Chapter 1

MOREAU'S PROXIMAL MAPPINGS AND CONVEXITY IN HAMILTON-JACOBI THEORY

R. T. Rockafellar *

Department of Mathematics, University of Washington, Seattle, WA 98195-4350, USA

rtr@math.washington.edu

Abstract Proximal mappings, which generalize projection mappings, were introduced by Moreau and shown to be valuable in understanding the subgradient properties of convex functions. Proximal mappings subsequently turned out to be important also in numerical methods of optimization and the solution of nonlinear partial differential equations and variational inequalities. Here it is shown that, when a convex function is propagated through time by a generalized Hamilton-Jacobi partial differential equation with a Hamiltonian that concave in the state and convex in the co-state, the associated proximal mapping exhibits locally Lipschitz dependence on time. Furthermore, the subgradient mapping associated of the value function associated with this mapping is graphically Lipschitzian.

Keywords: Proximal mappings, Hamilton-Jacobi theory, convex analysis, subgradients, Lipschitz properties, graphically Lipschitzian mappings.

1. INTRODUCTION

In some of his earliest work in convex analysis, J.-J. Moreau introduced in [1], [2], the *proximal* mapping P associated with a lower semicontinuous, proper, convex function f on a Hilbert space $\mathcal{H}$, namely

$$P(z) = \operatorname{argmin}_x \Big\{ f(x) + \tfrac{1}{2}||x - z||^2 \Big\}. \tag{1.1}$$

It has many remarkable properties. Moreau showed that P is everywhere single-valued as a mapping from $\mathcal{H}$ into $\mathcal{H}$, and moreover is nonexpansive:

$$||P(z') - P(z)|| \leq ||z' - z|| \text{ for all } z, z'. \tag{1.2}$$

*Research supported by the U.S. National Science Foundation under grant DMS–104055

In this respect P resembles a projection mapping, and indeed when f is the indicator of a convex set C, P is the projection mapping onto C. He also discovered an interesting duality. The proximal mapping associated with the convex function f^* conjugate to f, which we can denote by

$$Q(z) = \operatorname{argmin}_x \Big\{ f^*(y) + \tfrac{1}{2}||y - z||^2 \Big\}, \tag{1.3}$$

which likewise is nonexpansive of course, satisfies

$$Q = I - P, \qquad P = I - Q. \tag{1.4}$$

In fact the mappings P and Q serve to parameterize the generally set-valued subgradient mapping ∂f associated with f:

$$y \in \partial f(x) \iff (x, y) = (P(z), Q(z)) \text{ for some } z, \tag{1.5}$$

this z being determined uniquely through (1.4) by $z = x+y$. Another important feature is that the *envelope* function associated with f, namely

$$E(z) = \min_x \Big\{ f(x) + \tfrac{1}{2}||x - z||^2 \Big\}, \tag{1.6}$$

is a finite convex function on $\mathcal{H}$ which is Fréchet differentiable with gradient mapping

$$\nabla E(z) = Q(z). \tag{1.7}$$

Our objective in this article is to tie Moreau's proximal mappings and envelopes into the Hamilton-Jacobi theory associated with convex optimization over absolutely continuous arcs $\xi : [0, t] \to I\!R^n$; for this setting we henceforth will have $\mathcal{H} = I\!R^n$. Let the space of such arcs be denoted by $\mathcal{A}^1_n[0, t]$.

The optimization problems in question concern the functions f_t on $I\!R^n$ defined by $f_0 = f$ and

$$f_t(x) = \min_{\xi \in \mathcal{A}^n[0,t]} \Big\{ f(\xi(0)) + \int_0^t L(\xi(\tau), \dot{\xi}(\tau))d\tau \Big\} \text{ for } t > 0, \tag{1.8}$$

which represent the propagation of f forward in time t under the "dynamics" of a Lagrangian function L.

A pair of recent articles [3], [4], has explored this in the case where L satisfies the following assumptions, which we also make here:

(A1) The function L is convex, proper and lsc on $I\!R^n \times I\!R^n$.

(A2) The set $F(x) := \operatorname{dom} L(x, \cdot)$ is nonempty for all x, and there is a constant ρ such that $\operatorname{dist}(0, F(x)) \leq \rho(1 + ||x||)$ for all x.

