Thai Journal of **Math**ematics
Special Issue: The 17$^{\text{th}}$ IMT-GT ICMSA 2021
Pages 54–61 (2022)

http://thaijmath.in.cmu.ac.th

ISSN 1686-0209

Lagrange Multipliers, an Elementary Practical Zeroth-Order Approach

Yutthapong Pianroj[1], Seppo Karrila[1], Kittisak Tinpun[2], Aniruth Phon-On[2,*]

[1] *Faculty of Science and Industrial Technology, Prince of Songkla University (PSU), Surat Thani campus, Surat Thani 84000, Thailand*
e-mail : seppo.karrila@gmail.com (SK); yutthapong.p@psu.ac.th (YP)

[2] *Department of Mathematics and Computer Science, Prince of Songkla University, Pattani campus, Pattani 94000*
e-mail : kittisak.ti@psu.ac.th (KT); aniruth.p@psu.ac.th (AP, corresponding author)*

Abstract While Lagrange multipliers are widely used to solve constrained optimization problems, their introduction in mathematical analysis is at a fairly advanced level coupled with the implicit function theorem, and then only for equality constraints. The case with inequality constraints tends to be left to the study of convex optimization, in which context the only allowed equality conditions are linear (called affine when there is a constant term). To gain rigor, one can lean on the equality constrained case treated in math courses, by use of slack variables. In both cases above, usually multivariate calculus and linear algebra are prerequisites, but then in practical applications it is recognized (with some exceptions) that the various assumptions and conditions cannot be checked ahead of just trying out to see if it works. We pursue the practical results here with minimal demands on background, and in particular the text includes no mention of a gradient and no differentiability is required – this is a zeroth-order approach. For this, we begin with inequality constraints, and then use the results to address equality constraints. The statements here take a simple form: If you can minimize the Lagrangian, then, and it is this fact that helps avoid complications with assumptions and conditions (that were unverifiable anyway, before trying out if it works). The proposed approach allows introducing the Lagrange multipliers in a rigorous manner at an early stage to mathematics, physics, or engineering students, or to non-mathematicians not eager to study advanced multivariate calculus.

MSC: 58E17; 49N15; 49J52; 91B02

Keywords: constrained nonsmooth optimization; zeroth-order theory; Lagrange multipliers; tutorial notes

Submission date: 15.03.2022 / Acceptance date: 31.03.2022

1. Introduction

In the following we follow a derivative-free zeroth-order approach to Lagrange multipliers for inequality constrained problems, effectively duplicating some of the results

*Corresponding author.

Published by The Mathematical Association of Thailand.

published by Everett in the 1960s [1], although not following that reference faithfully but only using its approach. That prior work seems to have been forgotten, and has not influenced teaching or textbooks much as regards the presentation of Lagrange multipliers. We then deviate further from that work and address the equality constrained case, also in a derivative-free manner. This approach makes it possible to introduce some constrained optimization tools with minimal demands on mathematical background. For convenience and clarity we name the problems as follows:

P1: Minimize $f(x) + g(x)$.
P2: Minimize $f(x)$ subject to $g(x) \leq g_{max}$.
P3: Minimize $f(x)$ subject to $g_i(x) \leq g_i(x^*)$ for all $i = 1, \ldots, m$.
P4: Minimize $f(x)$ subject to

$$g_i(x) \leq g_i(x^*) \quad \text{for} \quad u_i > 0; \qquad \text{and} \quad g_j(x) \geq g_j(x^*) \quad \text{for} \quad u_j < 0.$$

P5: Minimize $f(x)$ subject to $g_i(x) \leq c_i$ for all $i = 1, \ldots, m$.
P6: Minimize $f(x)$ subject to $h_k(x) = c_k$ for all k.
P7: Minimize $f(x)$ subject to $g_i(x) \leq 0$ for all $i = 1, \ldots, m$.

