Topology of Lagrange multipliers

C. Udrişte, M. Constantinescu, I. Ţevy, O. Dogaru

Abstract. We re-discuss the well-known programs with holonomic constraints insisting on the following issues: (i) the Lagrange-dual problem with weak respectively strong duality; (ii) the Wolfe-dual problem; (iii) the topology of Lagrange multipliers; (iv) the interpretation of Lagrange multipliers; (v) pertinent examples.

M.S.C. 2010: 49K35, 58A17.
Key words: optimization; dual problems; topology of Lagrange multipliers; meaning of Lagrange multipliers.

1 Classical Lagrange and Wolfe dual programs

In this Section, we re-discuss the well-known programs with holonomic constraints insisting on the following issues [1], [3], [4]: (i) the Lagrange-dual problem with weak respectively strong duality; (ii) the Wolfe-dual problem.

1.1 The Lagrange dual problem

Let D be a domain in R^n, let $x = (x^1, ..., x^n)$ be a point in D and $f : D \to R$ and $g_\alpha : D \to R$ be convex functions. Denote $g = (g_\alpha)$ and we introduce the set

$$\Omega = \{x \in D \,|\, g_\alpha(x) \leq 0,\, \alpha = 1, ..., m\} = \{x \in D \,|\, g(x) \preceq 0\}.$$

A complete notation for this set is $(\Omega, g, \preceq)$, but for short the sign $\preceq$ or the pair $(g, \preceq)$ are suppressed in the notation. Let us consider the *convex program*

$$(P) \qquad \min_x \{f(x) \,|\, x \in \Omega\}.$$

The *Lagrange function (or Lagrangian)* of (P)

$$L(x, \lambda) = f(x) + \sum_{\alpha=1}^{m} \lambda_\alpha g_\alpha(x) = f(x) + < \lambda, g >, x \in D, \lambda \succeq 0$$

BSG Proceedings 22. The International Conference "Differential Geometry - Dynamical Systems" DGDS-2014, September 1-4, 2014, Mangalia-Romania, pp. 84-97.

is convex in x and linear in λ.

Remark *We can create an original Riemannian geometry on the set of critical points, using similar ideas we shall develop in a further 3.2.1.*

For all $\lambda \succeq 0$, the inequality

$$\sum_{\alpha=1}^{m} \lambda_\alpha g_\alpha(x) \leq 0, \ \forall x \in \Omega$$

holds. Consequently

$$L(x,\lambda) \leq f(x), \ \forall x \in \Omega, \ \forall \lambda \succeq 0. \tag{1}$$

The equality holds iff (complementarity conditions)

$$\lambda_\alpha g_\alpha(x) = 0 \,(\text{for each } \alpha = 1, ..., m).$$

Let us introduce the *Lagrange dual function*

$$\psi(\lambda) = \inf_{x \in D}\{f(x) + \sum_{\alpha=1}^{m} \lambda_\alpha g_\alpha(x), \ x \in D, \ \lambda \succeq 0\}.$$

This function $\psi(\lambda)$ is concave, because it is a point-wise infimum of affine functions. Indeed, using the linearity of $L(x,\lambda)$ with respect to λ, and introducing $\lambda^1 \succeq 0$, $\lambda^2 \succeq 0$, and $0 \leq t \leq 1$, we have

$$\psi(t\lambda^1 + (1-t)\lambda^2) = \inf_{x \in D} L(x, t\lambda^1 + (1-t)\lambda^2)$$

$$= \inf_{x \in D}(tL(x,\lambda^1) + (1-t)L(x,\lambda^2)) \geq \inf_{x \in D}(tL(x,\lambda^1)) + \inf_{x \in D}((1-t)L(x,\lambda^2))$$

$$= t \inf_{x \in D} L(x,\lambda^1) + (1-t) \inf_{x \in D} L(x,\lambda^2) = t\psi(\lambda^1) + (1-t)\psi(\lambda^2).$$

Definition 1.1. The problem

$$\sup_{\lambda} \{\psi(\lambda) \,|\, \lambda \succeq 0\}$$

is the so-called *Lagrange dual problem* of (P).

The Lagrange dual problem can be called convex because it is equivalent to the convex problem

$$\inf_{\lambda} \{-\psi(\lambda) \,|\, \lambda \succeq 0\}.$$

The Lagrange-dual problem is also defined in this way if (P) is not convex. The following theorem holds also in that case.

Theorem 1.1. *(***weak duality***) The dual function yields lower bounds of the initial optimal value f_*, i.e., for any λ, we have $\psi(\lambda) \leq f_*$. In other words,*

$$\sup_{\lambda} \{\psi(\lambda) \,|\, \lambda \succeq 0\} \leq \inf_{x \in \Omega}\{f(x)\}.$$

Proof In the foregoing statements, we have the relation (1). Since $\Omega \subset D$, for each $\lambda \succeq 0$, we find

$$\psi(\lambda) = \inf_{x \in D} L(x, \lambda) \leq \inf_{x \in \Omega} L(x, \lambda) \leq \inf_{x \in \Omega} f(x).$$

Thus the statement in the theorem is true.

The problem of finding the best lower bound on f_* obtained from the Lagrange dual function is called the *Lagrange dual problem* for the original or primal problem.

The optimal values may be different. However, they are equal if (P) satisfies the Slater condition and has finite optimal value. This is the next result.

