KKT Conditions for variational problem involving higher-order derivatives

Lokesh Kumar[1], Sandeep Suman[2], Shubhra[3*]

[1]Department of Mathematics, S. M. College, TMBU, Bhagalpur, Bihar, 812001, India.
[2]University Department of Mathematics, TMBU, Bhagalpur, Bihar, 812007, India.
[3]Department of Mathematics, Patna University, Patna, Bihar, 800005, India.

*Corresponding author(s). E-mail(s): drshubhrasuman@gmail.com;
Contributing authors: lokeshkmath@gmail.com; drsandeepsuman@gmail.com;

Abstract

In this article, we will treat a generalization of the Lagrange problem of the calculus of variation by adding higher-order derivatives in the functional. We convert this to a mathematical programming problem and derive the KKT necessary conditions. The final result will be a generalization of Euler-Lagrange equations.

Keywords: KKT conditions, variational problem, mathematical programming, convex optimization

1 Introduction and Prelims

Mathematical programming[1] is concerned with the optimization of the function over some vector space. Mathematical programming problems often deal with constraints. On the other hand, variational problems are associated with the optimization of a functional. Classically these problems have no constraints. Mathematical programming and variational problems have different origins in history. A variational problem can be traced back to Newton and Bernoulli. Thus it is older than the modern optimization itself. Since both the fields have grown independently, each field can get advantage from the other field. In this article, we will handle a variational problem as a mathematical programming problem. This article also serves as an example of analyzing a variational problem as a mathematical programming problem.

The general optimization problem in which, the variables are continuous, can be written as follows

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \geq 0 \\ & h_i(x) = 0 \\ & x \in \mathbb{R}^k \end{aligned} \tag{1}$$

Here $x \in \mathbb{R}^k$, and $f, g_i, h_i : \mathbb{R}^k \to \mathbb{R}$ are scaler functions. This general optimization problem 1 can be classified based on the properties of the functions f, g_i, h_i and the domain of the decision variable x.

In order to further generalize this optimization problem we can write g_i as vector function $\mathbf{g} = (g_1, \ldots, g_m)$ and $\mathbf{h} = (h_1, \ldots, h_n)$, Now the above problem can be represented as

$$\min_{\mathbf{x}\in\mathbb{R}^k} F(\mathbf{x}) \quad \text{subject to } -\mathbf{g}(\mathbf{x}) \in \mathbb{R}^m_+ \qquad \mathbf{h}(\mathbf{x}) = \mathbf{0}. \tag{2}$$

Here $\mathbb{R}^m_+ = \{\mathbf{x} = (x_1, \ldots, x_m) \in \mathbb{R}^m \mid x_i \geq 0\}$ is called the non-negative orthant of $\mathbb{R}_m$[2].

In mathematical programming the decision variable can be a function, hence it generalizes the notion of optimization in which the decision variable are elements of real numbers. In variational problems the decision variable are itself functions. The space of continuous functions form a vector space, hence this problem lies under the mathematical programming problems. The constraints are inequality involving scalar functions of function. If we have a system of the constraints $g_i(x) \leq 0, 1 \leq i \leq m$, it means that

$$g_i(x)(t) \leq 0, \quad t \in [a, b] \tag{3}$$

If $\mathbf{S}$ is the set of all continuous functions, $\phi : [a, b] \subset \mathbb{R} \to \mathbb{R}^m$, such that,

$$\phi(t) \in \mathbb{R}^m_+, \quad t \in [a, b]$$

Then the constraints 3 in the vector form can be written as $-\mathbf{g}(x) \in \mathbf{S}$. This set $\mathbf{S}$ is a convex cone.

Now we will replace the decision variable $\mathbf{x}$ with the elements of arbitrary vector space $\mathbf{z} \in V$, then the general problem of mathematical programming will be like

$$\min_{\mathbf{z}\in V} \quad F(\mathbf{z}) \tag{4}$$

$$\text{subject to} \quad -\mathbf{g}(\mathbf{z}) \in \mathbf{S} \tag{5}$$

$$\mathbf{h}(\mathbf{z}) = \mathbf{0}$$

Where $\mathbf{F}$ is a scaler function, $g : V \to \mathbb{R}^m$ and $\mathbf{S}$ is a convex cone. Hence a mathematical programming problem often represents a large class of optimization problems compared to 1.

