On the Principle of Fermat-Lagrange for Mixed Smooth-Convex Extremal Problems

Jan Brinkhuis
Econometrisch Instituut
Erasmus Universiteit Rotterdam
Postbus 1738
3000 DR Rotterdam
The Netherlands
E-mail: brinkhuis@few.eur.nl

Abstract

A necessary condition - the Principle of Fermat-Lagrange - is offered for mixed smooth-convex optimization problems. This generalizes and unifies most of the known necessary conditions for concrete finite and infinite dimensional optimization problems of interest. The new idea in comparison with the unified version of Tikhomirov and others ([I-T], [A-T-F] and [T]) is that a geometrical construction of the principle is given. In the present set-up constraints are not mentioned explicitly, the feasibility set is allowed to vary in a non-standard way and the objective function is also allowed to vary. An equivalent analytical formulation is given as well; we propose a new standard form for optimization problems which allows greater flexibility.

Key Words. Optimization, Lagrange multiplier, principle of Lagrange, Pontrijagin's maximum principle, smooth-convex problems, perturbation function, Banach space, tangent space.

1 Introduction

The initial development of the theory of extremal problems is due to Fermat, Euler and Lagrange. The latter was the first to propose a general principle to analyse extremal problems, a far reaching extension of the method of Fermat. All the necessary conditions for concrete finite and infinite dimensional problems of interest which have appeared in the literature up to the present date are versions of this principle. Sometimes these versions are quite complicated. For example for problems of optimal control one has Pontrijagin's maximum principle. Even the form of this result is not very straightforward. What makes these results possible is the mixed smooth-convex structure of the problems. Tikhomirov and others have given strong unifying versions of these necessary conditions for optimization problems of mixed smooth-convex type ([I-T], [A-T-F] and [T]). We also refer to the work on non-smooth optimization by Clarke, Ioffe, Mordukhovich and Rockafellar ([C], [I], [M], [R]). However the theory of necessary conditions for optimization problems of mixed smooth-convex type is not yet in a satisfactory state. One open problem is how to unify this theory with the theory of sufficient conditions (including the Hamilton-Jacobi-Bellman theory). Another one is how to find satisfactory necessary and sufficient conditions for multidimensional variational problems. Therefore and because of the important role of necessary conditions in the theory and practice of optimization it is of considerable interest to give an exhaustive development of the principle of Lagrange. In the present paper I give a relatively straightforward geometrical realization of the idea of Lagrange, or rather the even older idea of Fermat, to be called the principle of Fermat-Lagrange. Also a new general standard form for optimization problems is proposed which allows greater flexibility than the usual one. For problems in this new standard form an analytical formulation of the principle is given.

Now we sketch the new, geometrical result given in the present paper. It turns out that almost all concrete finite and infinite dimensional extremal problems of interest fit essentially into the following framework. For suitable vector spaces X, U and Y and a suitable map F from $X \times U \times Y$ to $\mathbb{R} \cup \{\infty\}$, the problem to minimize $F(x, u, 0)$ is the given problem. Moreover there is a meaningful way to introduce a parameter from Y into the problem such that for each choice of parameter $y \in Y$ one gets the problem to minimize $F(x, u, y)$. Furthermore certain conditions on F are usually satisfied.The principle of Lagrange can be realized as follows. For each solution $(\hat{x}, \hat{u})$ of the given problem there exists an affine function A on $X \times U \times Y$ such that $A(x, \hat{u}, y)$ is a linearization in $(\hat{x}, 0)$ of $F(x, \hat{u}, y)$ in the spirit of smooth analysis, $A(\hat{x}, u, y)$ is a linearization in $(\hat{u}, 0)$ of $F(\hat{x}, u, y)$ in the spirit of convex analysis and $(\hat{x}, \hat{u})$ is a solution of the problem to minimize $A(x, u, 0)$. Of course then $A(x, u, 0)$ is constant. Thus the principle is presented as the result that the property of $(\hat{x}, \hat{u})$ to minimize $F(x, u, 0)$ is preserved under linearization. This way of stating the principle is clearly a straightforward generalization of the theorem of Fermat that the derivative of a smooth function in an optimum is zero. The slope of A in the Y-direction is a generalization of the usual Lagrange multiplier, also called shadow price or dual variable. The result generalizes and simplifies the results on mixed problems mentioned above. The special cases of purely convex (resp. smooth) linearization were dealt with in [B1] and [B2] (resp. in [B3]).