Theorem 1.2. (strong duality) *If the program (P) satisfies the Slater condition and has finite optimal value, then*

$$\sup_{\lambda} \{\psi(\lambda) \mid \lambda \succeq 0\} = \inf_{x \in D} \{f(x) \mid g(x) \preceq 0\}.$$

Moreover, then the dual optimal value is attained.

Proof Denote by f_* the optimal value of (P). Taking $a = f_*$ in the Convex Farkas Lemma, it follows that there exists a vector $\lambda_* = (\lambda_{1*}, ..., \lambda_{m*}) \geq 0$ such that

$$L(x, \lambda_*) = f(x) + \sum_{\alpha=1}^{m} \lambda_{\alpha *}\, g_\alpha(x) \geq f_*, \forall x \in D.$$

Using the definition of $\psi(\lambda_*)$ this implies $\psi(\lambda_*) \geq f_*$. By the weak duality theorem, it follows that $\psi(\lambda_*) = f_*$. This not only proves that the optimal values are equal, but also that λ_* is an optimal solution of the dual problem.

Remark Unlike in Linear Programming theory, the strong duality theorem cannot always be established for general optimization problems.

2 Topology of Lagrange multipliers

The aim of this Section is to give some original results regarding the topology of Lagrange multipliers set.

Let $f : D \subseteq R^n \to R$ and $g : D \subseteq R^n \to R^p$, $p < n$, of class C^2 with $rank\, J_g = p$ in D. Let $L(x, \lambda) = f(x) + \lambda \cdot g(x)$, with $\lambda \in R^p$. We recall that

$$H(x, \lambda) = \nabla f(x) + \lambda \cdot \nabla g(x) = 0$$

is the equation of critical points with respect to x of the Lagrange function.

Let $A = \{x \mid \exists \lambda \text{ cu } H(x, \lambda) = 0\}$ and $B = \{\lambda \mid \exists x \text{ cu } H(x, \lambda) = 0\}$. Introduce $h : A \to B$ such that $H(x, h(x)) = 0$. The function h is well defined since the equation $H(x, \lambda) = 0$ is linear in λ (system with unique solution). Hence, for any $\lambda \in B$, the set $h^{-1}(\lambda)$ is non-void, and it consists of all critical points corresponding to λ (set in which the nondegenerate critical points are isolated).

Proposition 2.1. *Let $\lambda_0 \in B$ such that there exists $x_0 \in h^{-1}(\lambda_0)$ with the property that x_0 is nondegenerate, i.e., the Hessian $d^2 f(x_0) + \lambda_0 \cdot d^2 g(x_0)$ is nondegenerate. Then h admits a differentiable section $s_{\lambda_0} : I_{\lambda_0} \to A$.*

Proof. Since $\frac{\partial H}{\partial x}(x_0, \lambda_0) = d^2 f(x_0) + \lambda_0 \cdot d^2 g(x_0)$ is non-degenerate, by hypothesis, there exists a neighborhood I_{λ_0} of λ_0 and a differentiable function $s_{\lambda_0} : I_{\lambda_0} \to A$ such that $H(s_{\lambda_0}(\lambda), \lambda) = 0$, $\forall \lambda \in I_{\lambda_0}$ and $s_{\lambda_0}(\lambda_0) = \lambda_0$. Moreover, the function s_{λ_0} is unique, with these properties. □

For any $\lambda \in B$, let S_λ be the set of all sections of h defined in a neighborhood of λ, set which is eventually void.

Remark 2.1. (i) If $h^{-1}(\lambda)$ contains at least one nondgenerate critical point, then S_λ is non-void. If $h^{-1}(\lambda)$ does not contain degenerate critical points, then the sets $h^{-1}(\lambda)$ and S_λ have the same cardinal and are discrete sets.

(ii) The set $C = \left\{ \lambda \in B \mid S_\lambda \neq \varnothing \right\}$ is open.

In the following, we suppose that the set S_λ is finite, for any $\lambda \in B$. We can define $f^* : B \to R$ by $f^*(\lambda) = \max_{s \in S_\lambda} f(s(\lambda))$, if $S_\lambda \neq \varnothing$ and $f^*(\lambda) = -\infty$, if $S_\lambda = \varnothing$.

Proposition 2.2. *(i) For any $\lambda \in B$, we have*

$$f^*(\lambda) \leq \sup_{x \in h^{-1}(\lambda)} f(x).$$

(ii) If $h^{-1}(\lambda)$ does not contain degenerate critical points, then

$$f^*(\lambda) = \sup_{x \in h^{-1}(\lambda)} f(x) = \max_{x \in h^{-1}(\lambda)} f(x).$$

Proof. (i) Let $s_0 \in S_\lambda$ cu $f(s_0(\lambda)) = \max_{s \in S_\lambda} f(s(\lambda)) = f^*(\lambda)$. Since $s_0(\lambda) \in h^{-1}(\lambda)$, it follows that $f^*(\lambda) \leq \sup_{x \in h^{-1}(\lambda)} f(x)$.