2. The approach

2.1. On minimizing sums of functions

Given real valued functions f and g defined in a shared domain that can be anything, even discrete, consider minimizing their sum $f(x) + g(x)$, and assume the least value is achieved at x^*. Then the two functions cannot at the same time get values less than those at x^* – otherwise their sum would also be less than the minimum found.

Let us visualize what is happening. For the sake of discussion, let the minimum of the sum equal c. Then all (f, g) value pairs are above and to the right of the $f + g = c$ line in a plot with axes for f and g, with x^* giving a point on this line of slope -1. Restricting g to values only up to this point, $g(x) \leq g(x^*)$, makes sure that points above and to right of the line will not come below $f(x^*)$. In other words

$$x^* \quad \text{minimizes} \quad f(x) \quad \text{subject to} \quad g(x) \leq g(x^*).$$

Alternatively, appealing to symmetry, we get the reciprocal statement (a case of Tikhonov reciprocity) that

$$x^* \quad \text{minimizes} \quad g(x) \quad \text{subject to} \quad f(x) \leq f(x^*).$$

So we have the following Lemma.

Lemma 2.1. *If x^* minimizes $f(x) + g(x)$, then it also minimizes $f(x)$ among those x satisfying $g(x) \leq g(x^*)$.*

Proof. Assume that x^* solves P1 so for all x

$$f(x) + g(x) \geq f(x^*) + g(x^*).$$

Then by using the constraint on g we have

$$f(x) \geq f(x^*) + g(x^*) - g(x) \geq f(x^*)$$

proving the assertion. ∎

This means that if we can solve the unconstrained problem P1, we also gain a solution to the constrained problem P2, for the specific case $g_{max} := g(x^*)$. This is significant because P1 is typically easier to solve than P2.

However, the reverse does not hold: solutions to P2 need not be solutions to P1, as demonstrated in this example.

Example 2.2. Suppose f gets the constant value 1, while g gets values from 0 to 10. Then the minimum of their sum is 1, and this requires that $g(x^*) = 0$. In contrast, $f(x) = 1$ is constantly at its minimum, and ANY x with $0 \leq g(x) \leq 10$ would minimize f, so we clearly can have solutions to P2 not found by solving P1.

Now we note that if, in 2.1, we substitute for $g(\cdot)$ its scaled version $ag(\cdot)$ with $a > 0$, this should affect the solution x^* and through it the constraint, but the scaling cancels out from the latter. In other words, $g_{max} := g(x^*)$ can be manipulated by adjusting the coefficient a.

Let us now examine the effects of such scaling. The monotonic effects of the weight, and its role as a 'shadow price' in terminology used by economists, are the topics of the following theorem.

Theorem 2.3. *(Shadow prices) Let $0 \leq a < b$ and let x^* minimize $f(x) + ag(x)$ while y^* minimizes $f(y) + bg(y)$. Then $g(y^*) \leq g(x^*)$, $f(x^*) \leq f(y^*)$, and*

$$0 \leq a[g(x^*) - g(y^*)] \leq f(y^*) - f(x^*) \leq b[g(x^*) - g(y^*)].$$

Proof. We use the fact that x^* and y^* are minimizers, put on the left sides of these inequalities:

$$f(x^*) + ag(x^*) \leq f(y^*) + ag(y^*)$$

and

$$f(y^*) + bg(y^*) \leq f(x^*) + bg(x^*).$$

Adding these, canceling terms, and moving everything on one side gives

$$(a - b)[g(x^*) - g(y^*)] \leq 0.$$

Noting the $a-b < 0$ immediately gives the monotonic response of g-value at the optimum.