Finally we mention the following additional novel features of the present realization of Lagrange's principle: constraints are not mentioned explicitly, the feasibility set is allowed to vary in a non-standard way and the objective function is also allowed to vary. In all existing realizations the only perturbation allowed is a standard perturbation of equality constraints.

For some concrete problems the principle of Lagrange has been used as a heuristic principle: in cases where it is not known whether the condition provided by the principle of Lagrange is necessary for optimality. In combination with other methods this has led to a complete, rigorous solution. Therefore it is of interest to emphasize that the condition of Fermat-Lagrange is defined in the present paper in great generality compared with previous versions of the principle of Lagrange.

I would like to acknowledge an inspiring tutorial on extremal problems by Professor Vladimir Tikhomirov at the foot of the Euromast. I thank him also for suggesting an improvement of the analytical formulation of the Principle of Fermat-Lagrange.

2 Statement of the results.

In this section we give a precise formulation of the results of this paper. For each function h from a set S to $\mathbb{R} \cup \{\infty\}$ we let Γ_h be the finite part of the graph of h, that is $\Gamma_h = \{(s, r) \in S \times \mathbb{R} \mid r = h(s)\}$. We let $\mathrm{epi}(h)$ be the epigraph of h, that is $\mathrm{epi}(h) = \{(s, r) \in S \times \mathbb{R} \mid r \geq h(s)\}$. A vector x from a normed space X is said to be a tangent vector to a subset M of X in a point m of M if there is a map r from the interval $[-1,1]$ to X such that $m+\lambda x+r(\lambda)\in M$ for all $\lambda\in[-1,1]$ and $|r(\lambda)|=o(|\lambda|)$ for $\lambda\to 0$. The set T_mM of all tangent vectors to the set M in the point m is called the tangent space to M in m.

The set-up is as follows. X and Y are Banach spaces, U is a set provided with the trivial topology, F is a function from the cartesian product $X\times U\times Y$ to $\mathbb{R}\cup\{\infty\}$, called a *perturbation function*, and $(\hat{x},\hat{u})$ is an element of $X\times U$ with $F(\hat{x},\hat{u},0)\in\mathbb{R}$. We write $\hat{r}=F(\hat{x},\hat{u},0)$.

Let $F(\hat{u})$ be the function from $X\times Y$ to $\mathbb{R}\cup\{\infty\}$ defined by $F(\hat{u})(x,y)=F(x,\hat{u},y)$ for all (x,y) in $X\times Y$. For all $x\in X$ let $F(x)$ be the function from $U\times Y$ to $\mathbb{R}\cup\{\infty\}$ defined by $F(x)(u,y)=F(x,u,y)-\hat{r}$ for all (u,y) in $U\times Y$. Let σ (resp. τ) be the natural projection from $X\times Y\times\mathbb{R}$ (resp. $U\times Y\times\mathbb{R}$) onto $Y\times\mathbb{R}$. Let S and C be the subsets of $Y\times\mathbb{R}$ defined by $S=\sigma(T_{(\hat{x},0,\hat{r}),\ F(\hat{u})})$ and $C=\tau(\mathrm{epi}F(\hat{x}))$.

Let α be a non-trivial element of the dual space $(Y\times\mathbb{R})^*$. We say that *the condition of Fermat-Lagrange* holds for F and $(\hat{x},\hat{u})$ with multiplier α if $\alpha(S)=0$ and $\alpha(C)\subseteq\mathbb{R}_+$. We write $\alpha=(\eta,\eta_0)$ with $\eta\in Y^*$ and $\eta_0\in\mathbb{R}$. It follows readily from $\alpha(C)\subseteq\mathbb{R}_+$ that $\eta_0\geq 0$. Let π be the projection from $Y\times\mathbb{R}$ onto Y. We shall say that the *regularity condition holds* if the origin 0_Y lies in the interior of the subset $\pi(S+C)$ of Y.

Let k be a map from an open subset W of a Banach space A to another Banach space B and let $v\in W$. Then k is said to be strictly differentiable in v if there is a continuous linear map L from A to B such that $|k(s)-k(t)-L(s-t)|=o(|s-t|)$ for $s\to v$ and $t\to v$. Then L is uniquely determined. It is called the strict derivative of k in v and it is denoted by $k'(v)$.

Now we formulate a list of assumptions on F and $(\hat{x},\hat{u})$.