(2) By hypothesis, the sets S_λ şi $h^{-1}(\lambda)$ have the same cardinal, hence $h^{-1}(\lambda)$ is finite. Let $y_0 \in h^{-1}(\lambda)$ with $f(y_0) = \max_{x \in h^{-1}(\lambda)} f(x)$. Since (y_0, λ) is a non-degenerate critical point, there exists $s_1 \in S_\lambda$ with $s_1(\lambda) = y_0$. Then, it follows that

$$\max_{s \in S_\lambda} f(s(\lambda)) \geq f(y_0) = \max_{x \in h^{-1}} (\lambda) f(x).$$

□

Proposition 2.3. *Let $\lambda_0 \in B$ such that $h^{-1}(\lambda_0)$ does not contain degenerate critical points. Suppose, also, that $f|_{h^{-1}(\lambda_0)}$ is injective. Then there exists $s_0 \in S_{\lambda_0}$, $s_0 : I_0 \to A$ such that $f^*(\lambda) = f(s_0(\lambda))$, for any $\lambda \in I_0$.*

Proof. Let $s_0 \in S_{\lambda_0}$, $s_0 : I_0 \to A$ such that $f^*(\lambda_0) = f(s_0(\lambda_0))$. Then $f^*(\lambda_0) = f(s_0(\lambda_0)) > f(s(\lambda_0))$, $\forall s \in S_{\lambda_0}$, $s : I_s \to A$. Since f is continuous and the set S_{λ_0} is finite, it follows that we can restrict the neighborhood I_0 such that $f(s_0(\lambda)) > f(s(\lambda))$, $\forall \lambda \in I_0, \forall s \in S_{\lambda_0}$, i.e., $f^*(\lambda) = f(s_0(\lambda)), \forall \lambda \in S_{\lambda_0}$. □

3 The meaning of Lagrange multiplier

In our mostly geometrical discussion, λ is just an artificial variable that lets us compare the directions of the gradients without worrying about their magnitudes. To express mathematically the meaning of the multiplier, we write the constraint in the form $g(x) = c$ for some constant c. This is mathematically equivalent to our usual

$g(x) = 0$, but allows us to easily describe a whole family of constraints. For any given value of c, we can use Lagrange multipliers to find the optimal value of $f(x)$ and the point where it occurs. Call that optimal value f_*, occurring at coordinates x_0 and with Lagrange multiplier λ_0. The answers we get will all depend on what value we used for c in the constraint, so we can think of these as functions of c : $f_*(c), x_0(c), \lambda_0(c)$. Of course, $f(x)$ only depends on c because the optimal coordinates x_0 depend on c: we could write it as $f_*(c)$.

To find how the optimal value changes when we change the constraint, just take the derivative

$$\frac{df_*}{dc} = \frac{\partial f_*}{\partial x_0^i}\frac{dx_0^i}{dc} = \nabla f_* \cdot \frac{dx_0}{dc}.$$

Use the equation of critical points to substitute $\nabla f_* = -\lambda_0 \nabla g_0$ and obtain

$$\frac{df_*}{dc} = -\lambda_0 \nabla g_0 \cdot \frac{dx_0}{dc} = -\lambda_0 \frac{dg_0}{dc}\,.$$

But the constraint function $g_0 = g(x_0(c))$ is *always* equal to c, so $dg_0/dc = 1$. Thus, $df_*/dc = -\lambda_0$. That is, the Lagrange multiplier is the *rate of change of the optimal value with respect to changes in the constraint.*

Of course, f_* depends on c through of λ, and then $\dfrac{df_*}{dc} = \dfrac{df_*}{d\lambda}\dfrac{d\lambda}{dc}$. We can define $c(\lambda)$ by Cauchy problem

$$(EC) \qquad \frac{dc}{d\lambda} = -\frac{1}{\lambda}\frac{df_*}{d\lambda}\,, \quad c(\lambda_0) = 0\,.$$

Then another Lagrange dual function may be

$$(LDF) \qquad \varphi(\lambda) = f_*(x_0(\lambda)) + \lambda\, c(\lambda)\,.$$

Proposition *If optimum points are critical points, both Lagrange dual functions give the same solution. Hence strong duality holds.*

Proof Indeed, using (EC) we have

$$\varphi'(\lambda) = \frac{df_*}{d\lambda} + c(\lambda) + \lambda\frac{dc}{d\lambda} = c(\lambda)$$

and $\varphi'(\lambda) = 0$ implies $c(\lambda) = 0$, that is for λ_0.

Often the Lagrange multiplier have an interpretation as some quantity of interest:

(i) λ is the rate of change of the quantity being optimized as a function of the constraint variable since $\frac{\partial L}{\partial c} = \lambda$;

(ii) by the envelope theorem the optimal value of a Lagrange multiplier has an interpretation as the marginal effect of the corresponding constraint constant upon the optimal attainable value of the original objective function: if we denote values at the optimum with an asterisk, then it can be shown that

$$\frac{d}{dc} f_* = \frac{d}{dc} f(x(c)) = \lambda_*.$$

For details regarding classical theory of programs see [1], [3], [4].

If we have more constraints $g_\alpha(x) = c_\alpha\,, \alpha = 1, ..., m$, then the Lagrange function is $L(x, \lambda) = f(x) + \sum_{\alpha=1}^m \lambda_\alpha (g_\alpha(x) - c_\alpha)$ and the system of critical points is

$$\frac{\partial f}{\partial x^i} + \sum_{\alpha=1}^m \lambda_\alpha \frac{\partial g_\alpha}{\partial x^i} = 0\,.$$

