Novel insights into the multiplier rule

Jan Brinkhuis
Erasmus University Rotterdam

Vladimir Protasov
Moscow State University

29 September 2005

Econometric Institute Report EI-2005-39

We present the Lagrange multiplier rule, one of the basic optimization methods, in a new way. Novel features include:

- Explanation of the true source of the power of the rule: reversal of tasks, but not the use of multipliers.
- A natural proof based on a simple picture, but not the usual technical derivation from the implicit function theorem.
- A practical method to avoid the cumbersome second order conditions.
- Applications from various areas of mathematics, physics, economics.
- Some hints on the use of the rule.

I. Lagrange multiplier rule.

Suppose we need to find the maximal or minimal value of a function of n variables $f_0(x^1, \ldots, x^n)$ (objective function), and we know that the variables are linked by several constraints $f_j(x^1, \ldots, x^n) = 0,\ 1 \le j \le m$.

There are two principal approaches to solve this problem. The first one is to try to eliminate m variables by means of the given constraints, and then to find local maxima (minima) in the usual way, by putting partial derivatives equal to zero. For many problems this straightforward method is very inconvenient. First, the equations $f_j(x^1, \ldots, x^n) = 0$ may be too difficult to eliminate variables. For example, if f_j is a polynomial of the third power or higher in some variable. Even if we succeed in expressing some variable in others, the derivative of this expression may be complicated. Another difficulty is that if a problem possesses some symmetry in the variables (in practice this happens very often), then this symmetry will be completely lost after elimination of variables.

The second approach, invented in 1797 by Joseph Louis Lagrange, suggests the following rule:

Each local minimum/maximum of a function f_0 under the equality constraints $f_j(x^1, \ldots, x^n) = 0,\ 1 \le j \le m$, is a solution of the system of equations

$$\mathcal{L}_{x^i}(x) \;=\; 0\,, \quad 1 \le i \le m, \tag{1}$$

where $\mathcal{L} = \mathcal{L}(x, \lambda)$ denotes the Lagrange function $\sum_{j=0}^{m} \lambda_j f_j(x)$ for a suitable nonzero choice of $\lambda = (\lambda_0, \ldots, \lambda_m)$, and $\mathcal{L}_{x^i}$ is the derivative with respect to the variable x^i.

This is the *Lagrange multiplier rule.* Of course, all partial derivatives of all functions should exist and be continuous. Note that we get a system of equations where the number of unknowns is one more than the number of equations. However, we can get rid of one unknown: we can always assume without loss generality that λ_0 is either 0 or 1, as the stationarity equations (1) are homogeneous in the λ's. We will always put $\lambda_0 = 1$ and not display the routine verification that $\lambda_0 = 0$ is impossible. The bad case $\lambda_0 = 0$ may occur, but mostly in artificially concocted examples.

II. The secret of the power of the multiplier rule.

What is the source of the power of the multiplier rule? It is the simple but clever idea of reversing the natural order of the main tasks, elimination and differentiation. This turns the hardest task, elimination, from a nonlinear problem into a linear one. It is *not* the 'miraculous' use of multipliers, which is usually given. These serve only to put the rule in a more user-friendly form. Let us illustrate this with the following numerical example:

Example 1 *What is the smallest possible value of the function* $f(x,y) = 5x^2 + 2xy + 3y^2$ *if* $g(x,y) = 7x^2 + 2xy + 4y^2 - 1 = 0$*?*

Solution. We form the Lagrangian $\mathcal{L}(x,y,\lambda) = f(x,y) + \lambda g(x,y)$. Differentiating with respect to x and y we get the following system:

$$\begin{cases} 10x + 2y + \lambda(14x + 2y) = 0 \\ 6y + 2x + \lambda(8y + 2x) = 0, \end{cases}$$

which implies $\frac{10x+2y}{14x+2y} = \frac{6y+2x}{8y+2x}$. From this we immediately get $\frac{y}{x} = -1$ or 2. Substituting into $g(x,y) = 0$ we obtain several suspicious points (x,y), from which we get the points of minimum $(x,y) = (-\frac{1}{3}, \frac{1}{3})$ and $(x,y) = (\frac{1}{3}, -\frac{1}{3})$ by a comparison of f-values.