(A3) There are constants α and β and a coercive, proper, nondecreasing function θ on $[0, \infty)$ such that $L(x, v) \geq \theta(\max\{ 0, ||v|| - \alpha||x|| \}) - \beta||x||$ for all x and v.

The convexity of $L(x,v)$ with respect to (x,v) in (A1), instead of just with respect to v, is called *full convexity*. It opens the way to broad use of the tools of convex analysis. The properties in (A2) and (A3) are dual to each other in a sense brought out in [3] and provide coercivity and other needed features of the integral functional.

It was shown in [3, Theorem 2.1] that, under these assumptions, f_t is, for every t, *a lower semicontinuous, proper, convex function on* $I\!R^n$ *which depends epi-continuously on* t (i.e., the set-valued mapping $t \mapsto \operatorname{epi} f_t$ is continuous with respect to $t \in [0,\infty)$ in the sense of Kuratowsk-Painlevé set convergence [6]). The topic we wish to address here is how, in that case, the associated proximal mappings

$$P_t(z) = \operatorname{argmin}_x \Big\{ f_t(x) + \tfrac{1}{2}||x-z||^2 \Big\}, \tag{1.9}$$

with $P_0 = P$, and envelope functions

$$E_t(z) = \min_x \Big\{ f_t(x) + \tfrac{1}{2}||x-z||^2 \Big\}, \tag{1.10}$$

with $E_0 = E$, behave in their dependence on t. It will be useful for that purpose to employ the notation

$$\bar{P}(t,z) = P_t(z), \qquad \bar{E}(t,z) = E_t(z). \tag{1.11}$$

Some aspects of this dependence can be deduced readily from the epi-continuity of f_t in t, for example the continuity of $\bar{P}(t,z)$ and $\bar{E}(t,z)$ with respect to $t \in [0,\infty)$; cf. [6, 7.37, 7.38]. From that, it follows through the nonexpansivity of $\bar{P}(t,z)$ in z and the finite convexity of $\bar{E}(t,z)$ in z, that both $\bar{P}(t,z)$ and $\bar{E}(t,z)$ are continuous with respect to $(t,z) \in [0,\infty) \times I\!R^n$.

At the end of our paper [7] in an application of other results about variational problems with full convexity, we were able to show more recently that $\bar{E}(t,z)$ is in fact *continuously differentiable with respect to* (t,z), not just with respect to z, as would already be a consequence of the convexity and differentiability of E_t, noted above. But this property does not, by itself, translate into any extra feature of the dependence of $\bar{P}(t,z)$ on t, beyond the continuity we already have at our disposal.

The following new result which we contribute here thus reaches a new level, moreover one where $\bar{P}$ and $\bar{E}$ are again on a par with each other.

Theorem 1. *Under (A1), (A2) and (A3), both* $\bar{P}(t,z)$ *and* $\nabla\bar{E}(t,z)$ *are locally Lipschitz continuous with respect to* (t,z)*. Thus,* $\bar{E}$ *is a function of class* $\mathcal{C}^{1+}$.

Our proof will rely on the Hamilton-Jacobi theory in [4] for the forward propagation expressed by (1.8). It concerns the characterization of the function

$$\bar{f}(t,x) = f_t(x) \ \text{ for } \ (t,x) \in [0,\infty) \times I\!R^n \tag{1.12}$$

in terms of a generalized "method of characteristics" in subgradient format.

2. HAMILTON-JACOBI FRAMEWORK

The Hamiltonian function H that corresponds to the Lagrangian function L is obtained by passing from the convex function $L(x,\cdot)$ to its conjugate:

$$H(x,y) := \sup_v \Big\{ \langle v,y\rangle - L(x,v) \Big\}. \tag{2.1}$$

Because of the lower semicontinuity in (A1) and the properness of $L(x,\cdot)$ implied by (A2), the reciprocal formula holds that

$$L(x,v) = \sup_y \Big\{ \langle v,y\rangle - H(x,y) \Big\}, \tag{2.2}$$

so L and H are completely dual to each other. It was established in [3] that a function $H : I\!R^n \times I\!R^n \to \overline{I\!R}$ is the Hamiltonian for a Lagrangian L fulfilling (A1), (A2) and (A3) if and only if it satisfies

