Note that semismoothness* does not require directional differentiability of the function in question. However, semismoothness* is still sufficient to yield convergence of Newton-type methods under suitable regularity conditions [23, 27]. In view of the above discussion, this is important because the Euclidean projector onto a closed convex set may not be directionally differentiable [40], in which case the arguments and methods based on (standard) semismoothness are invalidated.

5. The proximal value

The projection onto the epigraph of a function $f\in\Gamma_0(\mathbb{E})$ requires a particular value of λ so that the equation (2) holds. In this section we examine the variational properties of the value of the proximal map as a function of λ, i.e., the function

$$0<\lambda\mapsto f(P_\lambda f(\bar{x})), \tag{24}$$

where $\bar{x}\in\mathbb{E}$ is fixed. Note that this map is not generally convex, as illustrated by the following counterexample.

Example 2 (Nonconvexity of the proximal value). Define $f=|\cdot|+\delta_{[-1,1]}\in\Gamma_0(\mathbb{R})$. By Beck [4, Example 6.22],

$$P_\lambda f(x)=\min\{\max\{|x|-\lambda,0\},\,1\}\cdot\operatorname{sgn}(x)\quad\forall x\in\mathbb{R},\ \lambda>0.$$

Hence, for $\bar{x}=2$, we obtain the nonconvex function

$$f(P_\lambda f(\bar{x}))=\begin{cases}1 & \text{if } \lambda\in(0,1],\\ 2-\lambda & \text{if } \lambda\in(1,2],\\ 0 & \text{if } \lambda>2.\end{cases}$$

◇

The next result describes the monotonicity and continuity of the map (24).

Corollary 10 **(Monotonicity and continuity in λ)**. *Let $f\in\Gamma_0(\mathbb{E})$ and fix $\bar{x}\in\mathbb{E}$. Then*

(a) *$0<\lambda\mapsto f(P_\lambda f(\bar{x}))$ is decreasing (i.e., increasing as $\lambda\downarrow 0$);*

(b) $0 < \lambda \mapsto \|\bar{x} - P_\lambda f(\bar{x})\|$ *is increasing;*
(c) $\lim_{\lambda\to 0} f(P_\lambda f(\bar{x})) = f(P_{\mathrm{cl}\,(\mathrm{dom}\, f)}(\bar{x}))$.

Proof. Parts (a) and (b). Let $0 < \lambda < \mu$ and set $P(\lambda) := P_\lambda f(\bar{x})$, $P(\mu) := P_\mu f(\bar{x})$, and $\delta := \frac{1}{2}(\|P(\mu) - x\|^2 - \|P(\lambda) - x\|^2)$. Then from (21) of Lemma 5, we obtain

$$\frac{1}{\mu}\delta \le f(P(\lambda)) - f(P(\mu)) \le \frac{1}{\lambda}\delta.$$

As $\lambda < \mu$, this implies that $\delta \ge 0$, i.e., $\|P(\mu) - x\|^2 \ge \|P(\lambda) - x\|^2$, and hence $f(P(\mu)) \le f(P(\lambda))$.

Part (c). Let $\{\lambda_k\} \downarrow 0$. Then $p_k := P_{\lambda_k} f(\bar{x}) \to p := P_{\mathrm{cl}\,(\mathrm{dom}\, f)}(\bar{x})$; see Proposition 5. It follows that

$$\begin{aligned} f(p) &\ge \limsup_{k\to\infty} \left[f(p_k) + \tfrac{1}{2\lambda_k}\left(\|\bar{x} - p_k\|^2 - \|\bar{x} - p\|^2\right) \right] \\ &\ge \limsup_{k\to\infty} f(p_k) \\ &\ge \liminf_{k\to\infty} f(p_k) \\ &\ge f(p). \end{aligned}$$

Here the first inequality uses that $f(p) + \frac{1}{2\lambda_k}\|\bar{x} - p\|^2 \ge f(p_k) + \frac{1}{2\lambda_k}\|\bar{x} - p_k\|^2$ for all $k \in \mathbb{N}$, by definition of p_k. The second is due to $\|\bar{x} - p_k\| \ge \|\bar{x} - p\|$, by the definition of p and since $p_k \in \mathrm{dom}\, f$. The last one is just lower semicontinuity of f. ☐

As we did with the Moreau envelope and proximal map, we define the extension of the map (24) to include negative values of λ:

$$\eta^f_{\bar{x}} : \lambda \in \mathbb{R} \mapsto \begin{cases} f(P_\lambda f(\bar{x})) & \text{if } \lambda > 0, \\ f(P_{\mathrm{cl}\,(\mathrm{dom}\, f)}(\bar{x})) & \text{if } \lambda \le 0. \end{cases}$$

We call this the *proximal value function*. Observe that

$$\eta^f_{\bar{x}}(\lambda) = e_\lambda f(\bar{x}) - (1/2\lambda)\|\bar{x} - P_\lambda f(\bar{x})\|^2 \quad (\lambda > 0). \tag{25}$$

We use Corollary 10 to derive the following result.

Corollary 11 **(Continuity properties of the proximal value)**. *Let $f \in \Gamma_0(\mathbb{E})$ and fix $\bar{x} \in \mathbb{E}$. Then the following hold:*
(a) *$\eta^f_{\bar{x}}$ is decreasing, continuous (possibly in an extended real-valued sense), and finite-valued if (and only if) $P_{\mathrm{cl}\,(\mathrm{dom}\, f)}(\bar{x}) = \bar{x} \in \mathrm{dom}\, f$.*
(b) *$\eta^f_{\bar{x}}$ is locally Lipschitz on $\mathbb{R}_{++}$.*
(c) *If $\bar{x} \in \mathrm{dom}\, \partial f$, then the assertion in (b) holds on $\mathbb{R}$.*

Proof. Set $\eta := \eta^f_{\bar{x}}$. Parts (a) and (b). The fact that η is decreasing follows from Corollary 10(a). Now consider (25). By Corollary 6, the map $0 < \lambda \mapsto e_\lambda f(\bar{x})$ is convex and finite-valued, hence locally Lipschitz. By Corollary 7(a), this conclusion also holds for $0 < \lambda \mapsto \frac{1}{2\lambda}\|x - P_\lambda f(\bar{x})\|^2$. This gives the local Lipschitz continuity of η on $\mathbb{R}_{++}$. The continuity at 0 is due to Corollary 10(c).

Part (c). By Parts (a) and (b), and because η is constant (and finite by assumption) on $\mathbb{R}_-$, we only need to be concerned about the desired properties at 0. To this end, let $\mu > \lambda$. If $\lambda < 0$, then

$$\left|\frac{\eta(\mu) - \eta(\lambda)}{\mu - \lambda}\right| \le \left|\frac{\eta(\mu) - \eta(0)}{\mu - 0}\right|.$$

Thus we can restrict ourselves to the case $0 \le \lambda < \mu$. Set $P(\tau) := P_\tau f(\bar{x})$ for all $\tau > 0$ and $P(0) := \bar{x}$. Then by Corollary 7(c), there exist positive scalars ε and κ such that

$$\|P(\mu) - P(\lambda)\| \le \kappa(\mu - \lambda) \quad \forall 0 \le \lambda \le \mu \le \varepsilon. \tag{26}$$

For $0 < \lambda < \mu \le \varepsilon$, we have

$$\begin{aligned}|\eta(\lambda) - \eta(\mu)| &= \eta(\lambda) - \eta(\mu)\\ &\le \tfrac{1}{2\lambda}\left(\|P(\mu) - \bar{x}\|^2 - \|P(\lambda) - \bar{x}\|^2 - \|P(\mu) - P(\lambda)\|^2\right)\\ &= \tfrac{1}{2\lambda}\left[(\|P(\mu) - \bar{x}\| - \|P(\lambda) - \bar{x}\|)\cdot(\|P(\mu) - \bar{x}\| + \|P(\lambda) - \bar{x}\|) - \|P(\mu) - P(\lambda)\|^2\right]\\ &\le \tfrac{1}{2\lambda}\|P(\mu) - P(\lambda)\|\cdot(\|P(\mu) - \bar{x}\| + \|P(\lambda) - \bar{x}\| - \|P(\mu) - P(\lambda)\|)\\ &\le \tfrac{\kappa}{2\lambda}|\mu - \lambda|\left(\|P(\mu) - \bar{x}\| + \|P(\lambda) - \bar{x}\| - (\|P(\mu) - \bar{x}\| - \|P(\lambda) - \bar{x}\|)\right)\\ &= \tfrac{\kappa}{\lambda}\|\bar{x} - P(\lambda)\|\cdot|\mu - \lambda|\\ &\le \kappa^2|\mu - \lambda|.\end{aligned}$$

Here, the first identity follows from Corollary 10(a), where the first inequality uses Lemma 5(a). The rest the follows from the reverse triangle inequality and (26), recalling that $P(0) = \bar{x}$. □

REMARK 2. The requirement that $\bar{x} \in \partial f$, made in Corollary 11, cannot be relaxed to $\bar{x} \in \operatorname{dom} f$. To see this, we again use Example 1(b), where

$$\eta_{\bar{x}}^f(\lambda) = \begin{cases} -(\lambda/2)^{\frac{1}{3}} & \text{if } \lambda \ge 0,\\ 0 & \text{if } \lambda < 0,\end{cases}$$

which is neither locally Lipschitz nor directionally differentiable at $\lambda = 0$. We also conclude from this example that the lack of calmness of the proximal map at $\lambda = 0$ is not necessarily compensated by applying f.