In this example we could have eliminated first, expressing y in x (or x in y) by means of the constraint $g(x,y) = 0$. However, substituting that expression into the equation $f'(x, y(x)) = 0$ we get an equation of sixth degree, and then it is not clear what to do next. Moreover, there will actually be two equations, because of the sign $\pm$ before the discriminant in the expression for y. Therefore, it is better to keep both variables and to use Lagrange multiplier rule. This example illustrate the following principle:

In many extremal problems it is better to keep all variables and to write the Lagrangian than to express some variables in the others.

Example 2 *[The Steiner problem]. The lengths of all sides of a quadrangle are fixed, but the sides are linked freely at the vertices, so the angles between them can vary. Which position of the sides corresponds to the largest area of the quadrangle?*

Solution. We have

$$\begin{cases} S(\alpha,\beta) \;=\; \frac{1}{2}\Big(ab\sin\alpha + cd\sin\beta\Big) \quad \rightarrow \quad \max \\ a^2 + b^2 - 2ab\cos\alpha \;=\; c^2 + d^2 - 2cd\cos\beta \end{cases} \tag{2}$$

where a, b, c, d are the sides of the quadrangle, α is the angle between a and b, β is the angle between c and d. The equality constraint comes from applying the cosine rule to both triangles with side a diagonal. Differentiating the Lagrangian $\mathcal{L}(\alpha,\beta,\lambda) = S(\alpha,\beta) + \lambda\Big(a^2 + b^2 - 2ab\cos\alpha - c^2 - d^2 + 2cd\cos\beta\Big)$ in α and β:

$$\begin{aligned} \mathcal{L}_\alpha = 0 \quad &\Leftrightarrow \quad \tfrac{1}{2}ab\cos\alpha + 2\lambda ab\sin\alpha = 0 \\ \mathcal{L}_\beta = 0 \quad &\Leftrightarrow \quad \tfrac{1}{2}cd\cos\beta - 2\lambda cd\sin\beta = 0 \end{aligned} \tag{3}$$

From the first equation we obtain $\tan\alpha = -\frac{1}{4\lambda}$ and from the second one $\tan\beta = \frac{1}{4\lambda}$. Thus, $\tan\alpha = -\tan\beta$ and hence $\alpha = \pi - \beta$. This means that the quadrangle *abcd* is inscribed in a circle. Thus,

Among all the quadrangles with given sides the inscribed quadrangle possesses the largest area.

We have not solved system (3) completely and have not found the multiplier λ. We have used λ only to draw the conclusion that $\tan\alpha = -\tan\beta$. We see that

In most problems there is no need to find the multipliers and all variables.

III. Natural proof of the multiplier rule.

The usual proof of the Lagrange multiplier rule proceeds by means of some formula manipulations using a technical result, the implicit function theorem. Let us give a novel proof based on a simple figure. A picture for the case of two variables and one constraint gives a full insight (cf. Fig 1).

We will use the following property of continuously differentiable function f at a point x:

$$\frac{|f(x_1) - f(x_2) - f'(x)(x_1 - x_2)|}{|x_1 - x_2|} \to 0 \quad \text{as } r \to 0,$$

provided $|x_i - x| < r,\ i = 0, 1$.

Proof of Lagrange multiplier rule. We argue by contradiction. Let $\widehat{x}$ be a point of local minimum of the function $f_0(x)$ under the conditions $f_j(x) = 0,\ j = 1, \ldots, m$, but the vectors $f_j'(\widehat{x}),\ 0 \le j \le m$ are linearly independent. Write $F = (f_1, \ldots, f_m)^T$ and let v be the orthogonal projection of $f_0'(\widehat{x})^T$ on $(\mathrm{im} F'(\widehat{x})^T)^\perp = \ker F'(\widehat{x})$. Then $v \ne 0$. By continuity, the matrix $F'(x)$ has rank m for all $x \in U$, where U is some ball with the center at $\widehat{x}$. The main task is proving that there exists $h_\alpha = o(\alpha),\ \alpha \to 0$ such that the point $x_\alpha = \widehat{x} - \alpha v + h_\alpha$ is admissible (satisfy the conditions $f_j(x) = 0,\ 1 \le j \le m$) for each sufficiently small $\alpha > 0$ (as usual, $o(\alpha)$ denotes a value, for which $o(\alpha)/\alpha \to 0$). We produce this element as a solution of the auxiliary problem

$$g_\alpha(h) = |F(\widehat{x} - \alpha v + h)| \to \min, \quad h \in (\ker F'(\widehat{x}))^\perp,\ |h| \le |\alpha v|.$$