Because the optimal coordinates x_0 and the optimal value f_* depend on vector c, taking the derivatives we have

$$\frac{\partial f_*}{\partial c_\alpha} = \frac{\partial f_*}{\partial x_0^i}\frac{\partial x_0^i}{\partial c_\alpha} = -\sum_{\beta=1}^m \lambda_\beta^0 \frac{\partial g_\beta}{\partial x_0^i}\frac{\partial x_0^i}{\partial c_\alpha} = -\sum_{\beta=1}^m \lambda_\beta^0 \frac{\partial g_\beta}{\partial c_\alpha} = -\sum_{\beta=1}^m \lambda_\beta^0\, \delta_{\alpha\beta} = -\lambda_\alpha^0\,.$$

Then we can define $c(\lambda)$ by the partial differential system, with initial condition, written in matrix language as

$$(EC) \qquad [\lambda_1\ ...\ \lambda_m] \begin{bmatrix} \frac{\partial c_1}{\partial \lambda_1} & ... & \frac{\partial c_1}{\partial \lambda_m} \\ ... & ... & ... \\ \frac{\partial c_m}{\partial \lambda_1} & ... & \frac{\partial c_m}{\partial \lambda_m} \end{bmatrix} = -\left[\frac{\partial f_*}{\partial \lambda_1}\ ...\ \frac{\partial f_*}{\partial \lambda_m}\right], \quad c(\lambda_0) = 0\,.$$

Then another Lagrange dual function may be

$$(LDF) \qquad \varphi(\lambda) = f_*(x_0(\lambda)) + \sum_{\alpha=1}^m \lambda_\alpha\, c_\alpha(\lambda)\,.$$

Deriving the β-th equation with respect to λ_α and the α-th equation with respect to λ_β , in the previous system, we obtain the complete integrability conditions $\dfrac{\partial c_\alpha}{\partial \lambda_\beta} = \dfrac{\partial c_\beta}{\partial \lambda_\alpha}$ (symmetric Jacobian matrix); consequently $c(\lambda)$ is the gradient of a scalar function, namely the Lagrange dual function $\varphi(\lambda)$.

Moreover, the previous square matrix being a symmetrical one, we can write the equation (EC) as

$$[\lambda^1\ ...\ \lambda^m] \begin{bmatrix} \dfrac{\partial c_1}{\partial \lambda^1} & ... & \dfrac{\partial c_m}{\partial \lambda^1} \\ ... & ... & ... \\ \dfrac{\partial c_1}{\partial \lambda^m} & ... & \dfrac{\partial c_m}{\partial \lambda^m} \end{bmatrix} = -\left[\frac{\partial f_*}{\partial \lambda^1}\ ...\ \frac{\partial f_*}{\partial \lambda^m}\right].$$

Consequently, in a regular case, we have the following situation: Solving a constrained optimum problem we obtain the optimal value as $f_* = f(c_1, ..., c_m)$. For the dual problem we use a $f_* = f(\lambda^1, ..., \lambda^m)$. If the correspondence between $(c_1, ..., c_m)$ and $(\lambda^1, ..., \lambda^m)$ is like a change of variables there hold the relations:

$$grad_c\, f_* = -\lambda\,; \quad c'(\lambda)\,\lambda = -grad_\lambda\, f_*\,,\ c'(\lambda) \in L(R^m,\, R^m)\,.$$

4 Examples and counter-examples

We solve pertinent examples of constrained optimization problems.

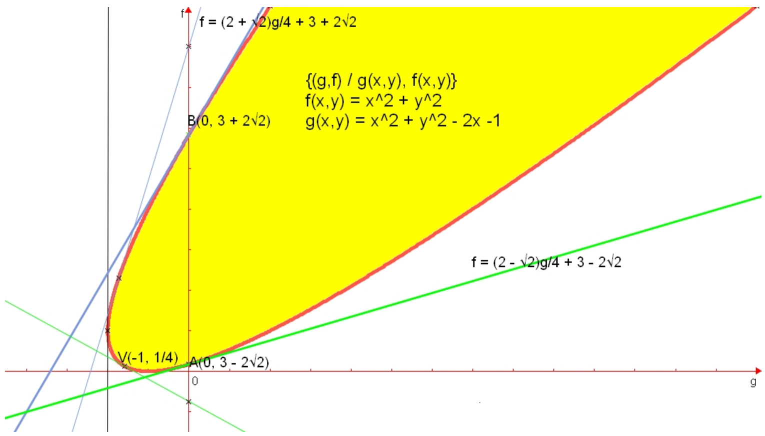

Figure 1: Geometry of Lagrange duality

(1) Let us consider the functions $f(x, y) = x^2 + y^2$ and $g(x, y) = x^2 + y^2 - 2x$ and the problem

$$\min f(x, y) \text{ constrained by } g(x, y) = c,\ c \geq -1.$$

The Lagrange function of this problem is

$$L(x, y, \lambda) = x^2 + y^2 + \lambda(x^2 + y^2 - 2x - c).$$

The critical points of the partial function $(x, y) \to L(x, y, \lambda)$ are the solutions of the system

$$\frac{1}{2}\frac{\partial L}{\partial x} = x + \lambda x - \lambda = 0,\ \frac{1}{2}\frac{\partial L}{\partial y} = y + \lambda y = 0.$$

It follows $x = \frac{\lambda}{\lambda+1}$, $y = 0$ and hence $f_* = \left(\frac{\lambda}{\lambda+1}\right)^2$. On the other hand, by restriction, in critical points, we have the relation

$$\left(\frac{\lambda}{\lambda+1}\right)^2 - 2\frac{\lambda}{\lambda+1} = c.$$

It follows

$$\frac{df_*}{dc} = \frac{df_*}{d\lambda}\ \frac{1}{\frac{dc}{d\lambda}}$$

and finally, we obtain the geometrical interpretation $\frac{df_*}{dc} = -\lambda$.