Under certain assumptions described by Corollary 12, we may interpret the extended proximal value function $\eta_{\bar{x}}^f$ as the derivative of the convex function

$$\bar{\phi}_{\bar{x}}^f : \lambda \in \mathbb{R} \mapsto \begin{cases} -\lambda e_\lambda f(\bar{x}) & \text{if } \lambda > 0,\\ -\frac{1}{2}d^2_{\operatorname{cl}(\operatorname{dom} f)}(\bar{x}) & \text{if } \lambda = 0,\\ -\lambda f(P_{\operatorname{cl}(\operatorname{dom} f)}(\bar{x})) - \frac{1}{2}d^2_{\operatorname{cl}(\operatorname{dom} f)}(\bar{x}) & \text{if } \lambda < 0;\end{cases} \tag{27}$$

cf. Attouch [2, Remark 3.32].

COROLLARY 12 **(The function $\bar{\phi}_{\bar{x}}^f$).** *Let $f \in \Gamma_0(\mathbb{E})$ and fix $\bar{x} \in \mathbb{E}$. Then the following hold:*

(a) *$\bar{\phi}_{\bar{x}}^f$ is proper, convex and continuous (possibly in an extended real-valued sense), and continuously differentiable on $\mathbb{R}_{++}$ with $\frac{d}{d\lambda}\bar{\phi}_{\bar{x}}^f(\lambda) = -f(P_\lambda f(\bar{x}))$ locally Lipschitz for all $\lambda > 0$.*

(b) *If $\bar{x} \in \operatorname{dom} f$, then $\bar{\phi}_{\bar{x}}^f$ is continuously differentiable on $\mathbb{R}$ with derivative given by*

$$\frac{d}{d\lambda}\bar{\phi}_{\bar{x}}^f(\lambda) = -\eta_{\bar{x}}^f(\lambda) = \begin{cases} -f(P_\lambda f(\bar{x})) & \text{if } \lambda > 0,\\ -f(P_{\operatorname{cl}(\operatorname{dom} f)}(\bar{x})) & \text{if } \lambda \le 0.\end{cases}$$

If, more strictly, $\bar{x} \in \operatorname{dom} \partial f$, then this derivative is locally Lipschitz on all of $\mathbb{R}$.

(c) *If $P_{\operatorname{cl}(\operatorname{dom} f)}(\bar{x}) \notin \operatorname{dom} f$, then $\operatorname{dom} \bar{\phi}_{\bar{x}}^f = \mathbb{R}_+$ and*

$$\partial\bar{\phi}_{\bar{x}}^f(\lambda) = \begin{cases} -f(P_\lambda f(\bar{x})) & \text{if } \lambda > 0,\\ \emptyset & \text{if } \lambda \le 0.\end{cases}$$

Proof. Set $\bar{\phi} := \bar{\phi}_{\bar{x}}^f(\lambda)$. Part (a). It is an easy computation to see that

$$0 < \lambda \mapsto -\bar{\phi}(\lambda) = \inf_u \left\{\lambda f(y) + \tfrac{1}{2}\|u - \bar{x}\|^2\right\}$$

is concave, i.e., $0 < \lambda \mapsto \bar{\phi}(\lambda)$ is convex. By setting $\bar{\phi}(0) = -\frac{1}{2} d^2_{\mathrm{cl}\,(\mathrm{dom}\, f)}(\bar{x})$ and using Proposition 4(a), we see that $\bar{\phi}$ is a continuous convex function on $\mathbb{R}_+$, which is linearly extended to $\mathbb{R}_-$. All in all, $\bar{\phi}$ is convex, proper and continuous (possibly in an extended real-valued) sense. From Corollary 6(b) (and the product rule) we infer, for all $\lambda > 0$, that

$$\bar{\phi}'(\lambda) = -e_\lambda f(\bar{x}) - \lambda \left(-\frac{1}{2} \left\| \frac{1}{\lambda}[\bar{x} - P_\lambda f(\bar{x})] \right\|^2 \right) = (1/2\lambda)\|\bar{x} - P_\lambda f(\bar{x})\|^2 - e_\lambda f(\bar{x}) = -f(P_\lambda f(\bar{x})),$$

where the last equality follows from (25). Hence, the local Lipschitz continuity follows from Corollary 11(b).

Part (b). Here we assume that $P_{\mathrm{cl}\,(\mathrm{dom}\, f}(\bar{x}) \in \mathrm{dom}\, f$. Then by definition of $\bar{\phi}$, we have $\bar{\phi}'(\lambda) = -f(P_{\mathrm{cl}\,(\mathrm{dom}\, f)}(\bar{x}))$ for all $\lambda < 0$. It remains to establish the case $\lambda = 0$. To this end, use the subgradient inequality to deduce that $g \in \partial \bar{\phi}(0)$ if and only if $\bar{\phi}(0) + \lambda g \le \bar{\phi}(\lambda)$ for all λ if and only if

$$g + e_\lambda f(\bar{x}) - \frac{1}{2\lambda} d^2_{\mathrm{cl}\,(\mathrm{dom}\, f)}(\bar{x}) \le 0 \quad \forall \lambda > 0, \tag{28a}$$

$$\lambda g + \lambda f(P_{\mathrm{cl}\,(\mathrm{dom}\, f)}(\bar{x})) \le 0 \quad \forall \lambda < 0, \tag{28b}$$

hold simultaneously. (The case with $\lambda = 0$ holds trivially.) From (28a), we infer that

$$\begin{aligned} g \le \inf_{\lambda > 0} \frac{\frac{1}{2} d^2_{\mathrm{cl}\,(\mathrm{dom}\, f)}(\bar{x}) - \lambda e_\lambda f(\bar{x})}{\lambda} &\overset{\text{(i)}}{=} \inf_{\lambda > 0} \frac{\bar{\phi}(\lambda) - \bar{\phi}(0)}{\lambda} \\ &\overset{\text{(ii)}}{=} \lim_{\lambda \downarrow 0} \frac{\bar{\phi}(\lambda) - \bar{\phi}(0)}{\lambda} \\ &\overset{\text{(iii)}}{=} \lim_{\lambda \downarrow 0} \frac{\frac{1}{2} d^2_{\mathrm{cl}\,(\mathrm{dom}\, f)}(\bar{x}) - \lambda e_\lambda f(\bar{x})}{\lambda} \\ &\overset{\text{(iv)}}{=} \lim_{\lambda \downarrow 0} \frac{-f(P_\lambda f(\bar{x}))}{1} \\ &\overset{\text{(v)}}{=} -f(P_{\mathrm{cl}\,(\mathrm{dom}\, f)}(\bar{x})). \end{aligned}$$

Here, (i) is simply the definition of $\bar{\phi}$; (ii) holds because $\bar{\phi}$ is convex [38, Theorem 23.1]; (iii) follows from the definition of $\bar{\phi}$; and (iv) follows from l'Hôpital's rule, which is applicable because the last limit exists by Corollary 10(c), which implies (v). Hence, (28a) is equivalent to $g \le -f(P_{\mathrm{cl}\,(\mathrm{dom}\, f}(\bar{x}))$. Combined with (28b), which is equivalent to $g \ge -f(P_{\mathrm{cl}\,(\mathrm{dom}\, f}(\bar{x}))$, establishes that $\partial \bar{\phi}(0) = \{-f(P_{\mathrm{cl}\,(\mathrm{dom}\, f}(\bar{x}))\}$. Thus, $P_{\mathrm{cl}\,(\mathrm{dom}\, f}(\bar{x}) \in \mathrm{dom}\, f$, $\bar{\phi}$ is differentiable, and hence continuously differentiable by convexity [38, Corollary 25.5.1]. The remainder follows from Corollary 11(c).