Existence of a global solution h_α follows from the Weierstrass theorem. We have to check that $F(x_\alpha) = 0$. Well, otherwise h_α would be a point of

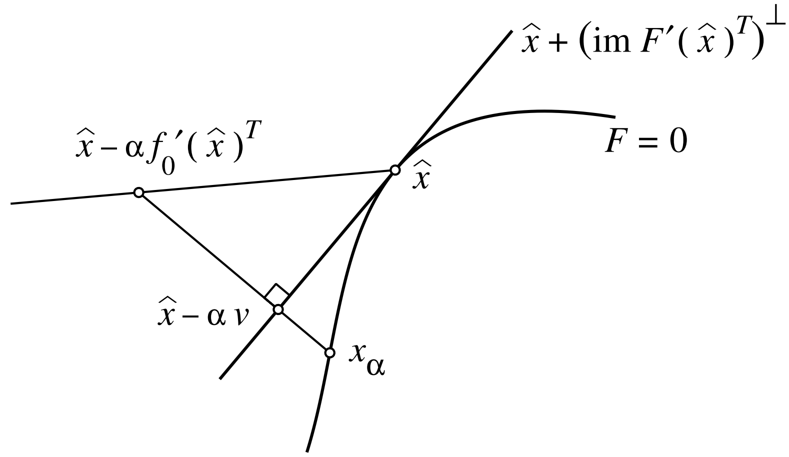

Figure 1: Idea proof multiplier rule.

differentiability of g_α. We are going to exclude this. To begin with, consider the case that h_α is an interior point, then

$$g'_\alpha(h) \;=\; 0 \;\Rightarrow\; |F(\widehat{x} - \alpha v + h)|^{-1} F(\widehat{x} - \alpha v + h)^T F'(\widehat{x} - \alpha v + h) \;=\; 0.$$

As $F'(\widehat{x} - \alpha v + h)$ has rank m, this leads to contradiction. Now consider the case that h_α is a boundary point, then

$$\frac{|F(x_\alpha) - F'(\widehat{x})h_\alpha|}{|-\alpha v + h_\alpha|} \;\to\; 0 \quad \text{and} \quad \frac{|F(\widehat{x} - \alpha v)|}{|\alpha v|} \;\to\; 0 \quad \text{as} \quad \alpha \to 0$$

and therefore $g_\alpha(h_\alpha) - g_\alpha(0) \;\geq\; C|\alpha||v| + o\,(\alpha)$, where $C > 0$ is a constant, for which $|F'(\widehat{x})x| \geq C|x|$ for all $x \in \ker F'(\widehat{x})^\perp$ and so we are again led to a contradiction: $g_\alpha(h_\alpha) \;>\; g_\alpha(0)$. This finishes the verification that $F(x_\alpha) = 0$. Moreover, we have $0 \;=\; F(x_\alpha) = F'(\widehat{x})h_\alpha + o(\alpha)$, and so $h_\alpha \;=\; o(\alpha)$. Now we compute

$$f_0(x_\alpha) \;=\; f_0(\widehat{x}) \;-\; \alpha f_0'(\widehat{x})v \;+\; o(\alpha) \;=\; f_0(\widehat{x}) \;-\; \alpha|v|^2 \;+\; o(\alpha).$$

It follows that $f_0(\widehat{x} - \alpha v + h_\alpha) \;<\; f_0(\widehat{x})$ for sufficiently small $\alpha > 0$. Therefore, $\widehat{x}$ is not a local solution of the given problem.

$\diamond$

IV. No need for second order conditions.

We recommend to complete the analysis of optimization problems instead by using that a continuous function on a nonempty compact (that is, closed and bounded) set in $\mathbb{R}^n$ attains its maximum and minimum (theorem of Weierstrass). In the case of a non-compact domain we have to do some preparations before we can use the Weierstrass theorem: we can usually show that there exists a compact subset C and a point p inside it such that all values taken by the objective function f outside C are larger than at p. For example this is the case if f is a *coercive* function. This means that $f(x)$ tends to $+\infty$ if x tends to the boundary of the domain of f or if $|x| \to \infty$.

Usually great emphasis is given to second order conditions, to be used instead of the Weierstrass theorem. These conditions are a considerable obstacle for everyone who wants to come to grips with optimization methods. Even their formulation, in terms of minors of bordered hessians, is fearsome, and their use leads to longwinded computations. Their achievement is not impressive: these conditions allow us to distinguish between *local* minima and maxima; they give no global information.