Rearranging the two inequalities and then joining them gives

$$a[g(x^*) - g(y^*)] \leq f(y^*) - f(x^*) \leq b[g(x^*) - g(y^*)]$$

This also shows the monotonic response claimed for f, since the bounds on left and right are both nonnegative. ∎

In Theorem 2.3, the first two inequalities are intuitively easy to accept and can be perceived in a practical way: b gives more weight on g than a does, so g gets better minimized with b than with a. Conversely minimizing f suffers more when g is given more weight. These results show that the value component responses to minimizing a weighted sum behave monotonically with relative weighing of two summands. However, these component responses are in opposite directions, so we have no result on whether the minimum of the weighted sum increases or decreases except if we know that g has a fixed sign at these optima. For example, if $g \leq 0$, then the weighted sum decreases as multiplier a is changed to the larger b. Its minimum must also decrease (perhaps not strictly, but it certainly does not increase). This is among the reasons to put an inequality constrained problem to the form P7, with all bounds at zero.

The more complicated last inequality shows that the two weights are prices paid for change in g, to obtain a change in f in other words, conversion factors. Note that each quantity in brackets is nonnegative. If there is no change in g from one optimal point to the other, then there is no change in f either. When there are changes we can divide to get

$$a \leq [f(y^*) - f(x^*)]/[g(x^*) - g(y^*)] \leq b.$$

In the limit this difference quotient could turn into a derivative, as a and b converge to a single Lagrange multiplier representing a sharp price that f should pay for unit change in g provided that the assumed optima exist.

Obviously Lemma 2.1 holds for longer sums with more terms. We rephrase (omitting multipliers to begin with) a theorem given by Everett in 1963 [1] as follows.

Theorem 2.4. *(Everett, 1963) Let x^* minimize the sum $L(x) = f(x) + \sum_i g_i(x)$ Then x^* also solves* **P3**.

In the proof of 2.1 one only needs to substitute for g the sum term above, and express the difference of sums as the sum of (positive) differences.

The strength of this fairly elementary observation is in the fact that it is amazingly general. We need not specify what kind of creatures the x are, or what types of functions we have this theorem is easy to recall precisely because it places no assumptions. Actually, the assumption is hidden, requiring that you can minimize $L(x)$, which in the next corollary takes the common form of a Lagrangian this is why we denote it with L.

An engineer would want the units to match before taking a sum and is familiar with conversion factors, while an economist tends to look at costs and uses prices to compare items on an equal footing. So now we get to the multipliers.

Corollary 2.5. *Let u_i be real numbers and let x^* minimize the weighted sum $L(x) = f(x) + \sum_i u_i g_i(x)$. Then x^* also solves* **P4**.

All one needs to note is that a negative multiplier turns an inequality around, while of course zero multipliers can be ignored. The coefficients u_i in $L(x)$ can be called Lagrange multipliers.

We also note that Everett claims *generalized* Lagrange multipliers, and interpret this as a reference to having a zeroth-order theory that does not require existence of derivatives (or gradients), while the conventional explanations or derivations of the multipliers do require them.

2.2. The difficulty with inactive constraints

Commonly we would have several given resource constraints, like availability of some raw materials, or labor, or quality requirements on product that is made. These are naturally posed as inequalities.

The constrained problem would then look like **P5**: Minimize $f(x)$ subject to $g_i(x) \leq c_i$ for all i, where the constraints can be reshaped to $\tilde{g}_i(x) := g_i(x) - c_i \leq 0$ to get exactly zero bounds.

To use the above connection with a weighted sum, we note that if x^* is the solution to P5, then some of the constraints are active or binding so that we are at the allowed limit, as in $\tilde{g}_1(x^*) = 0$, while some other constraints may remain inactive at the solution and could be entirely ignored. Of course the constraints that can be ignored should not participate in a sum that is minimized either in this sum they should have zero multiplier or weight.

The difficulty now is that we cannot know in advance which ones are active constraints at the eventual optimum, and partly it is this generic problem that makes inequality constrained optimization (or 'programming') difficult even when the constraints are linear (with a shift) so their combinations can be solved: typically there are far too many such potential combinations. This difficulty is why we pursue unconstrained formulations where the Lagrangian is minimized, to possibly find a solution of the original constrained problem.