Part (c). Here we assume that $P_{\mathrm{cl}\,(\mathrm{dom}\, f}(\bar{x}) \notin \mathrm{dom}\, f$. Suppose $g \in \partial \phi(0)$, i.e., analogous to some arguments in b),

$$g \le (1/2\lambda) d^2_{\mathrm{cl}\,(\mathrm{dom}\, f)}(\bar{x}) - e_\lambda f(\bar{x}) \quad \forall \lambda > 0.$$

On the other hand, using e.g., Corollary 10(b), we have

$$\begin{aligned} (1/2\lambda) d^2_{\mathrm{cl}\,(\mathrm{dom}\, f)}(\bar{x}) - e_\lambda f(\bar{x}) &= (1/2\lambda)\|\bar{x} - P_{\mathrm{cl}\,(\mathrm{dom}\, f)}(\bar{x})\|^2 - (1/2\lambda)\|\bar{x} - P_\lambda f(\bar{x})\|^2 - f(P_\lambda f(\bar{x})) \\ &\le -f(P_\lambda f(\bar{x})). \end{aligned}$$

Since $-f(P_\lambda f(\bar{x})) \to -\infty$ as $\lambda \downarrow 0$, this concludes the proof. □

5.1. Semismoothness of the proximal value function In view of the properties of the proximal value function, as outlined by Corollary 11, the question for *semismoothness* of $\eta^f_{\bar{x}}$ on $\mathbb{R}_{++}$ arises naturally. Now consider the expression (25). The map $0 < \lambda \mapsto e_\lambda f(\bar{x})$ is continuously differentiable by Corollary 6(a), hence semismooth [20, Proposition 7.4.5]. Moreover, the map $0 < \lambda \mapsto (1/2\lambda)\|\bar{x} - P_\lambda f(\bar{x})\|^2$ is semismooth if $0 < \lambda \mapsto P_\lambda f(\bar{x})$ is semismooth [20, Proposition 7.4.4]. Thus, when the latter holds, we can conclude that $\eta^f_{\bar{x}}$ is semismooth. We can in addition use Corollary 9, which establishes conditions for the semismoothness of the map $(x, \lambda) \in \mathbb{E} \times \mathbb{R}_{++} \mapsto P_\lambda f(x)$, to obtain the following result.

Proposition 7 **(Semismoothness of the proximal value function)**. *Let $f \in \Gamma_0(\mathbb{E})$ and $\bar{x} \in \mathbb{E}$. Then $\eta_{\bar{x}}^f$ is semismooth at $\bar{\lambda} > 0$ if ∂f is proto-differentiable and semismooth* at $\big(P_{\bar{\lambda}} f(\bar{x}), \frac{1}{\bar{\lambda}}[\bar{x} - P_{\bar{\lambda}} f(\bar{x})]\big)$. This is the case under either of the following conditions:*

(a) *(PLQ case) f is piecewise-linear quadratic.*

(b) *(C^2 case) f is twice continuously differentiable around $P_{\bar{\lambda}} f(\bar{x})$. In this case, $\eta_{\bar{x}}^f$ is continuously differentiable.*

6. Post-composition envelopes and proximal maps

Given functions $\psi \in \Gamma_0(\mathbb{E})$ and $g \in \Gamma_0(\mathbb{R})$, we consider the composition

$$(g \circ \psi)(x) := \begin{cases} g(\psi(x)) & \text{if } x \in \operatorname{dom} \psi, \\ +\infty & \text{otherwise.} \end{cases}$$

It is well known that $g \circ \psi$ is closed proper convex if g is increasing and that the intersection $\psi(\mathbb{E}) \cap \operatorname{dom} g$ is nonempty; see, for example, Hiriart-Urruty and Lemaréchal [25, Theorem B.2.1.7], who describe this operation as *post-composition*. We establish variational formulas for the Moreau envelope and proximal map of the composition $g \circ \psi$ under a regularity assumption involving the intersection of domains. These results provide us with tools to infer properties of projections onto the epigraph and level sets of a closed proper convex function, as covered in Section 7.

Proposition 8 **(Post-composition, Moreau envelopes, and proximal maps)**. *Let $g \in \Gamma_0(\mathbb{R})$ be increasing and let $\psi \in \Gamma_0(\mathbb{E})$ such that*

$$(\operatorname{ri} \operatorname{dom} g) \cap \psi(\operatorname{ri} \operatorname{dom} \psi) \neq \emptyset. \tag{29}$$

Then the following properties hold.

(a) *$e_1(g \circ \psi)(\bar{x}) = -\min_{\lambda \geq 0} \big\{ g^*(\lambda) + \bar{\phi}_{\bar{x}}^{\psi}(\lambda) \big\}$, where $\bar{\phi}_{\bar{x}}^{\psi}$ is given by (27).*

(b) *$P_1(g \circ \psi)(\bar{x}) = P_1(\bar{\lambda} \cdot \psi)(\bar{x})$ for every $\bar{\lambda} \in \operatorname{argmin}_{\lambda \geq 0} \big\{ g^*(\lambda) + \bar{\phi}_{\bar{x}}^{\psi}(\lambda) \big\} \neq \emptyset$.*

(c) *If $\psi(P_{\operatorname{cl}(\operatorname{dom}\psi)}(\bar{x})) \notin \partial g^*(0)$, then $\operatorname{argmin}_{\lambda \geq 0} \big\{ g^*(\lambda) + \bar{\phi}_{\bar{x}}^{\psi}(\lambda) \big\} \subset \mathbb{R}_{++}$. This is, in particular, the case if $P_{\operatorname{cl}(\operatorname{dom}\psi)}(\bar{x}) \notin \operatorname{dom} \psi$.*

Proof. Part (a). We find that

$$\begin{aligned} e_1(g \circ \psi)(\bar{x}) &= \min_{x \in \mathbb{E}} \big\{ \tfrac{1}{2}\|x - \bar{x}\|^2 + (g \circ \psi)(x) \big\} \\ &= -\big(\tfrac{1}{2}\|(\cdot) - \bar{x}\|^2 + g \circ \psi \big)^* (0) \\ &= \max_{y \in \mathbb{E}, \lambda \geq 0} -\big\{ g^*(\lambda) - \tfrac{1}{2}\|y\|^2 + \langle \bar{x}, y \rangle + (\lambda \cdot \psi)^*(-y) \big\} \\ &= \max_{\lambda \geq 0} \big\{ -g^*(\lambda) + \max_{y \in \mathbb{E}} \big[-\tfrac{1}{2}\|y\|^2 - \langle \bar{x}, y \rangle - (\lambda \cdot \psi)^*(-y) \big] \big\} \\ &= \max_{\lambda \geq 0} \ -g^*(\lambda) - \bar{\phi}_{\bar{x}}^{\psi}(\lambda). \end{aligned}$$

Here, the third identity uses [10, Corollary 3] with $f := \frac{1}{2}\|(\cdot) - \bar{x}\|^2$, $F := \psi$, and $K = \mathbb{R}_+$, realizing that (29) is equivalent to qualification condition [10, Equation (17)] because $\operatorname{dom} g - \mathbb{R}_+ = \operatorname{dom} g$, and observing that attainment is guaranteed by finiteness of the left-hand side. The last identity uses Fenchel duality [38, Theorem 31.1] and the definition of $\bar{\phi}_{\bar{x}}^{\psi}$ in (27).