V. Convincing applications of the multiplier rule.

Now turn to applications of the Lagrange multiplier rule that are more difficult to derive by other methods.

Geometry

Example 3 *Let a straight line and three points be given on the plane. Find (or characterize) the point on the line for which the sum of distances from this point to the three given points is minimal.*

Solution. Denote the three given points by x_1, x_2 and x_3 and the straight line by l. We obtain the following minimization problem:

$$(4) \qquad \begin{cases} f(x) \ = \ |x-x_1| + |x-x_2| + |x-x_3| \ \to \ \min \\ x \ \in \ l. \end{cases}$$

By coercivity of the function f the point of minimum x does exist. The derivative f' (gradient) of the length $|x-x_1|$ at x is a unit vector u_1 with the same direction as the vector $x - x_1$. Similarly we define u_2 and u_3. We write the condition $x \in l$ as a constraint $\langle x - x_0, n\rangle = c$, where x_0 vis a point on l, n

is a vector orthogonal to l, c is some constant, and $\langle \cdot, \cdot \rangle$ is the standard inner product. Differentiating the Lagrangian $\mathcal{L}(M, \lambda) = f(x) + \lambda(\langle x - x_0, n\rangle - c)$ we get $u_1 + u_2 + u_3 = -\lambda n$, which means that the sum $u_1 + u_2 + u_3$ is orthogonal to l. This is the same as saying that the sum of the projections of the vectors u_1, u_2, u_3 onto l (or the sum of cosines of angles formed by these vectors with the line l) is zero. This property characterizes the desirable point x.

The solution remains the same for an arbitrary number of points $x_1, \ldots, x_k$. In particular, for $k = 2$ we obtain a well-known elementary high-school problem. For $k \geq 3$ the solution, in general, cannot be constructed by compasses and ruler, and can only be characterized as we did above. The same principle of solution is illustrated by the following problems:

Example 4 Find (or characterize) the point on a plane, for which the sum of distances from this point to k given points in this plane is minimal.

Example 5 Find (or characterize) the point on a plane, for which the sum of distances from this point to three given points in three-dimensional space is minimal.

Physics

Example 6 *[Snellius' law]. A ray of light passes through the flat boundary line between two media. It comes to the boundary at an angle of incidence α_1 and leaves at an angle α_2 (both angles are taken with the normal to the boundary). Then $\frac{\sin\alpha}{v_1} = \frac{\sin\beta}{v_2}$, where v_1 and v_2 are the speeds of light in these two media.*

Proof. As we know, the ray of light travels between two points along a path that takes the minimal possible time. If we take points x_1 and x_2 on the ray, on different sides of the boundary, and denote the straight line of the boundary by l, then we obtain the following minimization problem:

$$\text{(5)} \qquad \begin{cases} f(x) \;=\; \frac{1}{v_a}\,|x - x_1| \;+\; \frac{1}{v_b}\,|x - x_2| \;\rightarrow\; \min\,, \\ x \in l \end{cases}$$

Arguing as in Example 3, we obtain that the point of minimum is characterized by the following property: the vector $\frac{1}{v_1}u_1 + \frac{1}{v_2}u_2$ is orthogonal to l. This is the same as saying that the sum of projections of $\frac{1}{v_1}u_1$ and $\frac{1}{v_2}u_2$ onto l is zero, that is, $\frac{\sin\alpha_1}{v_1} = \frac{\sin\alpha_2}{v_1}$.

◇

Linear algebra

Example 7 *A real symmetric $d \times d$-matrix A has a real eigenvalue.*

Proof. Consider the function $f(x) = \langle Ax\,, x\rangle$ and find its maximum on the unit sphere: $\langle Ax\,, x\rangle \to \max\,,\ \langle x,\, x\rangle = 1$. By the Weierstrass theorem a point of maximum does exists. Differentiating the Lagrangian gives $2Ax + 2\lambda x = 0$. Thus x is a real eigenvector corresponding to the eigenvalue $-\lambda$.

$\diamond$

The existence of a real eigenvalue for a symmetric matrix or, in other words, for a self-conjugate operator, is usually proved in a different way, by using the characteristic polynomial and the fundamental theorem of algebra. Our proof has one advantage: it can be extended to infinite dimensional operators without any change (Hilbert-Schmidt theorem, see, for instance [1]). Also this trick can be applied to prove the existence of solutions for some differential equations.

Inequalities.