Assumptions 2.1. There is a Banach space Z, a neighbourhood V of $\hat{x}$ in X and a map g from $V\times U\times Y\times\mathbb{R}$ to Z such that the following conditions hold:

i , $_F\cap(V\times U\times Y\times\mathbb{R})=$ zero (g), the zero-set of g.

ii $g(x,u,y,r)$ depends strictly differentiable on (x,y,r).

iii $g'_{(y,r)}\ (\hat{x},u,y,r)$ is surjective for all $(\hat{x},u,y,r)\in$ zero g

Remark: By the tangent space theorem (3.1) these conditions imply that S is a closed linear subspace of $Y\times\mathbb{R}$.

iv $\dim(Y \times \mathbb{R}/S) < \infty$.

v $\tau\,(\text{epi}\, F(x))$ is a convex subset of $Y \times \mathbb{R}$ for all $x \in V$.

Now we come to the main result.

Theorem 2.2. *Principle of Fermat-Lagrange for mixed smooth-convex extremal problems (geometrical formulation). Let F and $(\hat{x}, \hat{u})$ be as above. Make the assumptions from (2.1). Assume that $(\hat{x}, \hat{u})$ is a local minimum of the function $F(x, u, 0)$. Then there is a non-trivial element $\alpha = (\eta, \eta_0)$ in $(Y \times \mathbb{R})^*$ such that the condition of Fermat-Lagrange holds for F and $(\hat{x}, \hat{u})$ with multiplier α. One has $\eta_0 > 0$ for each such α if and only if the regularity condition holds.*

Remark 2.3. Now we show what is the relation between this result and the preliminary description of it given in section 1. Assume U is equipped with a vector space structure and let an affine function A on $X \times U \times Y$ which depends continuously on Y be given. A is called a mixed smooth-convex linearization of F in $(\hat{x}, \hat{u}, 0)$ if

$$F(\hat{x}, \hat{u}, 0) = A(\hat{x}, \hat{u}, 0),$$

$$T_{(\hat{x},0,\hat{r}),\ F(\hat{u})} \subseteq T_{(\hat{x},0,\hat{r}),\ A(\hat{u})}$$

and

$$\text{epi}\ F(\hat{x}) \subseteq \text{epi}\ A(\hat{x}).$$

Here $A(\hat{u})$ (resp. $A(\hat{x})$) is defined in the same way as $F(\hat{u})$ (resp. $F(\hat{x})$). Let $\alpha \in (Y \times \mathbb{R})^*$ be defined by

$$\alpha(y, r) = r + \hat{r} - A(\hat{x}, \hat{u}, y) \quad \forall y \in Y \quad \forall r \in \mathbb{R}.$$

Then the condition of Fermat-Lagrange holds for F and $(\hat{x}, \hat{u})$ with multiplier α if and only if A is a mixed smooth-convex linearization of F in $(\hat{x}, \hat{u}, 0)$ and if moreover $A(x, u, 0)$ is minimal in $(\hat{x}, \hat{u})$.

For use in the analysis of concrete extremal problems we give also an analytical formulation of theorem 2.2. If A and B are normed vectorspaces and $\Lambda : A \to B$ is a continuous linear operator, then the adjoint operator $\Lambda^* : B^* \to A^*$ is defined by $< \Lambda^* b^*, a > = < b^*, \Lambda a >$ for all $b^* \in B^*$ and $a \in A$. We will write $\hat{g}'_x$ for $g'_x(\hat{x}, \hat{u}, 0, \hat{r})$; in the same way we will use the notation $\hat{g}'_{y,r}$ and $\hat{g}'_r$.

Proposition 2.4. *Principle of Fermat-Lagrange (analytical formulation). Let F and $(\hat{x}, \hat{u})$ be as above. Make the assumptions from (2.1). Assume that $(\hat{x}, \hat{u})$ is a local minimum of the problem to minimize $r \in \mathbb{R}$ under the constraint that $g(x, u, 0, r) = 0$ for suitable $x \in X$ and $u \in U$. Then there is a non-trivial element $\zeta \in Z^*$ such that*

$$\hat{g}_x^{'*}\zeta = 0 \text{ and } \min_{c \in C}(\hat{g}_{y,r}^{'*}\zeta)(c) = 0.$$

Moreover $\hat{g}_r^{'*}\zeta > 0$ for each such that ζ if and only if the interior of the following subset of the image $\hat{g}'_{x,y,r}$ contains the origin

$$\{\hat{g}'_{x,y,r}(0, y, 0)\,|\exists u \in U \exists r \in \mathbb{R} \;\text{ such that } g(\hat{x}, u, y, r) = 0\}.$$

By a straightforward specialization of this result one obtains the following consequence. This is essentially theorem (P) from [T] (see p.48). In [T] it is made clear that theorem (P) unifies most of the known necessary conditions for concrete extremal problems.