Part (b). Note that by [10, Corollary 4],

$$\partial(g \circ \psi)(x) = \bigcup_{\lambda \in \partial g(\psi(x))} \partial(\lambda \cdot \psi)(x) \quad \forall x \in \operatorname{dom} g \circ \psi, \tag{30}$$

and observe that $\partial g(x)\subset\mathbb{R}_+$ because g is increasing. Next, observe that

$$\begin{aligned}\bar\lambda\in\operatorname{argmin}_{\lambda\geq 0}\{g^*(\lambda)+\bar\phi^{\psi}_{\bar x}(\lambda)\},\ \bar u=P_1(\bar\lambda\cdot\psi)(\bar x) &\overset{\text{(i)}}{\iff} 0\in\partial g^*(\bar\lambda)+\partial\bar\phi^{\psi}_{\bar x}(\bar\lambda),\ \bar u=P_1(\bar\lambda\cdot\psi)(\bar x)\\ &\overset{\text{(ii)}}{\iff} \psi(\bar u)\in\partial g^*(\bar\lambda),\ \bar u=P_1(\bar\lambda\cdot\psi)(\bar u)\\ &\overset{\text{(iii)}}{\iff} \bar\lambda\in\partial g(\psi(\bar u)),\ \bar u=P_1(\bar\lambda\cdot\psi)(\bar x)\\ &\overset{\text{(iv)}}{\iff} \bar\lambda\in\partial g(\psi(\bar u)),\ 0\in\bar u-\bar x+\partial(\bar\lambda\cdot\psi)(\bar x)\\ &\overset{\text{(v)}}{\implies} \bar u=P_1(g\circ\psi)(\bar x).\end{aligned}$$

Equivalence (i) is valid because $\operatorname{int}(\operatorname{dom} g^*)\subset\mathbb{R}_{++}\subset\operatorname{int}(\operatorname{dom}\bar\phi^{\psi}_{\bar x})$; see [10, Lemma 4] and Corollary 12, respectively. Corollary 12(b) justifies equivalence (ii). Equivalence (iii) is the inversion formula for the subdifferential [38, Corollary 23.5.1]. Equivalence (iv) uses the optimality conditions that uniquely determines $\bar u=P_1(\bar\lambda\cdot\psi)(\bar x)$. Implication (v) follows from (30) and the optimality conditions that uniquely determine $P_1(g\circ\psi)(\bar x)$. Taken together, we deduce that for any $\bar\lambda\in\operatorname{argmin}_{\lambda\geq0} g^*(\lambda)+\bar\phi^{\psi}_{\bar x}(\lambda)$, we have $P_1(g\circ\psi)(\bar x)=P_1(\lambda\cdot\psi)(\bar x)$. The fact that $\operatorname{argmin}_{\lambda\geq0}\left\{g^*(\lambda)+\bar\phi^{\psi}_{\bar x}(\lambda)\right\}\neq\emptyset$ follows from Part (a).

Part (c). Recall from Part (b) that $0\in\operatorname{argmin}_{\lambda\geq0}\{g^*+\phi^{\psi}_{\bar x}\}$ entails $0\in\partial g^*(0)+\partial\phi^{f}_{\bar x}(0)$. In view of Corollary 12(c), we must have $P_{\operatorname{cl}(\operatorname{dom}\psi)}(\bar x)\in\operatorname{dom}\psi$, in which case $\partial\phi^{\psi}_{\bar x}(0)=-\psi(P_{\operatorname{cl}(\operatorname{dom}\psi)}(\bar x))$, by Corollary 12(b). This proves the claim. ☐

7. Epigraphical and level-set projections

We are now equipped to answer the initial question about computing epigraphical and level-set projections via proximal mappings. Our approach is based on the Moreau envelopes of the indicator functions to the epigraph and level set of a function f, which we express as the post-compositions

$$\delta_{\operatorname{lev}_\alpha f}=(\delta_{\mathbb{R}_-})\circ(f(\cdot)-\alpha)\qquad\text{and}\qquad\delta_{\operatorname{epi} f}=(\delta_{\mathbb{R}_-})\circ(f(\cdot)-(\cdot)).$$

Proposition 8 provides the required tools.

Corollary 13 **(Level-set projection)**. *Let $f\in\Gamma_0(\mathbb{E})$, $(\bar x,\bar\alpha)\in\mathbb{E}\times\mathbb{R}$, and assume there exists $\hat x\in\mathbb{E}$ such that $f(\hat x)<\bar\alpha$. Then the following statements hold.*

(a) *(Dual representation of distance to level set)*

$$\tfrac12 d^2_{\operatorname{lev}_{\bar\alpha} f}(\bar x)=-\min_{\lambda\geq0}\left\{\bar\phi^{f}_{\bar x}(\lambda)+\bar\alpha\lambda\right\}.$$

(b) *(Projection onto level set)*

$$P_{\operatorname{lev}_{\bar\alpha} f}(\bar x)=\begin{cases}P_{\operatorname{cl}(\operatorname{dom} f)}(\bar x) & \text{if } f(P_{\operatorname{cl}(\operatorname{dom} f)}(\bar x))\leq\bar\alpha,\\ P_{\bar\lambda}f(\bar x) & \text{otherwise,}\end{cases}$$

for any positive $\bar\lambda$ in the optimal solution set

$$\operatorname*{argmin}_{\lambda\geq0}\ \{\bar\phi^{f}_{\bar x}(\lambda)+\bar\alpha\lambda\}=\{\lambda\geq0\mid f(P_\lambda f(\bar x))=\bar\alpha\}\neq\emptyset.$$

Proof. Set $g:=\delta_{\mathbb{R}_-}$ and $\psi: x\in\mathbb{E}\mapsto f(x)-\bar\alpha$. Then $g\in\Gamma_0(\mathbb{R})$ is increasing and $\psi\in\Gamma_0(\mathbb{E})$ with $\operatorname{dom}\psi=\operatorname{dom} f$ and $\delta_{\operatorname{lev}_{\bar\alpha} f}=g\circ\psi$. Now observe that (29) applied to this setting is equivalent to saying that there exists $\bar y\in\operatorname{ri}(\operatorname{dom} f)$ such that $f(\bar y)<\bar\alpha$. We (only) assume that there exists $\hat x\in\operatorname{dom} f$ such that $f(\hat x)<\bar\alpha$. However, take any $z\in\operatorname{ri}(\operatorname{dom} f)$, then, by the line segment principle [38, Theorem 6.1], we have $y_\lambda:=\lambda z+(1-\lambda)\hat x\in\operatorname{ri}(\operatorname{dom} f)$ for all $\lambda\in(0,1]$. Moreover, $f(y_\lambda)<\lambda f(z)+(1-\lambda)\bar\alpha\to\bar\alpha$ as $\lambda\downarrow0$. Hence there exists $\hat\lambda\in(0,1]$ sufficiently small such that $f(y_{\hat\lambda})<\bar\alpha$. Hence $\hat y:=y_{\hat\lambda}\in\operatorname{ri}(\operatorname{dom} f)$ with $f(\hat y)<\bar\alpha$, and (29) holds.

Part (a). For all $\lambda \geq 0$,

$$\begin{aligned}\bar{\phi}^{\psi}_{\bar{x}}(\lambda) &= \begin{cases} -\lambda e_\lambda \psi(\bar{x}) & \text{if } \lambda > 0, \\ -\frac{1}{2} d^2_{\mathrm{cl}\,(\mathrm{dom}\,\psi)}(\bar{x}) & \text{if } \lambda = 0, \end{cases} \\ &= \begin{cases} -\lambda (e_\lambda f(\bar{x}) - \bar{\alpha}) & \text{if } \lambda > 0, \\ -\frac{1}{2} d^2_{\mathrm{cl}\,(\mathrm{dom}\,f)}(\bar{x}) & \text{if } \lambda = 0, \end{cases} \\ &= \bar{\phi}^{f}_{\bar{x}}(\lambda) + \bar{\alpha}\lambda.\end{aligned}$$

Use Proposition 8(a) and the fact that $g^* = \delta_{\mathbb{R}_+}$ to deduce that

$$\tfrac{1}{2} d^2_{\mathrm{lev}_{\bar{\alpha}} f}(\bar{x}) = e_1 \delta_{\mathrm{lev}_{\bar{\alpha}} f}(\bar{x}) = e_1 (g \circ \psi)(\bar{x}) = -\min_{\lambda \geq 0} \left\{ \bar{\phi}^{f}_{\bar{x}}(\lambda) + \bar{\alpha}\lambda \right\}.$$

Part (b). The equality of the two sets in question is clear from the (necessary and sufficient) optimality conditions and Corollary 12. The rest follows from Proposition 8, Parts (b) and (c) because $P_{\mathrm{lev}_{\bar{\alpha}} f}(\bar{x}) = P_1 (g \circ \psi)(\bar{x})$. ☐

Corollary 14 **(Epigraphical projection)**. *Let $f \in \Gamma_0(\mathbb{E})$ and $(\bar{x}, \bar{\alpha}) \in \mathbb{E} \times \mathbb{R}$. Then the following statements hold.*

(a) *(Dual representation of distance to epigraph)*

$$\tfrac{1}{2} d^2_{\mathrm{epi}\, f}(\bar{x}, \bar{\alpha}) = -\min_{\lambda \geq 0} \left\{ \bar{\phi}^{f}_{\bar{x}}(\lambda) + \bar{\alpha}\lambda + \tfrac{1}{2}\lambda^2 \right\}.$$

(b) *(Projection onto epigraph)*

$$P_{\mathrm{epi}\, f}(\bar{x}) = \begin{cases} [P_{\mathrm{cl}\,(\mathrm{dom}\,f)}(\bar{x}),\ \bar{\alpha}] & \text{if } f(P_{\mathrm{cl}\,(\mathrm{dom}\,f)}(\bar{x})) \leq \bar{\alpha}, \\ [P_{\bar{\lambda}} f(\bar{x}),\ \bar{\alpha} + \bar{\lambda}] & \text{otherwise}, \end{cases}$$

where $\bar{\lambda} > 0$ is the unique solution of the strongly convex optimization problem

$$\min_{\lambda \geq 0}\ \tfrac{1}{2}\lambda^2 + \bar{\alpha}\lambda + \bar{\phi}^{f}_{\bar{x}}(\lambda).$$

Equivalently, λ is the unique root of the strictly decreasing function $0 < \lambda \mapsto f(P_\lambda f(\bar{x})) - \lambda - \bar{\alpha}$.