Example 8 *[Cauchy's inequality for the arithmetic and geometric mean].*
For any nonnegative $x_1, \ldots, x_n$ we have $x_1^n + \ldots + x_n^n \ge nx_1 \cdots x_n$.

Proof. Denote $x_1^n + \ldots + x_n^n = a$ and consider the following problem:
$nx_1x_2\cdots x_n \to \max\,;\quad x_1^n + \cdots + x_n^n = a$ on the domain $x_i \ge 0,\ 1 \le i \le n$.
By the Weierstrass theorem there exists a point of maximum $(x_1, \ldots, x_n)$. Clearly, all x_i are strictly positive, otherwise $nx_1x_2\cdots x_n = 0$ and this is obviously not a maximum. Differentiating the Lagrangian we get

$$nx_1 \cdots x_n + \lambda n x_i^n \;=\; x_i \mathcal{L}_{x_i} \;=\; 0, \qquad i = 1, \ldots, n,$$

which implies $x_1 = \cdots = x_n$. Thus, at the point of maximum, $x_1 = \cdots = x_n$. But in this case $nx_1x_2\cdots x_n = x_1^n + \ldots + x_n^n = a$, hence at all other points $nx_1x_2\cdots x_n < a$, which proves the inequality.

$\diamond$

Inequalities are always connected with the solution of optimization problems. Therefore, they are a remarkable testing ground for the general theory. Most – maybe all – inequalities from the classical monographs [2] and [3] can be established in an efficient way using optimization methods. For example the following well-known inequalities:

[The Hölder inequality]. For every $p > 1$ and for all positive $x_1, \dots, x_n, y_1, \dots, y_n$,

$$\sum_{k=1}^{n} x_k y_k \;\le\; \Big(\sum_{k=1}^{n} x_k^p\Big)^{1/p} \Big(\sum_{k=1}^{n} y_k^q\Big)^{1/q},$$

where $q = p/(p-1)$. This inequality becomes an equality precisely if $x_k^p = \lambda y_k^q$ for some λ and all k. This follows directly from the multiplier rule.

[Inequality of Hadamard]. Let $A = (a_{ij})_{i,j=1}^n$ be a square matrix of order n. Then

$$(\det A)^2 \;\le\; \prod_{i=1}^{n} \big(\sum_{j=1}^{n} a_{ij}^2\big).$$

Olympiad problems

Many elementary problems from mathematical olympiads for high school students and for university students can easily be solved by the Lagrange multiplier rule. This concerns even olympiads of very high level. Many examples of this kind can be found in [4], [5]. It is most astonishing that problems, requiring tricky and sophisticated solutions, are carried out by the multiplier rule in a standard routine procedure.

Example 9 *[International Mathematical Olympiad, Washington, 1980]. Inside a given triangle find a point such that the sum of ratios of the sides of the triangle to the distances from this point to the lines containing these sides is minimal.*

Solution. Let a, b, c be the length of the sides, x, y, z be the distances from a point P to these sides respectively. To find a link between x, y and z, connect the point P with the vertices of the triangle by segments. Then three triangles appear and the sum of their areas is equal to the area S of the original triangle. Thus we have the problem

$$(6) \qquad \begin{cases} \frac{a}{x} + \frac{b}{y} + \frac{c}{z} \;\to\; \min \\ ax + by + cz \;=\; 2S \end{cases}$$

Differentiating the Lagrangian gives $-\frac{a}{x^2} + \lambda a \;=\; -\frac{b}{y^2} + \lambda b \;=\; -\frac{c}{z^2} + \lambda c \;=\; 0$, which immediately implies $x = y = z$. Thus, the point P is the center of the inscribed circle of our triangle.

Example 10 *[American Mathematical Olympiad, 1979].* Given an angle with vertex K and a point M inside the angle. Find points A and B on the sides of the angle such that the segment AB passes through the point M and the value $\frac{1}{MA} + \frac{1}{MB}$ is maximal.

Example 11 *[Soros Olympiad, Russia, 1997].* The distances from a point to the three vertices of a rectangular triangle are $2, 5$ and 10 (2 is the distance to the right angle). How large can the area of this triangle be?