The dual function is

$$\psi(\lambda) = L(x(\lambda), y(\lambda), \lambda) = -\frac{\lambda^2}{\lambda+1} - \lambda c.$$

The value $\psi(\lambda)$ is a minimum for $\lambda > -1$ and a maximum for $\lambda < -1$, in the initial problem. The condition of extremum (critical point), $\psi'(\lambda) = 0$, is equivalent to $(c+1)(\lambda+1)^2 = 1$ and the dual problem has the same solution as the primal one.

On the other hand the equation (EC) for this problem is

$$\frac{d\mathbf{c}}{d\lambda} = -\frac{1}{\lambda}\frac{d}{d\lambda}\left(\frac{\lambda}{\lambda+1}\right)^2, \quad \mathbf{c}(\lambda_0) = 0\,, \text{ where } \lambda_0 + 1 = \pm\frac{1}{\sqrt{c+1}}\,.$$

We find $\mathbf{c}(\lambda) = \frac{1}{(\lambda+1)^2} - 1 - c$ and the dual Lagrange function

$$\varphi(\lambda) = \left(\frac{\lambda}{\lambda+1}\right)^2 + \lambda\left(\frac{1}{(\lambda+1)^2} - 1 - c\right) = \psi(\lambda)$$

as the above one.

The Geometry of Lagrange duality is suggested in Fig. 1.

(2) **A problem with two constraints** Solve the following constrained optimum problem:

$$f(x,y,z) = xyz = extremum$$

constrained by

$$g_1(x,y,z) = x + y - a = 0,\ g_2(x,y,z) = xz + yz - b = 0\,.$$

The Lagrange function is

$$L(x,y,z,\lambda,\mu) = xyz + \lambda(x+y-a) + \mu(xz+yz-b)$$

and the feasible solution of the problem is, only,

$$x = y = \frac{a}{2} = -2\mu\,,\ z = \frac{b}{a} = \frac{\lambda}{\mu}\,,\ \lambda = -\frac{b}{4}\,,\ \mu = -\frac{a}{4}\,,$$

$$f_* = \frac{ab}{4} = 4\lambda\mu\,.$$

The Lagrange dual function is $\psi(\lambda,\mu) = -4\lambda\mu - a\lambda - b\mu\,.$

The partial differential system which defines $c_1(\lambda,\mu)$ and $c_2(\lambda,\mu)$ becomes, in this case,

$$(EC) \qquad [\lambda \ \ \mu]\begin{bmatrix} \frac{\partial c_1}{\partial\lambda} & \frac{\partial c_1}{\partial\mu} \\ \frac{\partial c_2}{\partial\lambda} & \frac{\partial c_2}{\partial\mu} \end{bmatrix} = -[4\mu \ \ 4\lambda]\,,\ c_1\left(-\frac{b}{4},-\frac{a}{4}\right) = c_2\left(-\frac{b}{4},-\frac{a}{4}\right) = 0\,.$$

Taking into account that $\frac{\partial c_1}{\partial\mu} = \frac{\partial c_2}{\partial\lambda}$, we obtain two quasilinear PDEs

$$\lambda\frac{\partial c_1}{\partial\lambda} + \mu\frac{\partial c_1}{\partial\mu} = -4\mu,\ \lambda\frac{\partial c_2}{\partial\lambda} + \mu\frac{\partial c_2}{\partial\mu} = -4\lambda\,,$$

with solutions, respectively

$$c_1(\lambda,\mu) = -4\mu + \alpha\left(\frac{\lambda}{\mu}\right),\ c_2(\lambda,\mu) = -4\lambda + \beta\left(\frac{\lambda}{\mu}\right),$$

$\alpha\,,\beta$ arbitrary functions. The condition $\dfrac{\partial c_1}{\partial \mu} = \dfrac{\partial c_2}{\partial \lambda}$ is verified, for instance, if α and β are constant functions. Using the initial conditions, we find finally

$$c_1(\lambda,\mu) = -4\mu - a\,,\ c_2(\lambda,\mu) = -4\lambda - b\,,$$

$$\varphi(\lambda,\mu) = 4\lambda\mu + \lambda(-4\mu - a) + \mu(-4\lambda - b) = \psi(\lambda,\mu)\,.$$

(3) **A strange problem** Solve the following constrained optimum problem:

$$f(x,y,z) = xyz = extremum$$

constrained by

$$g_1(x,y,z) = x + y + z - a = 0\,,$$

$$g_2(x,y,z) = xy + xz + yz - b = 0\,.$$

So the Lagrange function is

$$L(x,y,z,\lambda,\mu) = xyz + \lambda(x + y + z - a) + \mu(xy + xz + yz - b)$$

and one from the solutions of the problem is, for instance,

$$\mu = \frac{-a - \sqrt{a^2 - 3b}}{3}\,,\quad \lambda = \frac{2a^2 - 3b + 2a\sqrt{a^2 - 3b}}{9} = \mu^2\,,$$

$$x = y = \frac{a + \sqrt{a^2 - 3b}}{3} = -\mu\,,\quad z = \frac{a - 2\sqrt{a^2 - 3b}}{3} = a + 2\mu\,,$$

with the extremum value

$$f_* = \frac{1}{27}\left(-2a^3 + 9ab - 2(a^2 - 3b)^{3/2}\right) = \lambda(a + 2\mu)\,,$$

only if $a^2 - 3b \geq 0$.