(H1) $H(x,y)$ is concave in x, convex in y, and everywhere finite,
(H2) There are constants α and β and a finite, convex function φ such that

$$H(x,y) \;\leq\; \varphi(y) + (\alpha||y|| + \beta)||x|| \quad \text{for all } \; x, \; y.$$

(H3) There are constants γ and δ and a finite, concave function ψ such that

$$H(x,y) \;\geq\; \psi(x) - (\gamma||x|| + \delta)||y|| \quad \text{for all } \; x, \; y.$$

The convexity-concavity of H in (H1) is a well known counterpart to the full convexity of L under the “partial conjugacy” in (2.1) and (2.2). (cf. [5]). It impliers in particular that H is locally Lipschitz continuous; cf. [5, §35]. The growth conditions in (H2) and (H3) are dual to (A3) and (A2), respectively. This duality underscores the refined nature of (A2) and (A3); they are tightly intertwined. They have also been singled out because of the role they can play in control theory of fully convex type. For example, L satisfies (A1), (A2), (A3), when it has the form

$$L(x,v) = g(x) + \min_u \{ \, h(u) \,|\, Ax + Bu = v \} \tag{2.3}$$

for matrices $A \in I\!R^{n\times n}$, $B \in I\!R^{m\times n}$, a finite convex function g and a lower semicontinuous, proper convex function h that is coercive, or equivalently, has finite conjugate h^*. Then $f_t(x)$ is the minimum in the problem of minimizing

$$f(\xi(0)) + \int_0^t \{g(\xi(\tau)) + h(\omega(\tau))\} dt$$

over all summable control functions $\omega : [0,t] \to I\!R^m$ such that

$$\dot{\xi}(\tau) = A\xi(\tau) + B\omega(\tau) \;\text{ for a.e. } \tau, \quad \xi(t) = x.$$

The corresponding Hamiltonian in this case is

$$H(x,y) = \langle Ax, y\rangle - g(x) + h^*(y). \tag{2.4}$$

In the control context, backward propagation from a terminal time would be more natural than forward propagation from time 0, but it is elementary to reformulate from one to the other. Forward propagation is more convenient mathematically for the formulas that can be developed.

For any finite, concave-convex function H on $I\!R^n \times I\!R^m$, there is an associated *Hamiltonian dynamical sytem*, which can be written as the differential inclusion

$$\dot{\xi}(\tau) \in \partial_y H(\xi(\tau),\eta(\tau)), \quad -\dot{\eta}(\tau) \in \partial_x H(\xi(\tau),\eta(\tau)) \text{ for a.e. } \tau, \tag{2.5}$$

where ∂_y refers to subgradients in the convex sense in the y argument, and ∂_x refers to subgradients in the concave sense in the x argument. In principle, the candidates ξ and η for a solution over an interval $[0,t]$ could just belong to $\mathcal{A}^1_n[0,t]$, but the local Lipschitz continuity of H, and the local boundedness it entails for the subgradient mappings that are involved [5, §35], guarantee that ξ and η belong to $\mathcal{A}^\infty_n[0,t]$, i.e., that they are Lipschitz continuous.

Dynamics of the kind in (2.5) were first introduced in [8] for their role in capturing optimality in variational problems with fully convex Lagrangians. In the present circumstances where (A1), (A2) and (A3) hold, it has been established in [3] that

$$\xi \text{ solves (1.8)} \iff \begin{cases} \xi(t) = x \text{ and } (\xi,\eta) \text{ solves (2.5)} \\ \text{for some } \eta \text{ with } \eta(0) \in \partial f(\xi(0)) \,. \end{cases} \tag{2.6}$$

Again, ∂f refers to subgradients of the convex function f in the traditions of convex analysis. Another powerful property obtained in [3], which helps in characterizing the functions f_t and therefore $\bar{f}$, is that

$$y \in \partial f_t(x) \iff \begin{cases} (\xi(t),\eta(t)) = (x,y) \text{ for some } (\xi,\eta) \\ \text{satisfying (2.5) with } \eta(0) \in \partial f(\xi(0)). \end{cases} \tag{2.7}$$

Yet another property from [3], which we can take advantage of here, is that, for any $(x_0,y_0) \in I\!R^n \times I\!R^n$, the Hamiltonian system has at least one trajectory pair (ξ,η) that starts from (x_0,y_0) and continues forever, i.e., for the entire time interval $[0,\infty)$. This implies further that any trajectory up to a certain time t can be continued indefinitely beyond t. Such trajectories need not be unique, however.