Proof. Analogous to the proof of Corollary 13, we define closed proper convex functions $g := \delta_{\mathbb{R}_-}$ and $\psi : (x, \alpha) \in \mathbb{E} \times \mathbb{R} \mapsto f(x) - \alpha$ so that $\delta_{\mathrm{epi}\, f} = g \circ \psi$. Therefore,

$$\psi(\mathrm{ri}\,(\mathrm{dom}\,\psi)) = \psi(\mathrm{ri}\,(\mathrm{dom}\,f) \times \mathbb{R}) = f(\mathrm{ri}\,(\mathrm{dom}\,f)) - \mathbb{R} = \mathbb{R},$$

and thus the qualification condition (29) is trivially satisfied in this setting.

Part (a). Note that $e_\lambda \psi(x, \alpha) = e_\lambda f(x) + e_\lambda(-\mathrm{id}\,)(\alpha)$ for all $\lambda > 0$ [4, Theorem 6.58], and since $\mathrm{dom}\,\psi = \mathrm{dom}\,f$,

$$\bar{\phi}^{\psi}_{\bar{x},\bar{\alpha}}(\lambda) = \bar{\phi}^{f}_{\bar{x}}(\lambda) + \bar{\alpha} \cdot \lambda + \tfrac{1}{2}\lambda^2 \quad (\lambda \geq 0).$$

Apply Proposition 8(a) to obtain the desired result.

Part (b). Apply Proposition 8(b), observing that $P_{\mathrm{epi}\, f}(\bar{x}, \bar{\alpha}) = P_1 \delta_{\mathrm{epi}\, f}(\bar{x}, \bar{\alpha})$ and $P_1(\lambda \cdot \psi)(\bar{x}, \bar{\alpha}) = [P_1(\lambda f)(\bar{x}), \bar{\alpha} + \lambda]$ for all $\lambda \geq 0$ [4, Theorem 6.6]. The fact that $\bar{\lambda} > 0$ is due to Proposition 8(c). ☐

Remark 3 (Prior work). The level-set projection result Corollary 13 encompasses the result described by Beck [4, Theorem 6.30]. For epigraphical projection, Corollary 14 generalizes Beck [4, Theorem 6.36] to include functions that aren't finite-valued. For functions $f \in \Gamma_0(\mathbb{E})$ with open domain, Chierchia et al. [11, Proposition 1] describe an alternative formula for epigraphical projections via proximal maps.

Algorithm 1 SC^1 Newton method for minimizing θ_ξ

(S.0) Choose $\lambda_0, \delta > 0$, $\{\varepsilon_k\} \downarrow 0$, and let $\beta, \sigma \in (0,1)$. Set $k := 0$.
(S.1) If $|\theta'(\lambda_k)| \le \delta$: STOP.
(S.2) Choose $g_k \in \partial_B(\theta'_\xi)(\lambda_k)$ and set

$$\Delta_k := P_{[-\lambda_k,\infty)}\left(-\frac{\theta'_\xi(\lambda_k)}{g_k + \varepsilon_k}\right).$$

(S.3) Set

$$t_k := \max_{l \in \mathbb{N}_0} \left\{\beta^l \;\middle|\; \theta_\xi(\lambda_k + \beta^l \Delta_k) \le \theta_\xi(\lambda_k) + \beta^l \sigma \theta'_\xi(\lambda)\Delta_k \right\}.$$

(S.4) Set $\lambda_{k+1} := \lambda_k + t_k \Delta_k$, $k \leftarrow k+1$, and go to (S.1).

7.1. An SC^1 optimization framework In this section we present a unified algorithmic framework for computing projections onto the level sets and the epigraph of a closed proper convex function. Corollaries 13 and 14, respectively, guide us in how to compute these projections. For a given $f \in \Gamma_0(\mathbb{E})$ and $(\bar{x}, \bar{\alpha}) \in \mathbb{E} \times \mathbb{R}$ such that $f(\bar{x}) > \bar{\alpha}$, the epigraphical and level-set projections, respectively, correspond to the proximal map of f with parameter λ that solves the scalar problem

$$\min_{\lambda \ge 0} \; \theta_\xi(\lambda) \quad (\xi \in \{\mathrm{epi}, \mathrm{lev}\}), \tag{31}$$

for $\theta_\xi : \mathbb{R} \to \overline{\mathbb{R}}$ given by

$$\theta_\xi(\lambda) = \begin{cases} \bar{\phi}^f_{\bar{x}}(\lambda) + \bar{\alpha}\lambda & \text{if } \xi = \mathrm{lev}, \\ \bar{\phi}^f_{\bar{x}}(\lambda) + \bar{\alpha}\lambda + \frac{1}{2}\lambda^2 & \text{if } \xi = \mathrm{epi}. \end{cases} \tag{32}$$

Corollary 12 asserts that θ_ξ is convex, continuous (possibly in an extended real-valued sense), and continuously differentiable with monotonically increasing, locally Lipschitz derivative on $\mathbb{R}_{++}$. In particular, for any $\lambda > 0$,

$$\theta'_\xi(\lambda) = \begin{cases} -\eta^f_{\bar{x}}(\lambda) + \bar{\alpha} & \text{if } \xi = \mathrm{lev}, \\ -\eta^f_{\bar{x}}(\lambda) + \bar{\alpha} + \lambda & \text{if } \xi = \mathrm{epi}, \end{cases} \tag{33}$$

The minimization of ϕ_η could be accomplished using bisection if an upper bound on the optimal λ is available. However, the semismoothness of the derivative (33), described by Proposition 7, allows us to tap into the powerful SC^1 optimization framework [20, 36] that operates on functions $\theta : \mathbb{R} \to \overline{\mathbb{R}}$ that are *semismoothly differentiable* (i.e., SC^1), which means that at points $\bar{\lambda} \in \operatorname{int}(\operatorname{dom}\theta)$, the gradient θ' exists, and it is locally Lipschitz around $\bar{\lambda}$ and semismooth at $\bar{\lambda}$. The semismooth method, outlined by Algorithm 1, applies to the problem (31) whenever conditions (A1) and (A2) of Pang and Qi [36] hold, which is the case when $\bar{x} \in \operatorname{dom}\partial f$; see Corollary 12.

Algorithm 1 uses the notion of a *Bouligand subdifferential*, which for a function $\phi : \mathbb{R}^n \to \overline{\mathbb{R}}$ that is locally Lipschitz at a point $\bar{x} \in \operatorname{int}(\operatorname{dom}\phi)$, is defined at $\bar{x}$ as $\partial_B\phi(\bar{x}) = \{v \mid \exists\{x_k \in D_\phi,\ x_k \to \bar{x}\} : \nabla\phi(x_k) \to v\}$, where D_ϕ is the set of points of differentiability of ϕ. The *Clarke subdifferential* [12] of ϕ at $\bar{x}$ is $\partial_C\phi(\bar{x}) := \operatorname{conv}\partial_B\phi(\bar{x})$, which coincides (on the interior of $\operatorname{dom}\phi$) with the convex subdifferential if ϕ is convex.