Corollary 2.5. *Let X, Y be Banach spaces, V an open subset of X, U a set, f (resp.h) a map from $V \times U$ to $\mathbb{R}$ (resp. to Y). Let $(\hat{x}, \hat{u})$ be a local minimum of the problem to minimize $f(x, u)$ under the restriction $h(x, u) = 0$. Let the Lagrange function be defined by $\mathcal{L}(x, u, \eta, \eta_0) = \eta_0 f(x, u) + \eta(h(x, u))$ for all $x \in V, u \in U, \eta \in Y^*$ and $\eta_0 \in \mathbb{R}$. We make the following assumptions:*

1. *$f(x, u)$ and $h(x, u)$ are strictly differentiable in the variable x.*
2. *the image of the derivative $h'_x(\hat{x}, \hat{u})$ is a closed subspace of Y of finite codimension.*
3. *the subset $\{(h(x, u), f(x, u) + t) \mid u \in U, t \geq 0\}$ of $Y \times \mathbb{R}$ is convex for all $x \in V$.*

Then there exist $\hat{\eta} \in Y^*$ and $\hat{\eta}_0 \in \mathbb{R}_+$, not both zero, such that the following two conditions hold:

(i) Stationarity condition. The vector $\hat{x}$ is stationary for the problem to minimize $\mathcal{L}(x, \hat{u}, \hat{\eta}, \hat{\eta}_0)$. This is equivalent to

$$\hat{\eta}_0 f'_x(\hat{x}, \hat{u}) + \hat{\eta} \circ h'_x(\hat{x}, \hat{u}) = 0.$$

(ii) Minimum principle. The element $\hat{u}$ is a solution of the problem to minimize $\mathcal{L}(\hat{x}, u, \hat{\eta}, \hat{\eta}_0)$.

Moreover each such $\hat{\eta}_0$ is nonzero if and only if the interior of $h(\hat{x}, U)$ as a subset of Y contains the origin.

3 Proof of the results

In this section we prove theorem (2.2). The proposition, the corollary and the statement in remark (2.3) are straightforward consequences taking into account the theorem on the tangent space, which we recall below. This is also the main tool in the proof of theorem (2.2). Proposition 2.4 follows from theorem 2.2 by virtue of theorem 3.1 and the lemma on the annihilator of the kernel of a regular operator (see p.81 in [A-T-F]).

Theorem 3.1. *([A-T-F] p. 109). Let X and Y be Banach spaces, let $x_0 \in X$, let U be a neighbourhood of x_0 in X and let F be a map from U to Y with $F(x_0) = 0$. Assume that F is strictly differentiable in the point x_0 and that the linear map $F'(x_0)$ is surjective. Then the zero set of F has as tangent space in the point x_0 the kernel of $F'(x_0)$.*

Proof of Theorem (2.2). Let F and $(\hat{x}, \hat{u})$ be as in section 2, assume that $(\hat{x}, \hat{u})$ is a local minimum of the function $F(x, u, 0)$ and let Z, V and g be given such that the assumptions from (2.1) hold.

We carry out the proof of the theorem in three steps.

Step 1. The theorem holds if the regularity condition is not satisfied.

Proof of step 1. We use the standard separation result that in a finite dimensional real vector space any convex set can be separated from any point which does not lie in its interior. Then, by virtue of assumptions (2.1)(iv) and (2.1)(v), it follows that there exists a non-trivial linear function β on $Y \times \mathbb{R}/(S + 0 \times \mathbb{R})$ with $\beta(t) \geq 0$ for all t in the image of C under the natural projection from $Y \times \mathbb{R}$ to $Y \times \mathbb{R}/(S + 0 \times \mathbb{R})$. By composition of β with this projection one gets a non-trivial continuous linear function α on $Y \times \mathbb{R}$ with $\alpha(S) = 0$, $\alpha(0 \times \mathbb{R}) = 0$ and $\alpha(C) \subseteq \mathbb{R}_+$. This proves the theorem in the present case. □

From now on we assume, as we may, that the regularity condition holds.

