Annals. Oper. Res. **101**(2001), 383–390

ENVELOPE THEOREMS IN DYNAMIC PROGRAMMING

FUAN ZHAO AND ADI BEN-ISRAEL

Abstract. The Envelope Theorem is a statement about derivatives along an optimal trajectory. In Dynamic Programming the Envelope Theorem can be used to characterize and compute the Optimal Value Function from its derivatives. We illustrate this here for the Linear-Quadratic Control Problem, the Resource Allocation Problem, and the Inverse Problem of Dynamic Programming.

1. Introduction

Consider an optimization problem

$$\begin{array}{ll} \text{opt} & f(\mathbf{u}, x) \\ \text{s.t.} & \mathbf{u} \in U \end{array} \tag{1}$$

where $x \in \mathbb{R}$ is a parameter, $U \subset \mathbb{R}^n$ is the feasible set (may depend on x), f is assumed differentiable and smooth as needed, and "opt" stands for "max" or "min". We assume that (1) has a unique optimal solution, $\mathbf{u}(x) \in U$, for all x in some set X. The **optimal value function** is then defined as

$$V(x) := f(\mathbf{u}(x), x) \tag{2}$$

Even if (1) does not have a solution, say the supremum

$$\sup \{f(\mathbf{u}, x) : \mathbf{u} \in U\}$$

is not attained, it may still be possible to have an envelope theorem for the optimal value function, see [5] and [11]. Nonsmooth analysis is needed if the solution is not unique, see [3] and [4] for more information. In this paper we assume the existence of a unique solution, as well as differentiability of the optimal value function $V(x)$. The **Envelope Theorem**, [7], then states that

$$V'(x) = \frac{\partial f}{\partial x}(\mathbf{u}(x), x) \tag{3}$$

i.e. the slope of the optimal value function is given by the partial derivative of f w.r.t. x along the optimal trajectory $(\mathbf{u}(x), x)$.

This theorem is used widely in mathematical economics. For example, the profit of a competitive firm facing price vector $\mathbf{p}$ is

$$\begin{array}{rl} \pi(\mathbf{p}) = \sup & \{\mathbf{p} \cdot \mathbf{u} - c(\mathbf{u})\} \\ \text{s.t.} & \mathbf{u} \in U \end{array}$$

the Fenchel conjugate of the cost function $c(\cdot)$ (defined as ∞ outside the feasible set U). The envelope theorem states here that

$$\pi'(\mathbf{p}) = \mathbf{u}$$

the optimal quantity vector, see [9, Chapter 7] and[10] for details.

Key words and phrases. Envelope Theorem, Dynamic Programming, Resource Allocation, Inverse Problem.
Dedicated to Anthony V. Fiacco on the occasion of his 70th birthday.

If the problem is constrained,

$$\begin{array}{ll} \text{opt} & f(\mathbf{u},x) \\ \text{s.t.} & g(\mathbf{u},x)=0 \end{array} \tag{4}$$

consider the Lagrangian,

$$L(\mathbf{u},x,\lambda) := f(\mathbf{u},x) - V(x) + \lambda g(\mathbf{u},x) \tag{5}$$

and the first order conditions for optimality (assuming a suitable constraint qualification holds)

$$L_{\mathbf{u}}(\mathbf{u},x,\lambda) = f_{\mathbf{u}}(\mathbf{u},x) + \lambda g_{\mathbf{u}}(\mathbf{u},x) = 0 \tag{6a}$$

$$L_x(\mathbf{u},x,\lambda) = f_x(\mathbf{u},x) - V'(x) + \lambda g_x(\mathbf{u},x) = 0 \tag{6b}$$

$$L_\lambda(\mathbf{u},x,\lambda) = g(\mathbf{u},x) = 0 \tag{6c}$$

where subscripts denote partial derivatives. The envelope theorem for this problem is given by (6b),

$$V'(x) = f_x(\mathbf{u}(x),x) + \lambda(x) g_x(\mathbf{u}(x),x) \tag{7}$$

along the optimal trajectory $(\mathbf{u}(x), x, \lambda(x))$.