Subgradients of the value function $\bar{f}$ in (1.12) must be considered as well. The complication there is that $\bar{f}(t,x)$ is only convex with respect to x, not (t,x). Subgradient theory beyond convex analysis is therefore essential. In this respect, we use $\partial\bar{f}$ to denote subgradients with respect to (t,x) in the broader sense of variational analysis laid out, for instance, in [6]. These avoid the convex hull operation in the definition utilized earlier by Clarke and are merely "limiting subgradients" in that context.

The key result from [3] concerning subgradients $\partial \bar{f}$, which we will need to utilize later, reveals that, for $t > 0$,

$$(s, y) \in \partial \bar{f}(t, x) \iff y \in \partial f_t(x) \text{ and } s = -H(x, y). \tag{2.8}$$

Observe that the implication "⇒" in (2.8) says that $\bar{f}$ satisfies a subgradient version of Hamilton-Jacobi partial differential equation for H and the initial condition $\bar{f}(0, x) = f(x)$. It becomes the classical version when $\bar{f}$ is continuously differentiable, so that $\partial \bar{f}(t, x)$ reduces to the singleton $\nabla \bar{f}(t, x)$. This subgradient version turns out, in consequence of other developments in this setting, to agree with the "viscosity" version of the Hamilton-Jacobi equation, but is not covered by the uniqueness results that have so far been been achieved in that setting. The uniqueness of $\bar{f}$ as a solution, under our conditions (H1), (H2), (H3), and the initial function f follows, nonetheless, from independent arguments in variational analysis; cf. [9], [10].

By virtue of its implication "⇐" in our context of potential nonsmoothness, (2.8) furnishes more than just a generalized Hamilton-Jacobi equation. Most importantly, it can be combined with (2.7) to see that

$$(s, y) \in \partial \bar{f}(t, x) \iff \begin{cases} \exists(\xi, \eta) \text{ satisfying (2.6) over } [0, t] \\ \text{such that } (\xi(t), \eta(t)) = (x, y) \\ \text{and } -H(\xi(t), \eta(t)) = s. \end{cases} \tag{2.9}$$

This constitutes a generalized "method of characteristics" of remarkable completeness, and in a global pattern not dreamed of in classical Hamilton-Jacobi theory, where everything depends essentially on the implicit function theorem with its local character. Instead of relying on such classical underpinnings, the characterization in (2.9) is based on convex analysis and extensive appeals to duality.

Proof of Theorem 1. We concentrate first on the claims about $\bar{P}$, which we already know to have the property that

$$||\bar{P}(t, z') - \bar{P}(t, z)|| \leq ||z' - z|| \text{ for all } z, z' \in \mathbb{R}^n,\ t \in [0, \infty). \tag{2.10}$$

To confirm the local Lipschitz continuity of $\bar{P}(t, z)$ with respect to (t, z), it will be enough, on this basis, to demonstrate local Lipschitz continuity in t with a constant that is locally uniform in z. Therefore, we fix any $t^* \in [0, \infty)$ and $z^* \in \mathbb{R}^n$, and take

$$x^* = \bar{P}(t^*, z^*), \quad y^* = \bar{Q}(t^*, z^*), \tag{2.11}$$

where

$$\bar{Q}(t, z) = Q_t(z) \text{ for } Q_t = I - P_t. \tag{2.12}$$

Fix any $(t^*, z^*) \in [0, \infty) \times I\!R^n$ along with a compact neighborhood $T_0 \times Z_0$ of this pair. The mapping

$$M : (t, z) \to (\bar{P}(t, z), \bar{Q}(t, z)) \in I\!R^{2n}, \qquad (2.13)$$

which we already know is continuous, takes $T_0 \times Z_0$ into a compact set $M(T_0, Z_0) \subset I\!R^{2n}$. Utilizing the fact that H is locally Lipschitz continuous on $I\!R^{2n}$, we can select compact subsets U_0 and U_1 of $I\!R^n$ such that $M(T_0, Z_0) \subset U_1 \subset \operatorname{int} U_0$ and furthermore

$$(x, y) \in U_0,\ u \in \partial_x H(x, y),\ v \in \partial_y H(x, y) \quad \Longrightarrow \quad \begin{cases} ||u|| \leq \kappa, \\ ||v|| \leq \kappa. \end{cases}$$