Dynamical systems

Example 12 *[Birkhoff theorem]. For an arbitrary bounded convex body in $\mathbb{R}^2$ with a smooth boundary and for any $n \geq 3$ there exists a billiard with n vertices (a billiard is a polygon having its vertices on the boundary and possessing the property that two sides going from each vertex form equal angles with the boundary at this vertex).*

Proof. Denote the body by M and its boundary by ∂M. Consider the set of all polygons having n vertices, all lying on ∂M. Obviously this set is compact if we allow vertices to coincide and consecutive sides to lie on a common straight line. Therefore there exists a polygon of maximal perimeter. This is a desirable billiard. In the first place, it has exactly n different vertices, otherwise one can add extra vertices and the perimeter increases. Take now an arbitrary triple of consecutive vertices x_1, x_2, x_3 of this polygon and denote by l the tangent line to the curve ∂M at the point x_2. The point x_2 is a solution for the following maximization problem:

$f(x) = |x - x_1| + |x - x_3| \ \rightarrow \ \max, \ x \in \partial M$. Solving this as in Example 3, we obtain that the vectors $x - x_1$ and $x - x_3$ form equal angles with l. Therefore this polygon is a billiard.

$\diamond$

Billiards are phenomena used in ergodic theory, the study of dynamical systems and classical mechanics (see, for instance, [7]). In the proof of the Birkhoff theorem we have seen that for smooth convex curves billiards correspond to the inscribed polygons of maximal length. However, for acute triangles billiards correspond to the inscribed triangles of not maximal but *minimal* length (this is a triangle with vertices at the bases of altitudes [8]). How to explain this contradiction? Let us leave this question to the reader. We only note that the smoothness of the boundary is essential in the proof, it should be at least differentiable everywhere! For non-smooth convex curves Birkhoff theorem may fail. For instance, an obtuse triangle has no billiards of three vertices. It is still an open problem, whether it is true that any obtuse triangle has *at least one* billiard, not necessarily triangular.

Finance.

Example 13 [Risk minimization.] A person is planning to divide her savings among three mutual funds having expected returns of 10%, 10%, and 15%. Her goal is a return of at least 12%, while minimizing her risk. The risk function for an investment in this combination of funds is

$$200x_1^2 + 400x_2^2 + 100x_1x_2 + 899x_3^2 + 200x_2x_3,$$

where x_i is the proportion of her savings in fund i. Determine the proportions that should be invested in each fund. Would it help if she could go short, that is, if the x_i are allowed to be negative?

Economics.

We give a model that tries to capture that there can sometimes be a difference of interest between stakeholders and shareholders of a firm, which might lead to problems.

Example 14 [Shareholders versus stakeholders]. A firm has total revenue $TR = 40Q - 4Q^2 + 2A$, where Q is its output and A is its advertising expenditure. Its total costs are $TC = 2Q^2 + 20Q + 1 + 4A$. To encourage the managers, that is, the stakeholders, to perform well, their salary is linked to how well the firm is doing. For practical reasons it is made to depend on the total revenue of the firm, but not on the total costs. However, the profit of the firm is of importance as well: to be more concrete, the shareholders will not accept a profit of less than 3.

What will be the best choice of output and advertising expenditure from the point of view of the managers? Is this choice also optimal from the point of view of the shareholders?

The next example concerns a well-known application to economics [9]. The Lagrange multiplier rule gives the best explanation for the following fact:

Example 15 [Consumption problem.] A consumer maximizes his utility $U(x_1, \dots, x_n)$ subject to the budget constraint $p_1 x_1 + \cdots + p_n x_n = m$. Then in the optimal situation the marginal rate of substitution $\frac{\partial U(x)}{\partial x_k} / \frac{\partial U(x)}{\partial x_j}$ equals the price ratio $\frac{p_k}{p_j}$.

Bargaining.

Example 16 [Nash bargaining]. J.Nash has given a convincing answer to the question what is fair bargain. Three plausible axioms characterize a unique bargain as the solution of an optimization problem [10]. We consider a numerical example. Two individuals argue over which point $x = (x_1, x_2)^T$ satisfying the inequality $3x_1^2 + 4x_2^2 \leq 10$ should be adopted. If they agree on x, then the first individual receives a utility of x_1 units and the second a utility of x_2 units. If they fail to agree, it is understood that the result will be a given status quo point $s = (1, 1)^T$. The Nash bargaining solution is the solution of the problem

$$f(x) = (x_1 - s_1)(x_2 - s_2) \to \max, \quad 3x_1^2 + 4x_4^2 \leq 10,\ x_i \geq s_i,\ i = 1, 2.$$

Determine the Nash bargaining solution.

VI. Special tricks.

Let us now give some special tricks, which are common knowledge among users of the multiplier rule, but which by a conspiracy of silence are never written down.