Now we note that the zero-bounded constraints allow extending our shadow price result as follows. In the Lagrangian, the non-binding terms are dropped out, since those constraints are ignorable. The rest that are binding contribute only zeroes. Therefore $f(x^*) = L(x^*)$. Now changing one multiplier as in Theorem 2.3 will affect the rest of the Lagrangian, but in it only $f(x^*)$ remains – therefore Theorem 2.3 remains valid despite having multiple constraints.

An alternative way to view this is to have the other constraints implemented by restricting the allowed domain of x, so those inequalities can thereafter be forgotten. Since our zeroth-order derivations assume nothing about the domain, this approach is allowed and reduces the number of inequality constraints to one – and again the shadow price effect is valid based on Theorem 2.3.

2.3. The equality constrained case

We now revert our attention to the equality constrained case, for some easy progress. The equality constrained problem is here defined as **P6**: minimize $f(x)$ subject to $h_k(x) = c_k$ for all k in a finite index set. Every one of the equality conditions is required at an acceptable solution. A straightforward approach would be to eliminate one component of x vector per equality constraint, then optimize f over only the remaining free variables. However, this is seldom possible even in principle, since we lack analytic solutions to even simple appearing equations. It is much easier to form a weighted sum and seek its minimum.

Let us recall Corollary 2.5 to Everetts theorem, in which zero weights are ignored as meaningless.

Let u_k be real numbers and let x^* minimize the weighted sum

$$L(x) = f(x) + \sum_k u_k h_k(x).$$

Then x^* also minimizes $f(x)$ subject to

$$h_k(x) \leq h_k(x^*) \quad \text{for} \quad u_k > 0 \quad \text{and} \quad h_j(x) \geq h_j(x^*) \quad \text{for} \quad u_j < 0.$$

If we now require that $h_k(x^*) = c_k$ for all k, then x^* satisfies the equality conditions, and the corollary ensures it is optimal in the set allowed by the inequalities, therefore certainly optimal in the smaller set allowed by only equalities. We have reached a slightly unusual form of the Lagrange multiplier theorem, namely:

Theorem 2.6. *If for some choice of real multipliers u_i, the minimizer x^* of the Lagrangian*

$$L(x) = f(x) + \sum_i u_i h_i(x)$$

satisfies the equality conditions $h_i(x^*) = c_i$ *for all* i*, then it is also a solution to the equality constrained problem* **P6** *above.*

Please note that the multipliers need not be nonnegative, they can be any real numbers. Also note that we have not needed any assumptions about existence of derivatives, or even about the type of domain in which x lives. In case both equality and inequality constraints are given, the multipliers for inequalities have a sign restriction as seen in Corollary 2.5, while those for equalities are not restricted as seen above.

2.4. The complementarity in KKT conditions

This optional section takes the derivative of a square.

Let us first rename the functions $\tilde{g}_i$ back to g_i so that we have

P7: Minimize $f(x)$ subject to $g_i(x) \leq 0$ for all i.

As a reminder, this guarantees monotone response of the Lagrangian to the multipliers, since now the sign of every g_i is fixed at an allowed optimal solution. These inequality constraints can be modified with 'slack variables' s_i that can be chosen to make every one of the constraints binding (i.e., the inequality is actually equality) at a suitably chosen solution. Obviously the modification does not essentially change the problem: solution of either one generates a corresponding solution of the other.

P7-b: Minimize $f(x)$ subject to $h_i(x, s) = g_i(x) + s_i^2 = 0$ for all i.

Appealing to Theorem 2.6, we take multipliers λ_i and assume that x^* with s^* minimize the Lagrangian

$$L(x, s) = f(x) + \sum_i \lambda_i h_i(x, s).$$

Requiring zero derivative with respect to s_i gives $\lambda_i s_i^* = 0$. In other words, either there is no slack and $s_i^* = 0$, or if there is slack, then the multiplier $\lambda_i = 0$. This is known as the complementarity condition in the celebrated Karush-Kuhn-Tucker optimality conditions, typically called KKT conditions. The previous section already discussed this less formally: inactive constraints should not enter the Lagrangian that is minimized.