Trajectories (ξ, η) to the Hamiltonian system in (2.5) are then necessarily Lipschitz continuous with constant κ over time intervals during which they stay inside U_0. It is possible next, therefore, to choose an interval neighborhood T_1 of t^* within T_0 such that any Hamiltonian trajectory (ξ, η) over T_1 that touches U_1 remains entirely in U_0 (and thus has the indicated Lipschitz property). Finally, we can choose a neighborhood $T \times Z$ of (t^*, z^*) within $T_1 \times Z_0$, with T_1 being an interval, such that $M(T, Z) \subset U_1$.

With these preparations completed, consider any $z \in Z$ and any interval $[t, t'] \subset T$, with $t < t'$. Let $(x, y) = M(t, z)$, so that

$$x = \bar{P}(t, z) = P_t(z), \qquad y = \bar{Q}(t, z) = Q_t(z),$$

and consequently

$$y \in \partial f_t(x), \qquad x + y = z,$$

from the basic properties of proximal mappings. Also, $(x, y) \in U_1$. By (2.7), there is a Hamiltonian trajectory (ξ, η) over $[0, t]$ with $(\xi(t), \eta(t)) = (x, y)$. It can be continued over $[t, t']$. Our selection of $[t, t']$ ensures that, during that time interval, both ξ and η are Lipschitz continuous with constant κ.

We also have $\eta(\tau) \in \partial f_t(\xi(\tau))$; this follows by applying (2.7) to the interval $[0, \tau]$ in place of $[0, t]$. Let $\zeta(\tau) = \xi(\tau) + \eta(\tau)$ for $\tau \in [t, t']$. Then $\zeta(t) = z$ and ζ is Lipschitz continuous with constant κ. Moreover

$$\xi(\tau) = P_\tau(\zeta(\tau)) = \bar{P}(\tau, \zeta(\tau)), \qquad \eta(\tau) = Q_\tau(\zeta(\tau)) = \bar{Q}(\tau, \zeta(\tau)),$$

again according to Moreau's theory of proximal mappings. Now, by writing

$$\bar{P}(t', z) - \bar{P}(t, z) = [P(t', \zeta(t)) - P(t', \zeta(t'))] + [P(t', \zeta(t')) - P(t, \zeta(t))],$$

where $||P(t', \zeta(t)) - P(t', \zeta(t'))|| \leq ||\zeta(t)) - \zeta(t')||$ and, on the other hand, $P(t, \zeta(t)) = \xi(t)$ and $P(t', \zeta(t')) = \xi(t')$, we are able to estimate that

$$||\bar{P}(t, z) - \bar{P}(t, z)|| \ \leq \ ||\zeta(t') - \zeta(t)|| + ||\xi(t') - \xi(t)|| \ \leq \ 2\kappa|t' - t|.$$

Because this holds for all $z \in Z$ and $[t, t'] \subset T$, we have the locally uniform Lipschitz continuity property that was required for $\bar{P}$ in its time argument.

Note that the local Lipschitz continuity of $\bar{P}$ implies the same propert for $\bar{Q}$, inasmuch as $\bar{Q}(t,z) = z - \bar{P}(t,z)$.