Step 2: if $(\overline{y}, \overline{r}) \in S$ and $(\overline{y}, \overline{\overline{r}}) \in C$, then $\overline{\overline{r}} \geq \overline{r}$.

Proof of step 2. Let $(\overline{y},\overline{r}) \in S$ and $(\overline{y},\overline{\overline{r}}) \in C$ be given. Choose $\overline{x} \in X$ with $(\overline{x},\overline{y},\overline{r}) \in T_{(\hat{x},0,\hat{r}),\ F(\hat{u})}$ and $\overline{u} \in U$ with $(\overline{u},\overline{y},\overline{\overline{r}}) \in \text{epi}F(\hat{x})$. We assume, as we may for the proof, that $(\overline{u},\overline{y},\overline{\overline{r}}) \in \ ,\ _{F(\hat{x})}$, that is, $\hat{r}+ \overline{\overline{r}}= F(\hat{x},\overline{u},\overline{y})$, or equivalently $g(\hat{x},\overline{u},\overline{y},\hat{r}+ \overline{\overline{r}}) = 0$. Let m be the dimension of $Y\times\mathbb{R}/(S+0\times\mathbb{R})$. Identify $Y \times \mathbb{R}/(S + 0 \times \mathbb{R})$ with $\mathbb{R}^m$. Let $z_1,\ldots,z_{2^m}$ be the corners of a nontrivial cube T in $\mathbb{R}^m$ with centre at the origin which is entirely located in the image of C in $Y \times \mathbb{R}/(S + 0 \times \mathbb{R})$. This is possible by the regularity condition.

These vectors have the following properties, clearly.

(α) $\displaystyle\sum_{i=1}^{2^m} z_i = 0$

(β) $\{z_i\}_{i=1}^{2^m}$ generates the whole vector space $\mathbb{R}^m$.

(γ) For each $i \in \{1,\ldots,2^m\}$ there exist $\tilde{u}_i \in U, \tilde{y}_i \in Y, \tilde{r}_i \in \mathbb{R}$ such that $F(\hat{x},\tilde{u}_i,\tilde{y}_i) = \tilde{r}_i$ - that is, $g(\hat{x},\tilde{u}_i,\tilde{y}_i,\tilde{r}_i) = 0$ - and such that the image of $(\tilde{y}_i,\tilde{r}_i)$ in $Y \times \mathbb{R}/(S + 0 \times \mathbb{R})$ equals z_i.

Choose such elements $\tilde{u}_i,\tilde{y}_i,\tilde{r}_i$ for all $i \in \{1,\ldots,2^m\}$. We define a function μ on $\mathbb{R}^{2^m+2}$ by $\mu(\alpha_0,\alpha_0',\alpha_1,\ldots,\alpha_{2^m}) = \alpha_0 + 2^m\alpha_0' + \displaystyle\sum_{i=1}^{2^m}\alpha_i$ for all $(\alpha_0,\alpha_0',\alpha_1,\ldots,\alpha_{2^m}) \in \mathbb{R}^{2^m+2}$. We define a map $\Psi = (\Psi_i)_{i=0,\ldots,2^m+2}$ from the cartesian product $\mathbb{R}^{2^m+2} \times V \times (Y \times \mathbb{R})^{2^m+2}$ to $Y \times Z^{2^m+2}$ by

$$\left[\Psi(\alpha,x,(y_j,r_j)_{j=1}^{2^m+2})\right]_i =$$

$$\begin{array}{lll} = & (1-\mu(\alpha))y_{2^m+1} + \sum_{j=1}^{2^m}(\alpha_0'+\alpha_j)y_j + \alpha_0 y_{2^m+2} & \text{if} \quad i = 0 \\ = & g(x,\tilde{u}_i,y_i,r_i) & \text{if} \quad 1 \le i \le 2^m \\ = & g(x,\hat{u},y_{2^m+1},r_{2^m+1}) & \text{if} \quad i = 2^m+1 \\ = & g(x,\overline{u},y_{2^m+2},r_{2^m+2}) & \text{if} \quad i = 2^m+2 \end{array}$$

for all $\alpha = (\alpha_0,\alpha_0',\alpha_1,\ldots,\alpha_{2^m}) \in \mathbb{R}^{2^m+2}, x \in V$ and $(y_j,r_j) \in Y \times \mathbb{R} \quad \forall j \in \{1,\ldots,2^m+2\}$.