A special case is the **resource allocation** problem

$$\begin{array}{ll} \text{opt} & f(\mathbf{u}) \\ \text{s.t.} & g(\mathbf{u})=x \end{array} \tag{8}$$

where x is the resource level, and $g(\mathbf{u})$ is the consumption associated with the decision $\mathbf{u}$. The envelope theorem (7) here gives

$$V'(x) = \lambda \tag{9}$$

showing that the Lagrange multiplier λ is the **marginal value** of the resource.

In this paper we apply the envelope theorem to Dynamic Programming, in particular to resource allocation and inverse problems.

2. An Envelope Theorem for Dynamic Programming

Consider a **Dynamic Programming** (DP) problem

$$\begin{array}{ll} \min & \sum_{k=1}^{N-1} f_k(u_k,x_k) \\ \text{s..t.} & x_{k+1} = T_k(u_k,x_k)\ ,\ k=1,2,\cdots,N-1 \\ & x_1 \text{ given} \end{array} \tag{10}$$

where, at the $k\,\underline{\text{th}}$ stage,

$x_k =$ the initial **state**, $f_k(u,x) =$ the **return**,
$u_k =$ the **decision**, $T_k(u,x) =$ the **state transformation**, leading to

$x_{k+1} =$ the next state, also denoted x_+. At the $N\,\underline{\text{th}}$ (and final) stage, there is no decision, and the return $f_N(x_N)$ depends only on the last state x_N.

The DP problem (10) is solved recursively using the **Bellman optimality principle**,

$$V_k(x) = \min_u \{f_k(u,x) + V_{k+1}(x_+)\}\ ,\ k=1,\cdots,N-1 \tag{11a}$$

$$V_N(x) = f_N(x) \tag{11b}$$

where V_k is the k $\underline{\text{th}}$ **value function**, assumed differentiable. We assume that (11a) has an optimal solution satisfying the first-order optimality conditions,

$$\frac{\partial f_k(u,x)}{\partial u} + \frac{\partial T_k(u,x)}{\partial u}\, V'_{k+1}(T_k(u,x)) \;=\; 0 \qquad (12a)$$

$$\frac{\partial f_k(u,x)}{\partial x} + \frac{\partial T_k(u,x)}{\partial x}\, V'_{k+1}(T_k(u,x)) \;=\; V'_k(x) \qquad (12b)$$

where (12b) is an envelope theorem. If the derivative $\dfrac{\partial T_k(u,x)}{\partial u}$ is nonzero, we can eliminate $V'_{k+1}(T_k(u,x))$ from (12) to get

$$\begin{aligned} V'_k(x) &= \frac{\partial f_k(u,x)}{\partial x} - \frac{\partial T_k(u,x)}{\partial x}\left(\frac{\partial T_k(u,x)}{\partial u}\right)^{-1}\frac{\partial f_k(u,x)}{\partial u} \qquad (13)\\ &= \left(1, -\frac{\partial T_k(u,x)}{\partial x}\left(\frac{\partial T_k(u,x)}{\partial u}\right)^{-1}\right)\nabla f_k(u,x)\ , \quad \text{where } \nabla f_k(u,x) = \begin{pmatrix} \dfrac{\partial f_k(u,x)}{\partial x} \\ \dfrac{\partial f_k(u,x)}{\partial u}\end{pmatrix} \end{aligned}$$

The corresponding result in the vector case (where $\mathbf{x}$ and $\mathbf{u}$ are vectors) is

$$\nabla V_k(\mathbf{x}) = \left(I, -\frac{\partial T_k(\mathbf{u},\mathbf{x})}{\partial \mathbf{x}}\left(\frac{\partial T_k(\mathbf{u},\mathbf{x})}{\partial \mathbf{u}}\right)^{-1}\right)\nabla f_k(\mathbf{u},\mathbf{x}) \qquad (14)$$

if the matrix $\dfrac{\partial T_k(\mathbf{u},\mathbf{x})}{\partial \mathbf{u}}$ is nonsingular (otherwise a generalized inverse can be used). We call (13), or (14), the **Envelope Theorem** for DP.

The differentiability of the optimal value function was studied in [1] and [8] for various dynamic models. Here we assume differentiability and study consequences of the envelope theorem along the optimal path.

3. Examples and Special Cases

The following examples illustrate the envelope theorem for DP.