2 Variational Problem

Mond and Hanson [3] discussed the duality for a modified Lagrange problem. In this problem, the functional dependents on the first-order derivative. Husain and Jabeen [4] generalized this problem by adding a second-order derivative in the functional. Here we will consider the problem in which the objective functional will depends on the derivatives of arbitrary order. Variational problems in higher-order are a special case of the variational problem involving vector function[5, Section 7.4].

In this section we will discuss a generalization of Lagrange problem including arbitrary order derivatives as follows

$$\begin{aligned} \min_{x} \quad & \int_I f(t, x, x', \ldots, x^p)dx \\ \text{subject to} \quad & x^i(a) = A_i, x^i(b) = B_i,\ i = 1, \ldots, p-1 \\ & Q(t, x, x', \ldots, x^p) \geq 0 \end{aligned} \tag{6}$$

Here,

1. $I = [a, b] \subset \mathbb{R}$
2. $f : I \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n ((p+1)-\text{times}) \to \mathbb{R}$ belongs to the set of continuously differentiable function.
3. X is the space of all piecewise smooth function, $x : I \to \mathbb{R}^n$ having derivative up to order p. In order to make this function continuous we use the following norm

$$||x|| = ||x||_\infty + ||Dx||_\infty + \cdots + ||D^p x||_\infty.$$

   Here D is the differentiation operator given by

$$y = Dx \Leftrightarrow \alpha + \int_I y(s)ds,$$

   hence, $D \equiv \dfrac{d}{dx}$ except at the discontinuities, where α is given boundary values.
4. $Q : I \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n ((p+1) - \text{times}) \to \mathbb{R}^m$ is also a continuously differentiable function.

The main idea to represent the problem as a mathematical programming problem, more particularly a convex optimization problem. Convex optimization [6] is currently the most important generalization of optimization problem for practical purposes. Now the KKT necessary conditions for this problem will be derived as the following theorem [7].

Theorem 1 (KKT Conditions)**.** *Let $\bar{x}$ is any optimal solution and solvable for the problem 6, then there exit Lagrange multiplier $\lambda : I \to \mathbb{R}^m$ is a piecewise smooth and it satisfies the following set of equations*

$$\begin{aligned} & f_x(t, \bar{x}, \bar{x}', \ldots, \bar{x}^p) - \lambda(t)^T Q_x(t, \bar{x}, \bar{x}', \ldots, \bar{x}^p) \\ & = \begin{bmatrix} D\left(f_{x'}(t, \bar{x}, \bar{x}', \ldots, \bar{x}^p) - \lambda(t)^T Q_{x'}(t, \bar{x}, \bar{x}', \ldots, \bar{x}^p)\right) \\ - D^2\left(f_{x''}(t, \bar{x}, \bar{x}', \ldots, \bar{x}^p) - \lambda(t)^T Q_{x''}(t, \bar{x}, \bar{x}', \ldots, \bar{x}^p)\right) \\ \vdots \\ + (-1)^{(p-1)} D^p\left(f_{x^p}(t, x, x', \ldots, x^p) - \lambda(t)^T Q_{x^p}(t, \bar{x}, \bar{x}', \ldots, \bar{x}^p)\right) \end{bmatrix} \end{aligned} \tag{7}$$

$$\begin{aligned} \lambda(t)^T Q_x(t, \bar{x}, \bar{x}', \ldots, \bar{x}^p) = 0, t \in I \\ \lambda(t) \geq 0, t \in I \end{aligned} \tag{8}$$

Proof. This variational problem can be written as a mathematical programming problem 4 as follows

$$\min_{\mathbf{x} \in X} F(\mathbf{x}) \qquad \text{subject to } \mathbf{G}(\mathbf{x}) \in \mathbf{S} \tag{9}$$