Remark 4. Because θ_ξ is convex and differentiable with locally Lipschitz derivative on $\mathbb{R}_{++}$, all elements in the Clarke subdifferential $\partial_C(\theta'_\xi)(\lambda)$ are nonnegative for all $\lambda > 0$ [20]. In the epigraphical case (i.e., $\xi = \mathrm{epi}$), the quadratic term in the expression for θ_{epi} in (32) implies that the elements are bounded below by 1. Thus, the sequence of regularization parameters $\{\varepsilon_k\} \downarrow 0$ in Algorithm 1 is not necessary, and in fact, if θ' is piecewise affine, the regularization could be eliminated by setting the constant regularization $\varepsilon_k := 0$ for all k, which would improve numerical convergence regardless of the optimality parameter $\delta > 0$.

Algorithm 2 Full-step SC^1 Newton method

(S.0) Choose $\lambda_0 > 0$, $\delta > 0$, and $\{\varepsilon_k\} \downarrow 0$. Set $k := 0$.
(S.1) If $|\theta'_\xi(\lambda_k)| \le \delta$: STOP.
(S.2) Choose $g_k \in \partial_C(\theta'_\xi)(\lambda_k)$ and set

$$\Delta_k := \max\left\{\frac{-\lambda_k}{2},\ \frac{-\theta'_\xi(\lambda_k)}{g_k+\varepsilon_k}\right\}.$$

(S.3) Set $\lambda_{k+1} := \lambda_k + \Delta_k$, $k \leftarrow k+1$, and go to (S.1).

7.1.1. The case where θ'_ξ is concave on $(0, \lambda_l)$ Corollaries 13 and 14 imply that there exists positive parameters $\lambda_l \le \lambda_u$ such that

$$[\lambda_l, \lambda_u] = \operatorname*{argmin}_{\lambda \ge 0} \theta_\xi = \left\{\lambda > 0 \;\middle|\; \theta'_\xi(\lambda) = 0\right\}, \tag{34}$$

for both the epigraphical and level-set cases. In the epigraphical case in particular, the solution is unique, and thus $\lambda_u = \lambda_l$; see Corollary 14(b). If the derivative ϕ_ξ is concave on the interval $(0, \lambda_\ell)$, it is possible to take a full Newton step at every iteration while respecting positivity of the iterates, thus saving the computational cost of a backtracking line-search. The simplified iteration is described by Algorithm 2.

For many important functions, e.g., the 1-norm or negative log, (and their spectral counterparts), the respective map θ'_ξ is concave on $\mathbb{R}_{++}$, but, as suggested above, we only need the following:

Assumption 1 **(Concavity $(0, \lambda_l)$)**. *The function θ'_ξ is concave on $(0, \lambda_l)$.*

Proposition 9 **(Convergence of Algorithm 2)**. *Under Assumption 1, the full-step Newton method from Algorithm 2 converges to a minimizer of θ_ξ.*

Proof. Set $\theta = \theta_\xi$. If $0 < \lambda_k < \lambda_l$ for some $k \in \mathbb{N}$, then by Corollary 11(a), $\theta'(\lambda_k) < 0$ by monotonicity of $-\theta'$. Therefore,

$$\lambda_{k+1} = \lambda_k - \frac{\theta'(\lambda_k)}{g_k + \varepsilon_k} > \lambda_k.$$

Since $-(g_k + \varepsilon_k)$ is a convex subgradient of $-(\theta' + \varepsilon_k(\cdot))$, the concavity of θ'_ξ implies that

$$-\theta'(\lambda_{k+1}) - \varepsilon_k(\lambda_{k+1} - \lambda_k) \ge -\theta'(\lambda_k) - (\lambda_{k+1} - \lambda_k)(g_k + \varepsilon_k) = 0,$$

and hence $\theta'(\lambda_{k+1}) < 0$, thus $0 < \lambda_k < \lambda_{k+1} < \lambda_l$. Consequently, by an inductive argument, $\{\lambda_k\}$ converges to some $\tilde\lambda$. Therefore, the sequence $\{g_k \in \partial_C(\theta_\xi)(\lambda_k)\}$ is bounded, and hence

$$0 = (\lambda_{k+1} - \lambda_k)(g_k + \varepsilon_k) + \theta'(\lambda_k) \to \theta'(\tilde\lambda),$$

which shows that $\tilde\lambda$ has the desired properties. We hence still need to cover the case where $\lambda_l < \lambda_k$ for all $k \in \mathbb{N}$. In view of (34), we can assume that $\lambda_u < \lambda_k$ for all $k \in \mathbb{N}$. (Otherwise, a solution has already been obtained.) Since $\theta'(\lambda_k) > 0$ here, we observe that

$$0 < \lambda_u < \lambda_{k+1} = \lambda_k + \max\left\{\frac{-\lambda_k}{2},\ \frac{-\theta'(\lambda_k)}{g_k + \varepsilon_k}\right\} \le \lambda_k,$$

hence the sequence $\{\lambda_k\}$ converges to some $\hat\lambda$. In particular, $\lambda_{k+1} = \frac{1}{2}\lambda_k$ only finitely many times. Hence, without loss of generality, $0 = (\lambda_{k+1} - \lambda_k)(g_k + \varepsilon_k) + \theta'(\lambda_k) \to \theta'(\hat\lambda)$, which gives $\theta'(\hat\lambda) = 0$ also here. ◻

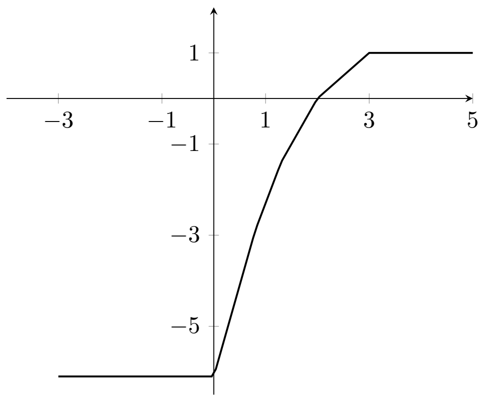


FIGURE 2. The function θ'_{epi} corresponding to the projection of point $\bar{x} = (-2, 0.8, 3, 1.3)$ onto the 1-norm unit ball.

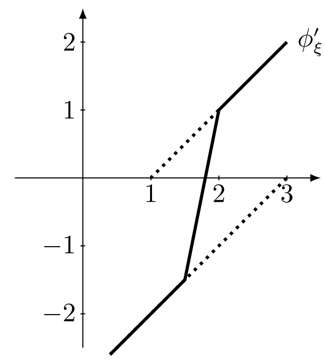


FIGURE 3. The function θ'_{epi} for Example 3, for which Algorithm 2 may cycle.

The next example illustrates that cycling may occur in Algorithm 2 if Assumption 1 fails.

EXAMPLE 3 (CYCLING). Consider the scalar function $f(x) = 2|x| + \delta_{[-1,1]}(x)$, and the task of projecting the $(\bar{x}, \bar{\alpha}) = (4, -1)$ onto epi f. Figure 3 illustrates the function θ'_{epi} whose root we seek. Then for λ_0 outside of the interval $[1.5, 2]$ the iterates λ_k $(k \in \mathbb{N})$ generated by Algorithm 2 oscillate between 1.5 and 3.

7.2. Numerical Experiments We present numerical experiments that hint at the computational effectiveness of the SC^1 optimization framework described in Section 7.1. The two experiments in this section were run on an Apple Macbook Air with a 1.8GHz Intel Core i5 and 8Gb RAM running OS 10.14.6. The code was written in C and available at `https://github.com/arielgoodwin/epi-proj`.

7.2.1. Level-set projection: 1-norm An important instance of the level-set case $(\xi = \text{lev})$ is the projection onto the unit 1-norm ball $\text{lev}_1\|\cdot\|_1 = \{x \in \mathbb{R}^n \mid \|x\|_1 \le 1\}$. The derivative of the corresponding function θ_{lev} reads

$$\theta'_{\text{lev}}(\lambda) = \begin{cases} 1 - \sum_{i=1}^n \max\{|x_i| - \lambda, 0\} & \text{if } \lambda \ge 0, \\ 1 - ||x||_1 & \text{if } \lambda < 0, \end{cases}$$

which is concave on $\mathbb{R}_+$ (as required) and piecewise affine, as shown by Fig. 2.