Example 1. A DP is said to have **linear dynamics** if the state transformation is linear in both variables,

$$T_k(\mathbf{u},\mathbf{x}) = A_k\mathbf{x} + B_k\mathbf{u} \qquad (15)$$

where A_k, B_k are given matrices, so that

$$\frac{\partial T_k}{\partial \mathbf{x}} = A_k\ ,\ \frac{\partial T_k}{\partial \mathbf{u}} = B_k\ .$$

Assuming B_k is nonsingular, the envelope theorem (14) gives,

$$\nabla V_k(\mathbf{x}) = (I, -A_k B_k^{-1})\nabla f_k(\mathbf{u},\mathbf{x}) \qquad (16)$$

Example 2. A DP is called **linear-quadratic**, or an **LQ problem**, if it has linear dynamics, and **quadratic objective**,

$$f_k(\mathbf{u},\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q_k \mathbf{x} + \frac{1}{2}\mathbf{u}^T R_k \mathbf{u}\ , \qquad (17)$$

then, by (16),

$$\nabla V_k(\mathbf{x}) \;=\; (I, -A_k B_k^{-1})\begin{pmatrix} Q_k\mathbf{x} \\ R_k\mathbf{u}\end{pmatrix}\ ,\ k = 1,\cdots,N-1\ , \qquad (18a)$$

$$\text{and } \nabla V_N(\mathbf{x}) \;=\; Q_N\mathbf{x}\ . \qquad (18b)$$

The envelope theorem (16) can be used to derive the following well known result.

Corollary 1. Given an LQ problem as above, assume the matrices Q_k, R_k are positive definite, for all k. Then the optimal value functions are quadratic,

$$V_k(\mathbf{x}) = \frac{1}{2}\,\mathbf{x}^T \Omega_k \mathbf{x}\ ,\ \ k = 1, \cdots, N\ , \tag{19}$$

and the optimal control $\mathbf{u}$ is linear in the state

$$\mathbf{u} = -(R_k + B_k^T \Omega_{k+1} B_k)^{-1} B_k^T \Omega_{k+1} A_k \mathbf{x}\ ,\ \ k = 1, \cdots, N-1\ , \tag{20}$$

where $\Omega_N = Q_N$ and

$$\Omega_k = Q_k + A_k^T \Omega_{k+1} A_k - A_k^T \Omega_{k+1} B_k \left(R_k + B_k^T \Omega_{k+1} B_k\right)^{-1} B_k^T \Omega_{k+1} A_k\ , \tag{21}$$

for $k = 1, \cdots, N-1$.

Proof. At the $N\,\underline{\text{th}}$ stage,

$$V_N(\mathbf{x}) = \frac{1}{2}\,\mathbf{x}^T Q_N \mathbf{x}$$

proving (19) for $k = N$. We prove (21) for $k < N$ by induction. Suppose V_{k+1} is quadratic,

$$V_{k+1}(\mathbf{x}_+) = \frac{1}{2}\,\mathbf{x}_+^T \Omega_{k+1} \mathbf{x}_+\ . \tag{22}$$

The optimality condition (12a) gives,

$$\begin{aligned} R_k \mathbf{u} + B_k^T \Omega_{k+1} A_k \mathbf{x} + B_k^T \Omega_{k+1} B_k \mathbf{u} &= 0\ ,\ \text{or} \\ \left(R_k + B_k^T \Omega_{k+1} B_k\right) \mathbf{u} + B_k^T \Omega_{k+1} A_k \mathbf{x} &= 0 \end{aligned}$$

proving (20) since $\left(R_k + B_k^T \Omega_{k+1} B_k\right)$ is positive definite.
The envelope theorem (18a) gives

$$\begin{aligned} \nabla V_k(\mathbf{x}) &= Q_k \mathbf{x} + A_k^T \nabla V_{k+1}(A_k \mathbf{x} + B_k \mathbf{u}) \\ &= Q_k \mathbf{x} + A_k^T \Omega_{k+1} \mathbf{x}_{k+1}\ ,\ \text{by}(22) \\ &= Q_k \mathbf{x} + A_k^T \Omega_{k+1} \left(A_k \mathbf{x} + B_k \mathbf{u}\right) \\ &= \left(Q_k + A_k^T \Omega_{k+1} A_k\right) \mathbf{x} + A_k^T \Omega_{k+1} B_k \mathbf{u} \\ &= \left(Q_k + A_k^T \Omega_{k+1} A_k - A_k^T \Omega_{k+1} B_k (R_k + B_k^T \Omega_{k+1} B_k)^{-1} B_k^T \Omega_{k+1} A_k\right) \mathbf{x}\ ,\ \text{by}(20)\ , \end{aligned}$$