Remark Another solution of the problem is

$$\mu = \frac{-a + \sqrt{a^2 - 3b}}{3}\,, \ldots \text{ and so on}$$

with the extremum value

$$f^* = \frac{1}{27}\left(-2a^3 + 9ab + 2(a^2 - 3b)^{3/2}\right) = \lambda(a + 2\mu)\,.$$

The interval $[f_*\,, f^*]$ solves the following algebraic problem: *Find the real numbers m such that the equation $t^3 - at^2 + bt - m = 0,\ a, b \in R$, has tree real roots.*

It is easily to verify that

$$\frac{\partial}{\partial a} f_*(a,b) = -\lambda \text{ and } \frac{\partial}{\partial b} f_*(a,b) = -\mu\,.$$

On the other hand, $\dfrac{\partial(\lambda,\mu)}{\partial(a,b)} = 0$ and f_* cannot be expressed as function of λ and μ only. Then we have to consider $f_* = f_*(a, b, \lambda(a,b), \mu(a,b))$ and the following relations

$$-\lambda = D_a f_* = \frac{\partial f_*}{\partial a} + \frac{\partial f_*}{\partial \lambda}\frac{\partial \lambda}{\partial a} + \frac{\partial f_*}{\partial \mu}\frac{\partial \mu}{\partial a}$$

$$-\mu = D_b f_* = \frac{\partial f_*}{\partial b} + \frac{\partial f_*}{\partial \lambda}\frac{\partial \lambda}{\partial b} + \frac{\partial f_*}{\partial \mu}\frac{\partial \mu}{\partial b}$$

which is easily to verify also (here $D_.$ is an operator of total derivative.)

Question Which is the dual Lagrange function $\psi(\lambda, \mu)$ in this case?

Solving the system of the critical points with respect to x, y and z we find, for instance, $x = y = -\mu$, z undeterminate and $\lambda = \mu^2$. With these, one obtains the dual Lagrange function

$$\psi(\lambda, \mu) = \chi(\mu) = -\mu^3 - a\mu^2 - b\mu\,.$$

Remark Although z is undeterminate, the dual Lagrange function does not depend upon z, because with the above solutions the derivative $\dfrac{\partial L}{\partial z}$ vanishes identically. The critical points condition for the dual Lagrange function

$$\frac{d\chi}{d\mu} = -(3\mu^2 + 2a\mu + b) = 0$$

gives us the same solutions as in primal problem.

Open problem How it means and how we find the functions c_1 and c_2 in the situation, like this, when $\dfrac{\partial(\lambda, \mu)}{\partial(a, b)} = 0$?

(4) Let us consider the functions $f(x, y) = x^2 + y^2$ and $g(x, y) = x + y$, with $(x, y) \in R^2$, and the problem

$$\min f(x, y) \text{ constrained by } g(x, y) \geq 1.$$

The Lagrange function is

$$L(x, y, \lambda) = x^2 + y^2 + \lambda(1 - x - y),\ (x, y) \in R^2,\ \lambda \geq 0.$$

The function $(x, y) \to L(x, y, \lambda)$ is convex. Consequently, it is minimal iff

$$\frac{\partial L}{\partial x} = 0, \frac{\partial L}{\partial y} = 0.$$

This holds if $x = \frac{\lambda}{2}$, $y = \frac{\lambda}{2}$. Substitution gives

$$\psi(\lambda) = \lambda - \frac{\lambda^2}{2}, \lambda \geq 0.$$

The dual problem $\max \psi(\lambda)$ has the optimal point $\lambda = 1$. Consequently, $x = y = \frac{1}{2}$ is the optimal solution of the original (primal) problem. In both cases the optimal value equals $\frac{1}{2}$, i.e., at optimality the duality gap is zero!

(5) Let us solve the program

$$\min x \text{ subject to } x^2 \leq 0,\ x \in R.$$

This program is not Slater regular. On the other hand, we have

$$\psi(\lambda) = \inf_{x \in R}(x + \lambda x^2) = \begin{cases} -\frac{1}{2\lambda} & for \quad \lambda > 0 \\ -\infty & for \quad \lambda = 0. \end{cases}$$

Obviously, $\psi(\lambda) < 0$ for all $\lambda \geq 0$. Consequently, $\sup\{\psi(\lambda) \,|\, \lambda \geq 0\} = 0$. So the Lagrange-dual has the same optimal value as the primal problem. In spite of the lack of Slater regularity there is no duality gap.

(6) **(Example with positive duality gap)** We consider the program

$$\min e^{-y} \text{ subject to } \sqrt{x^2+y^2} - x \leq 0,\, (x,y) \in R^2.$$

Here the feasible region is $\Omega = \{(x,y) \in R^2 \,|\, x \geq 0, y = 0\}$. Consequently this program is not Slater regular. The optimal value of the objective function is 1. The Lagrange function is

$$L(x,y,\lambda) = e^{-y} + \lambda(\sqrt{x^2+y^2} - x).$$

The Lagrange dual program can be written in the form

$$\sup \psi(\lambda) \text{ subject to } \lambda \geq 0.$$

Note that $L(x,y,\lambda) > 0$ implies $\psi(\lambda) \geq 0$. Now let $\epsilon > 0$. Fixing $y = -\ln \epsilon$ and $x = \frac{y^2-\epsilon^2}{2\epsilon}$, we find $\sqrt{x^2+y^2} - x = \epsilon$ and

$$L(x,y,\lambda) = (1+\lambda)\epsilon.$$

In this way,

$$\psi(\lambda) = \inf_{(x,y)\in R^2} L(x,y,\lambda) \leq \inf_{\epsilon>0} (1+\lambda)\epsilon = 0.$$

On the other hand, we also have $\psi(\lambda) \geq 0$, and consequently the optimal value of the Lagrange dual program is 0, and hence the minimal duality gap equals 1. Of course, here we have no strong duality here.