Here, the objective function is

$$F(\mathbf{x}) = \int_I f(t, x, x', \ldots, x^p) dx,$$

the set

$$\mathbf{S} = \{\phi : I \to \mathbb{R}^m \mid \phi(t) \in \mathbb{R}^m_+\},$$

forms a convex cone and the constraints is given by

$$G : X \to S, \quad \text{where } G(x)(t) = g(t, x, x', \ldots, x^p).$$

Since F and G are Fréchet differentiable, the KKT conditions for the problem 9 [2, Section 3.5] says that, for any optimal solution $\bar{\mathbf{x}}$ of it, we get $\mu \in \mathbf{S}^*, i.e.$, dual of $\mathbf{S}$, such that,

$$F'(\bar{\mathbf{x}}) - \mu^T G'(\bar{\mathbf{x}}) = 0 \tag{10}$$

$$\mu^T G(\bar{\mathbf{x}}) = 0 \tag{11}$$

$$\mu \geq 0 \tag{12}$$

where $F'(\bar{\mathbf{x}})$ and $G'(\bar{\mathbf{x}})$ represent Fréchet derivatives.

Now for $x, v \in X$, we have

$$F(\bar{\mathbf{x}}+\mathbf{v}) - F(\bar{\mathbf{x}}) = \int_I \begin{bmatrix} f_x(t,x,x',\ldots,x^p)v(t) + f_{x'}(t,x,x',\ldots,x^p)v'(t) \\ + \cdots + f_{x^p}(t,x,x',\ldots,x^p)v^p(t) \end{bmatrix} dt + \mathbf{0}||\mathbf{v}|| \tag{13}$$

Hence,

$$F(\bar{\mathbf{x}}+\mathbf{v}) - F(\bar{\mathbf{x}}) = \int_I \begin{bmatrix} f_x(t,x,x',\ldots,x^p)v(t) + f_{x'}(t,x,x',\ldots,x^p)v'(t) \\ + \cdots + f_{x^p}(t,x,x',\ldots,x^p)v^p(t) \end{bmatrix} dt \tag{14}$$

Now any μ in the dual space of $\mathbf{S}$,there is a measurable function $\lambda : I \to \mathbb{R}^m$ [2, Section 2], such that,

$$\langle \mu, v \rangle = \int_I \lambda(t)^T v(x) dt, \quad \forall v \in X \tag{15}$$

Also for $v \in X$, we have

$$\mu^T \mathbf{G}(\bar{\mathbf{x}})\mathbf{v} = \int_I \lambda(t)^T \begin{bmatrix} g_x(t,x,x',\ldots,x^p)v(t) + g_{x'}(t,x,x',\ldots,x^p)v'(t) \\ + \cdots + g_{x^p}(t,x,x',\ldots,x^p)v^p(t) \end{bmatrix} dt \tag{16}$$

Now using equations 14 and 16 in the first KKT equation 10, we get for all $v \in X$ after suppressing the arguments of all functions, we have

$$\begin{aligned} 0 &= F'(\bar{\mathbf{x}}) - \mu^T G(\bar{\mathbf{x}}) \\ &= \int_I [f_x v + f_{x'} v' + \cdots + f_{x^p} v^p] dt - \int_I \lambda^T [g_x v + g_{x'} v' + \cdots + g_{x^p} v^p] dt \\ &= \int_I (f_x - \lambda^T g_x) v dt + \int_I (f_{x'} - \lambda^T g_{x'}) v' dt + \cdots + \int_I (f_{x^p} - \lambda^T g_{x^p}) v^p dt \end{aligned}$$

using integration by parts, we get

$$\begin{aligned} &= \int_I (f_x - \lambda^T g_x) v dt \\ &\quad + \left[(f_{x'} - \lambda^T g_{x'}) v\right]_{t=a}^{t=b} - \int_I \left[D(f_{x'} - \lambda^T g_{x'})\right] v dt \\ &\quad \vdots \\ &\quad + \left[(f_{x^p} - \lambda^T g_{x^p}) v^{p-1}\right]_{t=a}^{t=b} - \int_I \left[D(f_{x^p} - \lambda^T g_{x^p})\right] v^{(p-1)} dt \end{aligned}$$

as $v^i(a) = v^i(b) = 0,\ 0 \le i \le p-1$, we get

$$
\begin{aligned}
&= \int_I \left[(f_x - \lambda^T g_x) - D(f_{x'} - \lambda^T g_{x'})\right] v dt \\
&\quad - \int_I \left[D(f_{x''} - \lambda^T g_{x''})\right] v' dt \\
&\quad \vdots \\
&\quad - \int_I \left[D(f_{x^p} - \lambda^T g_{x^p})\right] v^{(p-1)} dt
\end{aligned}
$$