First trick. *Find all variables, in which both the objective function and the constraints can be expressed in a simple and symmetric way.*

The first example is from mathematical economics [6]. The constraints are all linear, and the Lagrange multiplier rule allows us to keep their symmetry.

Example 17 *[Prediction of flows of cargo.] An investor wants to have information about the four flows of cargo within an area consisting of two zones, 1 and 2, including the flows within each zone. For both zones data are available to him, not only for the total flow originating in this zone,* $O_1 = 511$

and $O_2 = 1451$, *but also for the total flow with destination in this zone,* $D_1 = 1733$ *and* $D_2 = 229$. *Observe that*

$$S \;=\; O_1 + O_2 \;=\; D_1 + D_2 \;=\; \textit{total flow.}$$

However, the investor is not satisfied with this; he wants to have an estimation for T_{ij}, *the flow from* i *to* j, *measured in containers, for all* $i, j \in \{1, 2\}$. *What is the most probable distribution matrix* T_{ij} *given the available data and assuming that all units of cargo are distributed over the four possibilities with equal probability?*

Solution. We will use without proof the following approximate formula for the natural logarithm of the probability of a distribution matrix T:

$C - \sum_{i,j}[T_{ij}(\ln T_{ij}) - T_{ij}]$, where C is a constant which does not depend on the choice of T. It follows that the problem can be modeled as follows:

$$\begin{cases} \sum_{i,j}[T_{ij}(\ln T_{ij}) - T_{ij}] \;\to\; \min, \\ T_{i1} + T_{i2} = O_i, \quad T_{1j} + T_{2j} = D_j, \quad T_{ij} > 0 \quad \forall i, j. \end{cases}$$

Now we differentiate the Lagrangian function $\mathcal{L}_{T_{ij}} = 0 \;\Leftrightarrow\; \ln T_{ij} - \lambda_i - \lambda'_j = 0$. This yields $T_{ij} = e^{\lambda_i} e^{\lambda'_j}$. Substituting this in the four equality constraints of our problem we get $e^{\lambda_i} = \frac{O_i}{e^{\lambda'_1} + e^{\lambda'_2}}$ and $e^{\lambda'_j} = \frac{D_j}{e^{\lambda_1} + e^{\lambda_2}}$. Adding the first two equations gives $(e^{\lambda_1} + e^{\lambda_2})(e^{\lambda'_1} + e^{\lambda'_2}) = S$, which implies the required estimate: $T_{ij} = \frac{O_i D_j}{S}$.

The method we have just demonstrated is very flexible. For example, let us discuss the variant of the problem, for a general number of flows N, where we have the following additional information in advance: for a number of combinations of origin and destination we know that they have flow zero. Then we can proceed in the same way as above to write down an optimization problem, but now we add a constraint $T_{ij} = 0$ for each combination of origin i and destination j for which we know in advance that the flow will be zero. Then we apply the Lagrange multiplier rule and solve the resulting system of equations numerically.

Second trick. *In the solution of the Lagrange equations, you should ask yourself: which information the multipliers give on the solution. Usually it is not necessary to compute the Lagrange multipliers.*

In the following example the Lagrange equations give the information that all the variables x_i can be seen to satisfy the same equation, so each of them is contained in the set of roots of this equation.

Example 18 *If the sum of five values (not necessarily positive) is* 1*, and the sum of the squares is* 13*, what is the smallest possible value of the sum of the cubes?*

Solution. We have the following minimization problem:

$$\begin{cases} \sum\limits_{i=1}^{5} x_i^3 \ \to \ \min\,, \\ \sum\limits_{i=1}^{5} x_i^2 \ = \ 13\,, \quad \sum\limits_{i=1}^{5} x_i \ = \ 1. \end{cases}$$

Differentiating the Lagrangian we obtain the system of 5 equations:

$$3x_i^2 \, + \, 2\lambda_1 x_i \, + \, \lambda_2 \, = \, 0, \qquad i = 1, \ldots, 5\,. \tag{7}$$