3. Discussion

Practical implications

The common presentation (even introduction) of Lagrange multipliers tends to suffer from complications related to an approach from multivariate calculus. That by itself makes demands on background of the audience, while the above shows that a very elementary approach could be taken. Of course, on applying the last theorem above, to seek the minimizer of the Lagrangian one would (in analytical examples) start taking derivatives, one for each component in x, and setting them to zero. Then one would try to adjust the multipliers so that x^* satisfies the equality conditions. Finally, the candidate solutions found would have to be inspected. However, it may also be that a fully numerical approach is pursued, possibly entirely derivative-free. Various conditions could be listed that ensure existence of the Lagrange multipliers, such that allow finding an x^* that satisfies the constraints. However, usually such conditions cannot be checked beforehand anyway, instead one simply performs the same steps as above and perhaps does a post mortem check of the candidate solutions for such conditions. An exception is convex optimization [2, 3], on which there is extensive and well developed theory with various so-called constraint qualifications that guarantee success with the Lagrangian. However, in that context equality constraints are limited to only linear ones.

The reciprocal relation that almost automatically presented itself from symmetry of minimizing a sum has been studied, and relevant literature can be found by searching for Tikhonovs reciprocity principle.

Obviously a lot has been left out. The existence of Lagrange multipliers is one such aspect, and in mathematical analysis it can be pursued in connection with the multivariate implicit function theorem, for equality constraints [4]. Mathematical physics makes extensive use of Lagrange multipliers, and typically presents them without rigor, but with a multitude of convincing analytical examples [5] where direct elimination of variables is utterly impossible. We have not bothered to include any such examples here, they are easily available to teachers or students who want them.

With the emergence of machine learning, ever larger scale problems are tackled computationally. This has forced a move away from second order methods (that would require generating an enormous matrix of second derivatives) to first order methods, and to nonsmooth optimization. A zeroth-order approach to Lagrangian optimization can be defended and motivated by these trends. However, our motivation has mainly been a curiosity regarding what understanding of constrained optimization could be reached by such derivative-free approach, which poses minimal demands on the reader's background. It might be surprising that rigorous inequalities, such that also make sense intuitively and could be recalled with ease, were demonstrated for shadow prices of constraint bounds with zeroth-order methods.

4. Conclusion

We have walked through an inverted approach to Lagrange multipliers: instead of first presenting the equality constrained case, and either later or never introducing inequality constraints, we have started with the latter. It turns out that an elementary approach, purely algebraic and accessible with very modest background in math, allows reaching simple but quite powerful results for inequality constrained optimization. This material dates back to the 1960s but remains not widely known. As a curiosity, we touched also on reciprocity principles. What appears to be a novelty is to then revert to equality constraints, reaching a practical theorem very effortlessly. This allows presenting the Lagrange multipliers also for equality constrained optimization to a much wider audience than other approaches familiar to these authors, and in a form that is easy to recall since no list of assumptions or conditions is involved. Granted that, unfortunately, some of the magic in applying the Lagrange multipliers is lost, if they are given an easily understood explanation.

Acknowledgements

We would like to thank the referees for their comments and suggestions on the manuscript.

References

[1] H. Everett, Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources, Oper. Res. 11 (1963) 399417.

[2] S.P. Boyd, L. Vandenberghe, Convex optimization, Cambridge University Press; Cambridge, UK, New York, 2004.

[3] N. Lauritzen, Undergraduate convexity: from Fourier and Motzkin to Kuhn and Tucker, World Scientific; Singapore, Hackensack, NJ, 2013.

[4] J.C. Burkill, H. Burkill, A Second Course in Mathematical Analysis, Cambridge University Press; New York, NY (2002)

[5] G.B. Arfken, H.J. Weber, F.E. Harris, Mathematical Methods for Physicists, ELSEVIER INDIA; Place of publication not identified, 2012.