5 The Wolfe-dual problem

The *Lagrange dual program* can be written in the form

$$\sup_{\lambda\geq 0} \{ \inf_{x\in D} \{f(x) + \sum_{\alpha=1}^{m} \lambda_\alpha g_\alpha(x)\}\}.$$

Assume that $D = R^n$ and the functions $f, g_1, ..., g_m$ are continuously differentiable and convex. For a given $\lambda \geq 0$ the inner minimization problem is convex, and we can use the fact that the infimum is attained if and only if the gradient with respect to x is zero.

Definition 5.1. The problem

$$(WP) \qquad \sup_{x,\lambda} \{f(x) + \sum_{\alpha=1}^{m} \lambda_\alpha g_\alpha(x)\}$$

subject to

$$\frac{\partial f}{\partial x^i}(x) + \sum_{\alpha=1}^{m} \lambda_\alpha \frac{\partial g_\alpha}{\partial x^i}(x) = 0,\ \lambda \geq 0$$

is called the *Wolfe dual* of the program (P).

Obviously, the constraints in Wolfe dual are usually nonlinear. In such cases the Wolfe-dual is not a convex program.

The Wolfe dual has the weak duality property.

Theorem 5.1. (weak duality property) *Suppose that $D = R^n$ and the functions $f, g_1, ..., g_m$ are continuously differentiable and convex. If $\hat{x}$ is a feasible solution of (P) and $(\bar{x}, \bar{\lambda})$ is a feasible solution for (WP), then*

$$L(\bar{x}, \bar{\lambda}) \leq f(\hat{x}).$$

In other words, weak duality holds for (P) and (WP).

5.1 Example

(1) Let us consider the convex program

$$\min_{x,y} x + e^y \text{ subject to } 3x - 2e^y \geq 10,\ y \geq 0,\ (x, y) \in R^2.$$

Then the optimal value is 5 with $x = 4, y = 0$. The Wolfe dual of this program is

$$\sup_{x,y,\lambda} \{x + e^y + \lambda_1(10 - 3x + 2e^y) - \lambda_2 y\}$$

subject to

$$1 - 3\lambda_1 = 0,\ e^y + 2e^y\lambda_1 - \lambda_2 = 0,\ (x, y) \in R^2,\ \lambda \geq 0.$$

Obviously, the Wolfe dual program is not convex. It follows $\lambda_1 = \frac{1}{3}$ and the second constraint becomes $\frac{5}{3}e^y - \lambda_2 = 0$. Eliminating λ_1, λ_2 from the objective function, we find

$$g(y) = \frac{5}{3}\, e^y - \frac{5}{3}\, y\, e^y + \frac{10}{3}.$$

This function has a maximum when $g'(y) = 0$, i.e., $y = 0$ and $f(0) = 5$. Hence the optimal value of (WP) is 5 and then $(x, y, \lambda_1, \lambda_2) = (4, 0, \frac{1}{3}, \frac{5}{3})$.

Remark The substitution $z = e^y \geq 1$ makes the problem linear.

6 Minimax inequality

For any function ϕ of two vector variables $x \in X$, $y \in Y$, the *minimax inequality*

$$\max_{y \in Y} \min_{x \in X} \phi(x, y) \leq \min_{x \in X} \max_{y \in Y} \phi(x, y)$$

is true. Indeed, start from

$$\forall x, y : \min_{x' \in X} \phi(x', y) \leq \max_{y' \in Y} \phi(x, y')$$

and take the minimum over $x \in X$ on the right-hand side, then the maximum over $y \in Y$ on the left-hand side.

Weak duality is a direct consequence of the minimax inequality. To see this, start from the unconstrained formulation of Lagrange, and apply the above inequality, with $\phi = L$ the Lagrangian of the original problem, and $y = \lambda$ the Lagrange vector multiplier.

Acknowledgements

Partially supported by University Politehnica of Bucharest and by Academy of Romanian Scientists.

References

[1] H. Bonnel, *Analyse Fonctionnelle*, Maitrise de Mathématiques, Ingénerie Mathéematiques, Université de La Reunion, Faculté de Sciences et Technologies, 2015.

[2] G. Darboux, *Sur le problème de Pfaff*, Bull. Sci. Math. 6 (1882), 14-36, 49-68.

[3] R. B. Holmes, *Geometric Functional Analysis and Its Applications*, Springer-Verlag, New York, 1975.

[4] K. Roos, *Nonlinear Programming*, LNMB Course, De Uithof, Utrecht, TUDelft, February 6 - May 8, A.D. 2006.

[5] S. Sternberg, *Lectures on Differential Geometry*, Prentice Hall, 1964.

[6] O. Dogaru, V. Dogaru, *Extrema Constrained by C^k Curves*, Balkan Journal of Geometry and Its Applications, 4, 1 (1999), 45-42.