We observe that the vector $w = (0,\hat{x},(\tilde{y}_i,\tilde{r}_i)_{i=1}^{2^m},(0,\hat{r}),(\overline{y},\hat{r}+ \overline{\overline{r}}))$ lies in the zero-set of Ψ.

Claim: T_w zero$(\Psi) =$ Ker $\Psi'(w)$.

Proof of the claim: by theorem (3.1) it suffices to show that $\Psi'(w)$ is surjective. We choose $(\check{y}, \check{r}) \in S$ with $\sum_{i=1}^{2^m} \tilde{y}_i = \check{y}$, as we may by (α) and (γ). From the definition of Ψ one gets immediately that
$\Psi'(w)(\alpha, x, (y_i, r_i)_{i=1}^{2^m+2})$ is the vector v which is given componentwise as follows: $v_i =$

$$\begin{array}{lll} = & y_{2^m+1} + \alpha_0' \check{y} + \sum_{j=1}^{2^m} \alpha_j \tilde{y}_j + \alpha_0 \overline{y} & \text{if} \quad i = 0 \\ = & g'(\hat{x}, \tilde{u}_i, \tilde{y}_i, \tilde{r}_i)(x, y_i, r_i) & \text{if} \quad 1 \le i \le 2^m \\ = & g'(\hat{x}, \hat{u}, 0, \hat{r})(x, y_{2^m+1}, r_{2^m+1}) & \text{if} \quad i = 2^m + 1 \\ = & g'(\hat{x}, \overline{u}, \overline{y}, \hat{r} + \overline{\overline{r}})(x, y_{2^m+2}, r_{2^m+2}) & \text{if} \quad i = 2^m + 2. \end{array}$$

Here g' stands for $g'_{x,y,r}$. Now choose arbitrarily $y^{(0)} \in Y, r^{(1)} \in Z^{2^m}, r^{(2)} \in Z$ and $r^{(3)} \in Z$. We are going to show that by an appropriate choice of $(\alpha, x, (y_i, r_i)_{i=1}^{2^m+2})$ one can ensure that v equals $(y^{(0)}, r^{(1)}, r^{(2)}, r^{(3)})$.
To begin with we start with the choice $(\alpha, x, (y_i, r_i)_{i=1}^{2^m+2}) = 0$; we are going to change this choice step by step until we achieve our goal. We use assumption (2.1)(iii) to change the choices of $y_{2^{m+1}}$ and r_{2^m+1} in such a way that the 'third coordinate' equals $r^{(2)}$. Then we use (β) and (γ) to change the choices of the α_i $(1 \le i \le 2^m)$ in such a way that the first coordinate equals $y^{(0)}$ up to an element from the image of S under the natural projections from $Y \times \mathbb{R}$ to Y. This image equals the image of Ker $g'(\hat{x}, \hat{u}, 0, \hat{r})$ under the natural projections from $X \times Y \times \mathbb{R}$ to Y as a consequence of theorem 3.1. So by changing $(x, y_{2^m+1}, r_{2^{m+1}})$ suitably we can achieve that the first coordinate equals $y^{(0)}$ and the third one remains $r^{(2)}$. Finally, by assumption (2.1)(iii), we can change the (y_i, r_i) for $1 \le i \le 2^m$ and $i = 2^m + 2$ in such a way that the second coordinate becomes $r^{(1)}$ and the last one $r^{(3)}$. This finishes the proof of the claim. □

Now we continue the proof of step 2. As $(\check{y}, \check{r}) \in S$ there exists $\check{x} \in X$ with $g'(\hat{x}, \hat{u}, 0, \hat{r})(\check{x}, \check{y}, \check{r}) = 0$. We choose an arbitrary element $(\alpha_0', \alpha_0) \in \mathbb{R}^2$. We define the element $(x, y_{2^m+1}, r_{2^m+1})$ in $X \times Y \times \mathbb{R}$ to be the linear combination $-\alpha_0'(\check{x}, \check{y}, \check{r}) - \alpha_0(\overline{x}, \overline{y}, \overline{r})$. We choose $\alpha_i = 0 \;\; \forall i = 1, \ldots, 2^m$. We choose (y_i, r_i) such that $g'(\hat{x}, \tilde{u}_i, \tilde{y}_i, \tilde{r}_i)(x, y_i, r_i) = 0$ for all $i \in \{1, \ldots, 2^m\}$. We choose (y_{2^m+2}, r_{2^m+2}) such that $g'(\hat{x}, \overline{u}, \overline{y}, \hat{r} + \overline{\overline{r}})(x, y_{2^m+2}, r_{2^m+2}) = 0$. These choices can be made by assumption (2.1)(iii). Set $\alpha = (\alpha_0, \alpha_0', \alpha_1, \ldots, \alpha_{2^m})$. One readily verifies, using also the claim, that $(\alpha, x, (y_i, r_i)_{i=1}^{2^m+2})$ lies in the tangent space in w to the zero set of Ψ.