All attempts to solve this system involving the two equality constraints $\sum\limits_{i=1}^{5} x_i^2 = 13$, $\sum\limits_{i=1}^{5} x_i = 1$ would lead to an equation of a high power. Instead we observe that we are interested in finding x_i only and do not actually need to find λ's. All five variables x_i satisfy the same quadratic equation (7), which has at most two real roots. Denote these roots by a and b. Thus, $x_i \in \{a, b\}$ for every $i = 1, \ldots, 5$. Interchanging, if necessary, a and b, we obtain three possible cases: (1) all x_i equal to a; (2) four of the x_i equal to a and the remaining one is b; (3) three of these values equal to a and two ones equal to b. Invoking the equality constraints we see that the first case is impossible, since the system $5a^2 = 13$, $5a = 1$ has no solution. In the second case we come to the system $4a^2 + b^2 = 13$, $4a + b = 1$ that has two solutions $(a, b) = (1, -3)$ or $\left(-\frac{3}{5}, \frac{17}{5}\right)$. In the third case we have $3a^2 + 2b^2 = 13$, $3a + 2b = 1$, which gives $(a, b) = \left(\frac{3+8\sqrt{6}}{15}, \frac{1-4\sqrt{6}}{5}\right)$ or $\left(\frac{3-8\sqrt{6}}{15}, \frac{1+4\sqrt{6}}{5}\right)$. One of these four suspicious points must be a point of minimum. Substituting into the objective function (the sum of cubes) and comparing the values, we obtain that the point of minimum corresponds to the case $(a, b) = (1, -3)$. Thus, $x_1 = \ldots = x_4 = 1, x_5 = -3$, four other points of minimum are obtained by rearrangements of the variables x_i. The minimal value of the objective function is -23.

We have solved this problem just by the observation that all variables satisfy the same equation (even with unknown coefficients!). Afterwards it remained to investigate several cases. This trick is often very efficient, especially if the problem possesses some symmetry. The problems from the following examples exploit the same idea:

Example 19 *[The problem of the maximal entropy].* For n positive numbers $x_1, \dots, x_n$ such that $\sum_{k=1}^n x_k = 1$ find the minimal possible value of the sum $\sum_{k=1}^n x_k \ln x_k$ (minus this sum is the entropy)

Example 20 What are the largest and the smallest possible values of the sum of squares of n numbers, if the sum of the fourth powers equals 1?

Example 21 $\sum_{i=1}^5 x_i^4 \to \mathrm{extr},\ \sum_{i=1}^5 x_i = \sum_{i=1}^5 x_i^3 = 0,\ \sum_{i=1}^5 x_i^2 = 4.$

References

[1] A.N.Kolmogorov , S.V.Fomin, *Introductory real analysis*, Dover Publications inc., New York, 1975, 403 pp.

[2] G.Hardy, J.E.Littlewood, G.Polya, *Inequalities*, Cambridge University Press, 1934.

[3] O.Bottema, R.Z.Djordjevic, R.R.Janic, D.S.Mitrinovic, P.M.Vasic, *Geometric inequalities*, Wolters-Noordhoff Publishing, Groningen, 1969, 151 pp.

[4] V.M.Tikhomirov, *Stories about Maxima and Minima*, American Mathematical Association, 1990.

[5] J.Brinkhuis, V.Protasov, *Theory of extremum in simple examples*, Mathematical Education, 9 (2005), pp. 32-55 (in Russian).

[6] J.Brinkhuis, V.Tikhomirov, *Optimization: Insights and Applications*, Princeton Series in Applied Mathemathics, Princeton, N.J., 2005, 682 pp.

[7] Ya.G.Sinai, *Introduction to ergodic theory*, Translated by V. Scheffer. Mathematical Notes, 18. Princeton University Press, Princeton, N.J., 1976, 144 pp.

[8] H.S.M.Coxeter, S.Greitzer, *Geometry revisited*, The L. W. Singer Company, Random House, New York , 1967, 193 p..

[9] K.Sydsaeter, P.Hammond, *Essential Mathematics for Economic Analysis*, 1995.

[10] K.Binmore, *Fun and Games, A Text on Game Theory*, Lexington, 1992.

[11] A.M.Yaglom, I.M.Yaglom, *Challenging mathematical problems with elementary solutions. Vol. II. Problems from various branches of mathematics*, Translated from the Russian by James McCawley, Jr. Reprint of the 1967 edition. Dover Publications, Inc., New York, 1987. 214 pp.

Jan Brinkhuis, Erasmus Universiteit Rotterdam, Econometrisch Instituut, Faculteit der Economische Wetenschappen, H 11-16, Postbus 1738, 3000 DR Rotterdam, The Netherlands, e-mail: brinkhuis@few.eur.nl

Vladimir Protasov, Department of Mechanics and Mathematics, Moscow State University, Vorobyovy Gory 3, Moscow, 119992, Russia,
e-mail: vladimir_ protassov@yahoo.com