[7] O. Dogaru, I. Ţevy, *Extrema Constrained by a Family of Curves*, Proc. Workshop ob Global Analysis, Diff. Geom. and Lie Algebras, 1996, Ed. Gr. Tsagas, Geometry Balkan Press, 1999, 185-195.

[8] I. Ekeland, *Exterior Differential Calculus and Applications to Economic Theory*, Quaderni Scuola Normale Superiore di Pisa, 1998, Italy.

[9] R. Montgomery, *A Tour of Subriemanniene Geometries, Their Geodesics and Applications*, Mathematical Surveys and Monographs, 91, American mathematical Society, 2002.

[10] V. Radcenco, C. Udrişte, D. Udrişte, *Thermodynamic Systems and Their Interaction*, Sci. Bull. P.I.B., Electrical Engineering, vol. 53, no. 3-4 (1991), 285-294.

[11] Gr. Tsagas, C. Udrişte, *Vector Fields and Their Applications*, Geometry Balkan Press, Bucharest, 2002.

[12] C. Udrişte, O. Dogaru, *Mathematical Programming Problems with Nonholonomic Constraints*, Seminarul de Mecanică, Univ. of Timişoara, Facultatea de Ştiinţe ale Naturii, vol. 14, 1988.

[13] C. Udrişte, O. Dogaru, *Extrema with Nonholonomic Constraints*, Sci. Bull., Polytechnic Institute of Bucharest, Seria Energetică, Tomul L, 1988, 3-8.

[14] C. Udrişte, O. Dogaru, *Extreme condiţionate pe orbite*, Sci. Bull., 51 (1991), 3-9.

[15] C. Udrişte, O. Dogaru, *Convex Nonholonomic Hypersurfaces*, Math. Heritage of C.F. Gauss, 769-784, Ed. G. Rassias, World Scientific, 1991.

[16] C. Udrişte, O. Dogaru, I. Ţevy, *Sufficient Conditions for Extremum on Differentiable Manifolds*, Sci. Bull., Polytechnic Institute of Bucharest, Electrical Engineering, vol. 53, no. 3-4 (1991), 341-344.

[17] C. Udrişte, O. Dogaru, I. Ţevy, *Extremum Points Associated with Pfaff Forms*, Presented at the 90th Anniversary Conference of Akitsugu Kawaguchi's Birth, Bucharest, Aug. 24-29, 1992; Tensor, N.S., Vol. 54 (1993), 115-121.

[18] C. Udrişte, O. Dogaru, I. Ţevy, *Open Problem in Extrema Theory*, Sci. Bull. P.U.B., Series A, Vol. 55, no.3-4 (1993), 273-277.

[19] O. Dogaru, I. Ţevy, C. Udrişte, *Extrema Constrained by a Family of Curves and Local Extrema*, JOTA, vol. 97, no.3, June 1998, 605-621.

[20] C. Udrişte, O. Dogaru, I. Ţevy, *Extrema Constrained by a Pfaff System*, Hadronic J. Supplement, USA, 1991-Proc. Int. Workshop on Fundam. Open Problems in Math., Phys. and Other Sciences, Beijing, August 28, 1997.

[21] C. Udrişte, I. Ţevy, M. Ferrara, *Nonholonomic Economic Systems*, see [28], 139-150.

[22] C. Udrişte, I. Ţevy, *Geometry of test Functions and Pfaff Equations*, see [28], 151-165.

[23] C. Udrişte, O. Dogaru, I. Ţevy, *Extrema with Nonholonomic Constraints*, Geometry Balkan Press, Bucharest, 2002.

[24] C. Udrişte, O. Dogaru, M. Ferrara, I. Ţevy, *Pfaff Inequalities and Semi-curves in Optimum Problems*, Recent Advances in Optimization, pp. 191-202, Proceedings of the Workshop held in Varese, Italy, June 13/14th 2002, Ed. G.P. Crespi, A. Guerraggio, E. Miglierina, M. Rocca, DATANOVA, 2003.

[25] C. Udrişte, O. Dogaru, M. Ferrara, I. Ţevy, *Pfaff inequalities and semi-curves in optimum problems*, in Edt. G. P. Crespi, A. Guerraggio, E. Miglierina, M. Rocca, Recent Advances in Optimization, Proceedings of the Workshop held in Varese, Italy, June 13-14, 2002, pp. 191-202.

[26] C. Udrişte, O. Dogaru, M. Ferrara, I. Ţevy, *Extrema with constraints on points and/or velocities*, Balkan Journal of Geometry and Its Applications, 8, 1(2003), 115-123.

[27] C. Udrişte, O. Dogaru, M. Ferrara, I. Ţevy, *Nonholonomic optimization theory*, see [28], 177-192, Geometry Balkan Press, Bucharest, 2004.

[28] C. Udriste, M. Ferrara, D. Opris, *Economic Geometric Dynamics*, Geometry Balkan Press, Bucharest, 2004.

[29] Gh. Vrănceanu, *Leçons de Geometry Differentielle*, Editions de l'Academie Roumaine, Bucarest, 1957-1975.

Authors' address:

Constantin Udrişte, Mădălina Constantinescu, Ionel Ţevy *and* Oltin Dogaru
University Politehnica of Bucharest, Faculty of Applied Sciences
Department of Mathematics and Informatics, Splaiul Independentei 313
Bucharest 060042, Romania.

Email: udriste@mathem.pub.ro , madalina.const@yahoo.com
vascatevy@yahoo.fr , oltin.horia@yahoo.com