Turning now to the claims about $\bar{E}$, we observe, to begin with, that since $\nabla E_t = Q_t$ from proximal mapping theory, we have $\nabla_z \bar{E}(t,z) = \bar{Q}(t,z)$. A complementary fact, coming from [7, Theorem 4 and Corollary], is that

$$\frac{\partial \bar{E}}{\partial t}(t,z) = -H(x,y) \quad \text{for } (x,y) = (\bar{P}(t,z), \bar{Q}(t,z)).$$

In these terms we have

$$\nabla \bar{E}(t,z) = (-\,H(\bar{P}(t,z), \bar{Q}(t,z)), \bar{Q}(t,z)).$$

Since H is locally Lipschitz continuous, and both $\bar{P}$ and $\bar{Q}$ are locally Lipschitz continuous, as just verified, we conclude that $\nabla \bar{E}$ is locally Lipschitz continuous, as claimed, too. □

3. SUBGRADIENT GRAPHICAL LIPSCHITZ PROPERTY

The facts in Theorem 1 lead to a further insight into the subgradients of the function $\bar{f}$. To explain it, we recall the concept of a set-valued mapping $S : I\!R^p \to I\!R^q$ being *graphically Lipschitzian of dimension* d around a point $(\bar{u}, \bar{v})$ in its graph. This means that there is some neighborhood of $(\bar{u}, \bar{v})$ in which, under a smooth change of coordinates, the graph of S can be identified with that of a Lipschitz continuous mapping on a d-dimensional parameter space.

The subgradient mappings $\partial f : I\!R^n \rightrightarrows I\!R^n$ associated with lower semicontinuous, proper, convex functions f on $I\!R^n$, like here, are prime examples of graphically Lipschitzian mappings. Indeed, this property is provided by Moreau's theory of proximal mappings. In passing from the x, y, "coordinates" in which the relation $y \in \partial f(x)$ is developed, to the z, w, "coordinates" specified by $z = x + y$ and $w = x - y$, we get just the kind of representation demanded, since the graph of ∂f can be viewed parameterically in terms of the pairs $(P(z), Q(z))$ as z ranges over $I\!R^n$; cf. (1.5). Thus, ∂f is graphically Lipschitzian of dimension n, everywhere.

Is there an extension of this property to the mapping $\partial \bar{f}$ from $(0, \infty) \times I\!R^n$ to $I\!R \times I\!R^n$? The next theorem says yes.

Theorem 2. *Under (A1), (A2) and (A3), the subgradient mapping $\partial \bar{f}$ is everywhere graphically Lipschitzian of dimension $n+1$.*

Proof. We get this out of (2.8) and the parameterization properties developed in the proof of Theorem 1. These tell us that the representation

$$(-H(\bar{P}(t,z),\bar{Q}(t,z)),\bar{Q}(t,z)) \in \partial\bar{f}(t,\bar{P}(t,z))$$

fully covers the graph of $\partial\bar{f}$ in a one-to-one manner relative to $(0,\infty)\times I\!R^n$ as (t,z) ranges over $(0,\infty)\times I\!R^n$. This is an $n+1$-dimensional parameterization in which the mappings are locally Lipschitz continuous, so the assertion of the theorem is fully justified. □

References

[1] J.-J. MOREAU, Fonctiones convexes duales et points proximaux dans un espace hilbertien. *Comptes Rendus de l'Acadḿie des Sciences de Paris* **255** (1962), 2897–2899.

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

[3] R. T. ROCKAFELLAR, P. R. WOLENSKI, Convexity in Hamilton-Jacobi theory I: dynamics and duality. *SIAM Journal on Control and Optimization* **39** (2001), 1323–1350.

[4] R. T. ROCKAFELLAR, P. R. WOLENSKI, Convexity in Hamilton-Jacobi theory II: envelope representations. *SIAM Journal on Control and Optimization* **39** (2001), 1351–1372.

[5] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, 1970.

[6] R. T. ROCKAFELLAR, R. J-B WETS, *Variational Analysis*, Springer-Verlag, Berlin, 1997.

[7] R. T. ROCKAFELLAR, Hamilton-Jacobi theory and parametric analysis in fully convex problems of optimal control. *Journal of Global Optimization*, to appear.

[8] R. T. ROCKAFELLAR, Generalized Hamiltonian equations for convex problems of Lagrange. *Pacific Journal of Mathematics* **33** (1970), 411–428.

[9] G. GALBRAITH, *Applications of Variational Analysis to Optimal Trajectories and Nonsmooth Hamilton-Jacobi Theory*, Ph.D. thesis, University of Washington, Seattle (1999).

[10] G. GALBRAITH, The role of cosmically Lipschitz mappings in nonsmooth Hamilton-Jacobi theory (preprint).