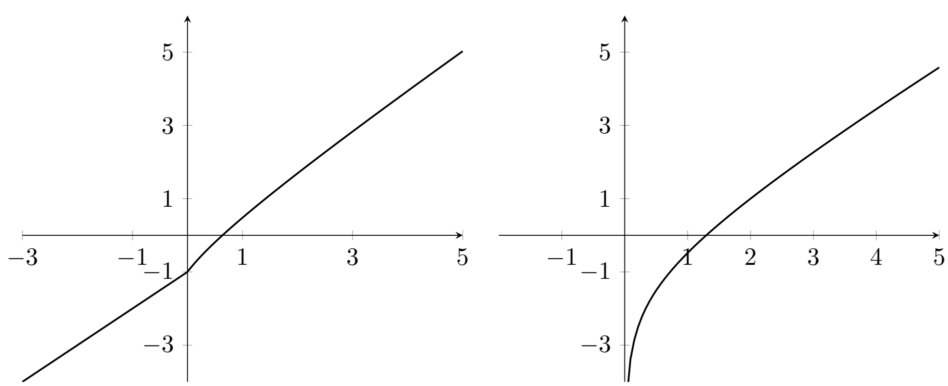
We implemented Algorithm 2 and compared it numerically to two state-of-the-art algorithms specifically tailored to 1-norm-ball projection, namely Condat's sorting-based method [16] as implemented in the code `condat_l1ballproject.c`, and Liu and Ye's improved bisection algorithm (IBIS) [30] implemented in the `eplb` module in SLEP [41].

The entries of the projected vectors $\bar{x} \in \mathbb{R}^n$ are drawn from a Gaussian distribution with zero mean and standard deviations $\sigma = \{0.1, 0.05, 0.01, 0.005\}$. The optimality tolerance was fixed at $\delta = 10^{-15}$, as in step (S.1) of Algorithm 2. Table 1 reports the average time required to compute the projection over 10^5 trials for vectors of dimension $n \in \{20, 10^3\}$, and over 500 trials for $n = 10^6$. The initial point $\lambda_0 > 0$ Algorithm 2 was chosen by sampling $\sqrt{n}\log n$ coordinates randomly from the vector $\bar{x}$ and setting λ_0 to be the largest of their absolute values. Observe that Algorithm 2 exhibits comparable performance relative to the specialized algorithms.

7.2.2. Level-set projection: negative sum-log We now consider the epigraphical projection for a function that is not polyhedral. Define the function $f : x \in \mathbb{R}^n \mapsto -\sum_{i=1}^n \log x_i$, where we take the negative logarithm to be ∞ outside the positive orthant. Figure 4 illustrate the function

n	Algorithm 2	Condat	IBIS	Algorithm 2	Condat	IBIS
	$\sigma = 0.1$			$\sigma = 0.05$		
20	1.94×10^{-6}	1.53×10^{-6}	1.83×10^{-6}	1.93×10^{-6}	1.41×10^{-6}	1.99×10^{-6}
10^3	3.33×10^{-5}	2.11×10^{-5}	3.65×10^{-5}	3.38×10^{-5}	2.23×10^{-5}	4.15×10^{-5}
10^6	2.08×10^{-2}	1.44×10^{-2}	2.89×10^{-2}	2.18×10^{-2}	1.44×10^{-2}	3.42×10^{-2}
	$\sigma = 0.01$			$\sigma = 0.005$		
20	2.05×10^{-6}	1.45×10^{-6}	1.87×10^{-6}	1.92×10^{-6}	1.36×10^{-6}	2.32×10^{-6}
10^3	3.14×10^{-5}	2.57×10^{-5}	4.07×10^{-5}	3.06×10^{-5}	2.68×10^{-5}	4.46×10^{-5}
10^6	1.93×10^{-2}	1.48×10^{-2}	3.73×10^{-2}	1.89×10^{-2}	1.50×10^{-2}	4.00×10^{-2}

TABLE 1. Average time (seconds) for projecting vectors onto the 1-norm unit ball in dimension n, with coordinates chosen using Gaussian distributions with standard deviation σ.



$(\bar{x}, \bar{\alpha}) = (+1, -1)$ $(\bar{x}, \bar{\alpha}) = (-1, -1)$

FIGURE 4. The graph of the function $\theta'_{\text{epi}}(\lambda)$ that corresponds to the base points $(\bar{x}, \bar{\alpha})$ shown for each figure. The left panel depicts the case where $P_{\text{cl}\,(\text{dom}\, f)}(\bar{x}) \in \text{dom}\, f$; the right panel depicts the case where $P_{\text{cl}\,(\text{dom}\, f)}(\bar{x}) \notin \text{dom}\, f$.

	$n = 1$	$n = 10^3$	$n = 10^6$
SSN	8.76×10^{-7}	1.69×10^{-4}	1.90×10^{-1}
Bisection 1	2.36×10^{-6}	1.88×10^{-3}	2.89
Bisection 2	2.66×10^{-6}	1.08×10^{-3}	1.16

TABLE 2. Time (seconds) for projecting vectors onto the epigraph of $f(x) = -\sum_{i=1}^{n} \log x_i$ in various dimensions n.

ϕ'_{epi} for the case when $P_{\text{cl}\,(\text{dom}\, f)}(\bar{x})$ is in, and not in, the domain of f. These functions are concave over $(0, \infty)$. Hence $-\theta'_\xi$ is convex over this interval and Algorithm 2 applies.

We numerically compare Algorithm 2 and the bisection method as solution approaches for (31). The coordinates of $\bar{x}$ were chosen uniformly at random on the interval $[-1, 1]$, and the value $\bar{\alpha}$ was chosen uniformly at random on the interval $[-2, -0.5]$. The initial value λ_0 was chosen to be $\sqrt{N}$. The termination condition for Algorithm 2 was $|\theta'_\xi(\lambda)| < 10^{-4}$, and the termination conditions for bisection was $|\theta'_\xi(\lambda)| < 10^{-4}$ (labeled *Bisection 1*) and $|b - a| < 10^{-8}$ (labeled *Bisection 2*), where a, b denote the endpoints of the bisection interval. Table 2 shows the average times over 10^5 trials when $n \in \{1, 10^3\}$, and over 500 trials when $n = 10^6$.

7.2.3. Discussion The numerical examples we presented extend easily to other useful cases involving matrices, such as the nuclear norm on $\mathbb{R}^{m \times n}$ and the barrier function $-\log\det$ on the space of symmetric matrices, using variational formulas that depend on matrix spectra [28, 29].

In these cases, the main computational effort involves computing singular value and eigenvalue decompositions, respectively, of the matrix iterates.

The cases where θ'_ξ does not satisfy either Assumption 1 or the domain condition $P_{\mathrm{cl}\,(\mathrm{dom}\, f)}(\bar{x}) \in \mathrm{dom}\, f$ lies outside the theoretical guarantees presented in this section, though the algorithms we present may still work in practice. In the case where $\mathrm{dom}\, f \subsetneq \mathbb{E}$ is open, the formula provided by Chierchia et al. [11, Proposition 1] is a viable option.

8. Final remarks Our analysis on the variational properties of epigraphical projections and infimal convolution is motivated by the authors' larger research interests on variations of first-order methods that operate in a lifted space. The promising work by Chierchia et al. [11] on epigraphical-projection methods for minimizing convex functions over p-norm constraints shows promise for this algorithmic approach, and we aim to develop methods for more general problem classes. We are also motivated by statistical M-estimation approaches that include as an additional unknown a particular parameter that characterizes data distribution [15]. The variational calculus that we derive is a useful tool for developing algorithmic approaches for solving these lifted M-estimation problems.

There are at least two avenues of future research that extend our analysis in this paper.

***K*-epigraphical projections.** A significant generalization of the post-composition operation defined in Section 6 occurs when we allow compositions of the form $f = g \circ H : \mathbb{E}_1 \to \overline{\mathbb{R}}$, where

- $K \subset \mathbb{E}_1$ a closed convex cone;
- $H : \mathbb{E}_1 \to \mathbb{E}_2$ K-convex, i.e., the K-*epigraph* $\{(X, Y) \mid Y - H(x) \in K\}$ is convex;
- $g \in \Gamma_0(\mathbb{E}_2)$ K-*increasing*, i.e., $g \le g((\cdot) + v)$ for all $v \in K$.

This *convex convex-composite* setting was studied by Burke et al. [10], and the required subdifferential formulas for the analysis are readily available. This may lead to a proximal calculus and ultimately to formulas and algorithms for projecting onto K-epigraphs, thus encompassing the study in Section 6.

Semismoothness* of subdifferential operators. The notion of semismooth* sets and maps is recent and still in development. One of the critical conditions in our study is the semismoothness* of the subdifferential operator ∂f, which also occurs in a recent report by Khanh et al. [27]. This suggests an important avenue of research that relaxes the overarching convexity assumption and, in particular, establishes verifiable sufficient conditions.