proving (19) and (21). □ □

The matrix Ω_k is computed recursively, by (21), and depends on A_k, B_k, Q_k, R_k and Ω_{k+1}. If the matrices A_k, B_k, Q_k, R_k are constant, (21) reduces to the **Riccati equation**

$$\Omega = Q + A^T \Omega A - A^T \Omega B (R + B^T \Omega B)^{-1} B^T \Omega A \tag{23}$$

for the limiting Ω.

Example 3 (Resource allocation)**.** In the **resource allocation** problem, the state at the $k\,\underline{\text{th}}$ stage is the available budget x, the decision is the allocation u to the $k\,\underline{\text{th}}$ project, and the return $f_k(u)$ typically depends only on the allocation, so that

$$\frac{\partial f}{\partial x} = 0\ . \tag{24}$$

The state transformation is

$$T_k(x, u) = x - g_k(u)\ ,\quad 0 \le g_k(u) \le x \tag{25}$$

where $g_k(u)$ is the budget consumed by the allocation u, in particular,

$$T_k(x,u) = x - u \ , \quad 0 \le u \le x \tag{26}$$

if $g_k(u) \equiv u$. From (24) and the envelope theorem (12b) we get

$$V_k'(x) = V_{k+1}'(x_+) \tag{27}$$

which combined with the optimality condition (12a) gives

$$V_k'(x) = f_k'(u)/g_k'(u) \ , \ \text{or} \tag{28}$$

$$V_k'(x) = f_k'(u) \ , \ \text{for the dynamics (26).} \tag{29}$$

Suppose an additional budgetary unit (marginal resource) becomes available at the k th stage. Then (29) seems to suggest that it is optimal to spend the marginal resource at the k th stage. This is a wrong interpretation: (29) merely says that, at optimum, one is indifferent between spending the additional unit at the k th stage, or spending it later.

4. An Inverse Problem

Given the optimal value function V_k, see (11a), we define an **inverse**

$$I_k(v) := \max_{\mathbf{x}} \{ \|\mathbf{x}\| : \ V_k(\mathbf{x}) \le v \} \tag{30}$$

where $\|\mathbf{x}\|$ is a norm of the vector $\mathbf{x}$. The inverse I_k does depend on the norm used, see also [2],[6]. The following properties of the inverse are obvious.

Lemma 1. (a) For all $\mathbf{x}$ such that $\|\mathbf{x}\| = s$,

$$I_k(V_k(\mathbf{x})) \ge s \tag{31}$$

(b) The inverse satisfies

$$I_k(v) = \max\{ \|\mathbf{x}\| : \ \exists\, \mathbf{u} \ni \ \|T_k(\mathbf{u},\mathbf{x})\| \le I_{k+1}(v - f_k(\mathbf{u},\mathbf{x}))\} \tag{32}$$

called an **inverse optimality principle**.

Example 4. Consider a resource allocation problem with objectives $f_k(u)$ and dynamics (25). Then the inverse optimality principle (32) is

$$I_k(v) = \max_{x \ge 0} \{ x : \ \exists\, u \ni \ 0 \le x - g_k(u) \le I_{k+1}(v - f_k(u))\} \tag{33}$$

which can be computed recursively as

$$I_k(v) = \max_{u \ge 0} \{ g_k(u) + I_{k+1}(v - f_k(u))\} \tag{34}$$

with the boundary condition $I_N(v) = f_N^{-1}(v)$.

If the functions I_k are differentiable, then the optimality conditions

$$g_k'(u) - I_{k+1}'(v - f_k(u)) f_k'(u) = 0$$

can be combined with the envelope theorem for (34),

$$I_k'(v) = I_{k+1}'(v - f_k(u))$$

to give

$$f_k'(u) = \frac{g_k'(u)}{I_k'(v)}$$

which, together with (28), gives

$$V_k'(x) = \frac{1}{I_k'(v)} \tag{35}$$

in agreement with the notion of inverse.