Therefore, by the definition of the tangent space, there exist o-functions $\rho_0(\cdot), \rho_0'(\cdot), \rho_i(\cdot)$ $1 \leq i \leq 2^m$ from $[-1,1]$ to $\mathbb{R}$,
(we write $\rho = (\rho_0, \rho_0', \rho_1, \ldots, \rho_{2^m})$), $r(\cdot)$ from $[-1,1]$ to V, $\sigma_i(\cdot)$ $1 \leq i \leq 2^m+2$ from $[-1,1]$ to Y, $\tau_i(\cdot)$ $1 \leq i \leq 2^m+2$ from $[-1,1]$ to $\mathbb{R}$ such that

$$0 = \Psi[w + \lambda(\alpha, x, (y_i, r_i)_{i=1}^{2^m+2}) \quad + \quad (\rho(\lambda), r(\lambda), (\sigma_i(\lambda), \tau_i(\lambda))_{i=1}^{2^m+2})], \quad \forall \lambda \in [-1,1]. \tag{1}$$

Then for all $\lambda \in [-1,1]$ the following elements all lie in $\tau(\text{epi } F(\hat{x}+\lambda x+r(\lambda)))$:

$$t_i = (\tilde{y}_i + \lambda y_i + \sigma_i(\lambda), \tilde{r}_i - \hat{r} + \lambda r_i + \tau_i(\lambda)) \quad \forall i \in \{1, \ldots, 2^m\},$$

$$t_{2^m+1} = (\lambda y_{2^m+1} + \sigma_{2^m+1}(\lambda), \lambda r_{2^m+1} + \tau_{2^m+1}(\lambda))$$

and

$$t_{2^m+2} = (\overline{y} + \lambda y_{2^m+2} + \sigma_{2^m+2}(\lambda), \overline{\overline{r}} + \lambda r_{2^m+2} + \tau_{2^m+2}(\lambda)).$$

Now choose $\varepsilon > 0$ such that $F(x, u, 0) \geq F(\hat{x}, \hat{u}, 0)$ for all $(x, u) \in X \times U$ with $|x - \hat{x}| < \varepsilon$.
Choose $\alpha_0 = 1$ and choose an arbitrary $\alpha_0' > 0$. Then there exists a $\lambda^{(1)} = \lambda_1(\alpha_0') \in <0, 1>$ such that for all $\lambda \in <0, \lambda^{(1)}>$:

(i) $|(\hat{x} + \lambda x + r(\lambda)) - \hat{x}| < \varepsilon$

(ii) $\gamma_{2^m+1} = 1 - \lambda(\alpha_0 + 2^m \alpha_0') - \rho_0(\lambda) - 2^m \rho_0'(\lambda) - \sum_{i=1}^{2^m} \rho_i(\lambda) \in \quad <0, 1>$

(iii) $\gamma_i = \lambda \alpha_0' + \rho_0'(\lambda) + \rho_i(\lambda) \in <0, 1> \quad \forall i \in \{1, \ldots, 2^m\}$

(iv) $\gamma_{2^m+2} = \lambda \alpha_0 + \rho_0(\lambda) \in <0, 1>$

By assumption (2.1)(v) it follows that for all $\lambda \in <0, \lambda^{(1)}>$ the following element lies in $\tau(\text{epi } F(\hat{x} + \lambda x + r(\lambda)))$:

$$\sum_{i=1}^{2^m+2} \gamma_i t_i$$

It is readily verified that its Y-coordinate is 0: this is essentially the first component of the vector equation (1).

Using the local minimality of $(\hat{x}, \hat{u})$ it follows that the second coordinate of this element is ≥ 0. Now divide this second coordinate by λ and take the limit $\lambda \downarrow 0$:

$$r_{2^m+1} + \alpha_0' \sum_{i=1}^{2^m} (\tilde{r}_i - \hat{r}) + \overline{\overline{r}} \geq 0.$$

We recall that $r_{2^m+1} = -\alpha_0' \check{r} - \overline{r}$ and that $\alpha_0' > 0$ is arbitrary.
Taking the limit $\alpha_0' \downarrow 0$ one gets $\overline{\overline{r}} \geq \overline{r}$. This concludes the proof of step 2. □

Step 3. The theorem holds under the regularity condition.