Acknowledgments M.P. Friedlander and T. Hoheisel are supported by NSERC Discovery grants, while A. Goodwin's work was partially supported by an NSERC summer research stipend. T. Hoheisel would like to thank Dr. Matus Benko, University of Vienna, for valuable discussions on semismoothness*.

References

[1] A.Y. Aravkin , J.V. Burke, D. Drusvyatskiy, M.P. Friedlander, and K.J. MacPhee: *Foundations of Gauge and Perspective Duality.* SIAM Journal on Optimization, 28(3), 2018, pp. 2406–2434.

[2] H. Attouch: *Variational Convergence for Functions and Operators.* Applied Mathematics Series, Pittman, Boston, 1984.

[3] H.H. Bauschke and P.L. Combettes: *Convex analysis and Monotone Operator Theory in Hilbert Spaces.* CMS Books in Mathematics, Springer, New York, 2nd Edition, 2017.

[4] A. Beck: *First-Order Methods in Optimization.* MOS-SIAM Series on Optimization, 2017.

[5] A. Beck and M. Teboulle: *Smoothing and first order methods: A unified framework.* SIAM Journal on Optimization 22 (2), 2012, pp. 557–580.

[6] M. Benko, H. Gfrerer, and J.V. Outrata: *Calculus for Directional Limiting Normal Cones and Subdifferentials.* Set-Valued and Variational Analysis 27, 2019, pp. 713–745.

[7] M. Bougeard, J.P. Penot, and A. Pommellet: *Towards minimal assumptions for the infimal convolution regularization.* Journal of Approximation Theory 64(3), 1991, pp. 245–270.

[8] J.V. Burke and T. Hoheisel: *Epi-convergent smoothing with applications to convex composite functions.* SIAM Journal on Optimization 23(3), 2013, pp. 1457–1479.

[9] J.V. Burke and T. Hoheisel: *Epi-convergence properties of smoothing by infimal convolution.* Set-Valued and Variational Analysis 25, 2017, pp. 1–23.

[10] J.V. Burke, T. Hoheisel, and Q.V. Nguyen: *A study of convex convex-composite functions via infimal convolution with applications.* Mathematics of Operations Research, to appear.

[11] G. Chierchia, N. Pustelnik, J.-C. Pesquet, B. Pesquet-Popescu: *Epigraphical projection and proximal tools for solving constrained convex optimization problems.* Signal, Image and Video Processing 9, 2015, pp. 1737–1749.

[12] F.H. Clarke: *Optimization and Nonsmooth Analysis.* John Wiley & Sons, New York, 1983.

[13] P.L. Combettes: *Perspective functions: properties, constructions, and examples.* Set-Valued and Variational Analysis 26, 2019, pp. 247–264.

[14] P.L. Combettes and C.L. Müller: *Perspective functions: proximal calculus and applications in high-dimensional statistics.* Journal of Mathematical Analysis and Applications 457(2), 2018, pp. 1283–1306.

[15] P.L. Combettes and C.L. Müller: *Perspective maximum likelihood-type estimation via proximal decomposition.* Electronic Journal of Statistics 14, 2020, pp. 207–238.

[16] L. Condat: *Fast projection onto the simplex and l1 ball.* Mathematical Programming, Series A, Springer, 2016, 158 (1), pp. 575–585.

[17] L. Condat: URL https://lcondat.github.io/software.html Last accessed January 27, 2021.

[18] A.L. Dontchev and R.T. Rockafellar: *Implicit Functions and Solution Mappings. A View from Variational Analysis.* Springer Series in Operations Research and Financial Engineering, Springer-Verlag New York, 2014.

[19] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra: *Efficient projections onto the l1-ball for learning in high dimensions.* ICML '08: Proceedings of the 25th international conference on Machine learning, ACM, New York, NY, USA, 2008, pp. 272–279.

[20] F. Facchinei and J.-S. Pang: *Finite-Dimensional Variational Inequalitites and Complementarity Problems, Volumes I and II*, Springer, New York, 2003.

[21] H. Gfrerer: *On directional metric subregularity and second-order optimality conditions for a class of nonsmooth mathematical programs.* SIAM Journal on Optimization 23(1), 2013, pp. 63–665.

[22] H. Gfrerer: *On directional metric regularity, subregularity and optimality conditions for nonsmooth mathematical programs.* Set-Valued and Variational Analysis 21, 2013, pp. 151–176.

[23] H. Gfrerer and J.V. Outrata: *On a semismooth* Newton method for solving generalized equations.* SIAM Journal on Optimization. 31(1), 2021, pp. 489–517.

[24] I. Ginchev and B.S. Mordukhovich: *Directional subdifferentials and optimality conditions.* Positivity 16, 2012, pp. 707–737.

[25] J.-B. Hiriart-Urruty and C. Lemaréchal: *Fundamentals of Convex Analysis.* Grundlehren Text Editions, Springer, Berlin, Heidelberg, 2001.

[26] T. Hoheisel: *Topics in Convex Analysis in Matrix Space.* Lecture Notes, Spring School on Variational Analysis, Paseky nad Jizerou, Czech Republic, 2019.

[27] P.D. Khanh, B.S. Mordukhovich, and V.T. Phat: *A generalized Newton method for subgradient systems.* arXiv:2009.10551, 2020.

[28] A.S. Lewis: *The convex analysis of unitarily invariant matrix functions.* Journal of Convex Analysis 2(1–2), 1995, pp. 173–183.

[29] A.S. Lewis:*Convex analysis on the Hermitian Matrices.* SIAM Journal on Optimization 6(1), 1996, pp. 164–177.

[30] J. Liu and J. Ye: *Efficient Euclidean projections in linear time.* Proceedings of the 26th Annual International Conference on Machine Learning, 2009, pp. 657–664.

[31] F. Meng, D. Sun, and G. Zhao: *Semismoothness of solutions to generalized equations and the Moreau-Yosida regularization.* Mathematical Programming 104, 2005, pp. 561–581.

[32] F. Meng, G. Zhao, M. Goh, and R. De Souza: *Lagrangian-dual functions and Moreau-Yosida regularization.* SIAM Journal on Optimization 19, 2008, pp. 39–61.

[33] A. Milzarek: *Numerical Methods and Second Order Theory for Nonsmooth Problems.* Dissertation, Technical University of Munich, 2016.

[34] B.S. Mordukhovich: *Variational Analysis and Applications.* Springer Monographs in Mathematics book series, Springer International Publishing AG, 2018.

[35] P. Neal and S. Boyd: *Proximal algorithms.* Foundations and Trends in Optimization 1(3), 2013, pp. 123–231.

[36] J.S. Pang and L. Qi: *A Globally convergent Newton method for convex SC^1 minimization problems.* Journal of Optimization Theory and Applications 85(3), 1995, pp. 633–648.

[37] L. Qi and J. Sun: *A nonsmooth version of Newton's method.* Mathematical Programming 58, 1993, pp. 353–367.

[38] R.T. Rockafellar: *Convex Analysis.* Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J. 1970.

[39] R.T. Rockafellar and R.J.-B. Wets: *Variational Analysis.* Grundlehren der Mathematischen Wissenschaften, Vol. 317, Springer-Verlag, Berlin, 1998.

[40] A. Shapiro: *Directionally nondifferentiable metric projection.* Journal of Optimization Theory and Applications 81(1), 1994, pp. 203–204.

[41] J. Liu, S. Ji, and J. Ye: SLEP: Sparse Learning with Efficient Projections, http://www.yelabs.net/software/SLEP/ Arizona State University, 2009.

[42] T. Strömberg: *The Operation of Infimal Convolution.* Dissertationes Mathematicae (Rozprawy Matematyczne) 352, 1996.

[43] M. Tofighi, K. Kose, and A.E. Cetin: *Denoising using projections onto the epigraph set of convex cost functions.* IEEE International Conference on Image Processing (ICIP), Paris, 2014, pp. 2709–2713.

[44] M. Tofighi, A. Bozkurt, K. Kose, and A.E. Cetin: *Deconvolution using projections onto the epigraph set of a convex cost function.* 22nd Signal Processing and Communications Applications Conference (SIU), 2014, pp. 1638–1641.

[45] P.-W. Wang, M. Wytock, and J.Z. Kolter: *Epigraph projections for fast general convex programming.* Proceedings of the 33rd International Conference on International Conference on Machine Learning 48, New York, 2016, pp. 2868–2877.