Example 5. Consider the LQ problem of Example 2 with

$$V(x) = \frac{1}{2}\mathbf{x}^T\Omega\mathbf{x}$$

by (19), dropping the subscript k. We make repeated use of the fact that Ω is positive definite. Then, by definition (30),

$$\begin{aligned} I(v) &= \max\{\, \|\mathbf{x}\| : V(\mathbf{x}) \leq v \,\} \\ &= \max\left\{\|\mathbf{x}\| : \mathbf{x}^T\Omega\mathbf{x} \leq 2v\right\} \\ &= \max\left\{\|\Omega^{-1/2}\,\mathbf{y}\| : \mathbf{y}^T\mathbf{y} \leq 2v\right\} \end{aligned} \tag{36}$$

where, since Ω is positive definite, its square root $\Omega^{1/2}$ is nonsingular. Using the Euclidean norm in (36), an optimal solution is

$$\mathbf{y}^* = \sqrt{2v}\,\mathbf{w}_n \tag{37}$$

where $\mathbf{w}_n$ is a (normalized) eigenvector of Ω corresponding to its smallest eigenvalue λ_n, which is positive since Ω is positive definite. Therefore

$$I(v) = \|\Omega^{-1/2}\mathbf{y}^*\| = \sqrt{\frac{2v}{\lambda_n}} \tag{38}$$

$$\therefore\ I'(v) = \frac{1}{\sqrt{2\lambda_n v}} \tag{39}$$

On the other hand, for $\mathbf{x}^* = \Omega^{-1/2}\,\mathbf{y}^*$, see (37),

$$\begin{aligned} \nabla V(\mathbf{x}^*) &= \Omega\mathbf{x}^*\ ,\ \text{by (19)} \\ &= \sqrt{2v}\,\Omega^{1/2}\,\mathbf{w}_n \\ \therefore\ \|\nabla V(\mathbf{x}^*)\| &= \sqrt{2\lambda_n v} \end{aligned}$$

a reciprocal of (39), illustrating the reciprocity relation (35) if the state $\mathbf{x}$ is a vector.

Acknowledgments

We thank the referees for constructive comments.

References

[1] A. Araujo and J. A. Scheinkman, *Smoothness, comparative dynamics, and the turnpike property*, Econometrica, Vol. **45** (1977), 601-620.

[2] A. Ben-Israel and S.D. Flåm, *Input optimization for infinite horizon discounted programs*, J. Optimiz. Th. Appl. **61** (1989), 347-357

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

[4] F. H. Clarke, Yu. S. Ledyaev, R. J. Stern and P. R. Wolenski, *Nonsmooth Analysis and Control Theory*, Springer, New York, 1998

[5] A. D. Ioffe, *Proximal analysis and approximate subdifferentials*, J. London Math. Soc. **41** (1990), 175–192

[6] S. Iwamoto, *Inverse theorem in dynamic programming* (parts I, II, III), J. Math. Anal. Appl. **58** (1977), 113-134, 249-279, 439-448

[7] P. Samuelson, *Foundations of Economic Analysis*, Harvard University Press, Cambridge, 1948

[8] J. A. Scheinkman and L. Weiss, *Borrowing constraints and aggregate economic activity*, Econometrica, Vol. **54** (1986), 23-45.

[9] E. Silberberg, *The Structure of Economics, A Mathematical Analysis*, 2nd edition, McGraw-Hill, New York, 1990

[10] H. R. Varian, *Microeconomic analysis*, Norton, New York, 1992

[11] Fuan Zhao, *Topics in Optimization and Optimal Control*, Ph.D. Thesis in Operations Research, Rutgers University (October 1998).

Fuan Zhao, INSTINET, A Reuters Co., 875 Third Ave, New York, NY 10022, USA
E-mail address: fuan_zhao@instinet.com

Adi Ben-Israel, RUTCOR–Rutgers Center for Operations Research, Rutgers University, 640 Bartholomew Rd, Piscataway, NJ 08854-8003, USA
E-mail address: adi.benisrael@gmail.com