Proof of step 3. Let i be the natural projection from $Y \times \mathbb{R}$ to $Y \times \mathbb{R}/S$. We begin by proving that $i(C)$ and $i(0 \times < -\infty, 0 >)$ are disjoint convex subsets of $Y \times \mathbb{R}/S$. The convexity statement follows from assumption (2.1)(v). To prove disjointness, we argue by contradiction. Assume $i(C)$ and $i(0 \times < -\infty, 0 >)$ are not disjoint. Then there exist $(\overline{y}, \overline{\overline{r}}) \in C$, $r < 0$, $(\overline{y}, \overline{r}) \in S$ with $(\overline{y}, \overline{\overline{r}}) = (0, r) + (\overline{y}, \overline{r})$. Therefore $\overline{\overline{r}} < \overline{r}$. On the other hand, by step 2, $\overline{\overline{r}} \geq \overline{r}$. This is the required contradiction. We apply to $i(C)$ and $i(0 \times < -\infty, 0 >)$ the well-known fact that in a finite dimensional vectorspace two disjoint convex sets can be separated by a hyperplane. It follows that there is a nontrivial linear function β on $Y \times \mathbb{R}/S$ such that $\beta(i(c)) \geq \beta(i(0, r))$ for all $c \in C$ and all $r < 0$. Taking the limit $r \uparrow 0$ we get $\beta(i(c)) \geq 0$ for all $c \in C$. Now we consider the nontrivial continuous linear function α on $Y \times \mathbb{R}$ defined by $\alpha = \beta o i$. For each $s \in S$ one has $\alpha(s) = \beta(i(s)) = \beta(0) = 0$. Moreover for each $c \in C$ one has $\alpha(c) = \beta(i(c)) \geq 0$. Therefore the condition of Fermat-Lagrange holds for F and $(\hat{x}, \hat{u})$ with multiplier α. We write $\alpha = (\eta, \eta_0)$ with $\eta \in Y^*$ and $\eta_0 \in \mathbb{R}$. To prove $\eta_0 > 0$ we argue by contradiction. Assume $\eta_0 = 0$. Take any $y \in Y$. By the regularity condition there exist $\varepsilon > 0$, $(c_1, c_2) \in C$, $(s_1, s_2) \in S$ and $t \in \mathbb{R}$ such that $(\varepsilon y, 0) + (s_1, s_2) + (0, t) = (c_1, c_2)$. Applying η we get - using $\alpha(s) = 0$ - that $\varepsilon \eta(y) = \eta(c_1) \geq 0$, so $\eta(y) \geq 0$. As $y \in Y$ is arbitrary and $\eta \neq 0$ we get a contradiction. □

References

[A-T-F] V.M. Alekseev, V.M. Tikhomirov and S.V. Fomin, *Optimal Control*, New York, 1987.

[B1] J. Brinkhuis, *Vertical halfspaces as solutions of dual extremal problems*, Econometric Institute Rotterdam, report 9540/B, 1995.

[B2] J. Brinkhuis, *Extremal problems and inclusion of perturbations in half-spaces*, Econometric Institute Rotterdam, report 9541/B, 1995.

[B3] J. Brinkhuis, *On a geometrical construction of the multiplier rule*, Econometric Institute Rotterdam, report 9721/B, 1997.

[C] F.H. Clarke, *Optimization and nonsmooth Analysis*, New York, John Wiley, 1983.

[I] A.D. Ioffe, *Euler-Lagrange and Hamiltonian Formalisms in Dynamic Optimization*, Trans. AMS

[I-T] A.D. Ioffe, V.M. Tikhomirov, *Theory of extremal problems*, Amsterdam, 1979.

[M] B. Mordukhovich, *Approximation Methods in Problems of Optimization and Control*, Nauka, Moscow, 1998 (Russian); translation to appear in Wiley Interscience.

[R] R.T. Rockafellar, *Equivalent subgradient versions of Hamiltonian and Euler-Lagrange equations in variational analysis*, SIAM, J Control Opt. (to appear).

[T] V.M. Tikhomirov, *Fundamental Principles of the Theory of Extremal Problems*, New York, 1986.