again integrating by parts, we get

$$
\begin{aligned}
&= \int_I \left[(f_x - \lambda^T g_x) - D(f_{x'} - \lambda^T g_{x'})\right] v dt \\
&\quad - \left[(f_{x''} - \lambda^T g_{x''}) v\right]_{t=a}^{t=b} + \int_I \left[D^2(f_{x''} - \lambda^T g_{x''})\right] v dt \\
&\quad \vdots \\
&\quad - \left[(f_{x^p} - \lambda^T g_{x^p}) v^{(p-2)}\right]_{t=a}^{t=b} + \int_I \left[D^2(f_{x^p} - \lambda^T g_{x^p})\right] v^{(p-2)} dt
\end{aligned}
$$

as $v^i(a) = v^i(b) = 0,\ 0 \le i \le (p-2)$, we get

$$
\begin{aligned}
&= \int_I \left[(f_x - \lambda^T g_x) - D(f_{x'} - \lambda^T g_{x'}) + D^2(f_{x''} - \lambda^T g_{x''})\right] v dt \\
&\quad + \int_I \left[D^2(f_{x'''} - \lambda^T g_{x'''})\right] v' dt \\
&\quad \vdots \\
&\quad + \int_I \left[D^2(f_{x^p} - \lambda^T g_{x^p})\right] v^{(p-2)} dt
\end{aligned}
$$

procceding in the same way untill we exhaust the derivatives of v, we get

$$
= \int_I \left[(f_x - \lambda^T g_x) - D(f_{x'} - \lambda^T g_{x'}) + \cdots + (-1)^p D^p(f_{x^p} - \lambda^T g_{x^p})\right] v dt
$$

Now, using the result of Valentine [8, Lemma 2][5, Lemma 7.3].

$$
\begin{aligned}
&(f_x - \lambda^T g_x) \\
&= D(f_{x'} - \lambda^T g_{x'}) - D^2(f_{x''} - \lambda^T g_{x''}) + \cdots + (-1)^{p-1} D^p(f_{x^p} - \lambda^T g_{x^p}), \quad t \in I
\end{aligned} \tag{17}
$$

The final equation we got is a linear differential equation of order p in $\lambda(.)$, therefore, it is solvable for piecewise continuous function $\lambda(.)$ and x. Using the result of valentine on equations 11 and 12, we will easily get other two conditions. This completes the proof of this theorem. □

3 Conclusions

Thus we have derived the necessary conditions for the variational problem 6 using the results of mathematical programming. We can also see that in the case of unconstrained optimization, we are left with the Euler-Lagrange equation [5].

Declarations

- The authors received no specific funding for this study.
- The authors declare that they have no conflicts of interest to report regarding the present study.
- No Human subject or animals are involved in the research.
- All authors have mutually consented to participate.
- All the authors have consented the Journal to publish this paper.
- Authors declare that all the data being used in the design and production cum layout of the manuscript is declared in the manuscript.

References

[1] Craven, B.D.: Mathematical Programming and Control Theory. Springer, New York (2012)

[2] Craven, B.D.: Control and Optimization vol. 16. CRC Press, Boca Raton, FL (1998)

[3] Mond, B., Hanson, M.A.: Duality for variational problems. Journal of Mathematical Analysis and Applications **18**(2), 355–364 (1967) https://doi.org/10.1016/0022-247X(67)90004-5

[4] Husain, Jabeen, Z.: On variational problems involving higher order derivatives. Journal of Applied Mathematics and Computing **17**(1-2), 433–455 (2005) https://doi.org/10.1007/bf02936067

[5] Burns, J.A.: Introduction to the Calculus of Variations and Control with Modern Applications. CRC Press, Boca Raton, FL (2013)

[6] Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge, UK (2004)

[7] Kuhn, H.W., Tucker, A.W.: Nonlinear programming. In: Kutateladze, S.S. (ed.) Traces and Emergence of Nonlinear Programming, pp. 247–258. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-08807-5_15

[8] University of Chicago. Department of Mathematics: Contributions to the Calculus of Variations. 1933–1937. University of Chicago Press, Chicago (1937). https://books.google.co.in/books?id=X90vtQEACAAJ