Mathematics and Statistics for Data Science Session 6 Optimization Techniques

Prof. Dr.-Ing. Paolo Mercorelli

November 29, 2020

Fundamental Optimization Techniques

- Lagrange Method
- Numerical Methods

An intuitive way to explain why Lagrange Method works and where the Lagrangian Function comes from

Lagrange Method

The method of Lagrange multipliers is the most used methods for solving optimization problems. In these notes I will try to explain it using an intuitive and common representation and didactical explanation.

Let us start with just a definition:
Gradient of a function

$$\nabla f(x_1, x_2, ..., x_n) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_n} \right)$$

is the vector which components are the partial derivatives.

Lagrange Method

Let us start with the enunciation of the Theorem

Let $f : A \subseteq \mathbb{R}^2 \to \mathbb{R}$ be a function defined in open interval $A \subseteq \mathbb{R}^2$ and let $g(x_1, x_2) = c$ be a constraint with "c" a given constant. Assuming that f and g are functions which admit continuous partial derivative on domain A.

Necessary and sufficient condition such that (x_{10}, x_{20}) is a relative max and min respectively to $g(x_{10}, x_{20}) - c = 0$ with "c" an constant and

1. $g(x_{10}, x_{20}) - c = 0$ and $\nabla g(x_{10}, y_{10}) \neq 0$
2. Let us define the following Lagrangian Function

$$L(x_1, x_2, \lambda) = f(x_1, x_2) - \lambda(g(x_1, x_2) - c)$$

is that there exists a real λ_0 such that:

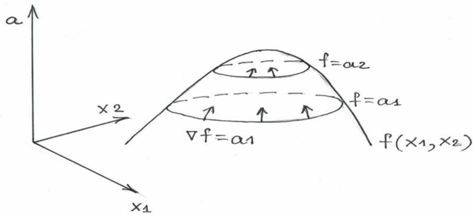
$$\nabla L = \nabla f(x_{10}, x_{20}) - \lambda_0 \nabla g(x_{10}, x_{20}) = 0$$

Geometric Property:

The most important geometric property about gradients is that they always point in the direction of a functions steepest slope at a given point and are orthogonal to the "level curves" and point toward the *concavity direction* of the surface.
In the picture we can find some "level curves" of considered function

With the terms "level curves" we intend the set of points, for instance, characterized by:

$$f(x_1, x_2) = a_1 \qquad f(x_1, x_2) = a_2$$

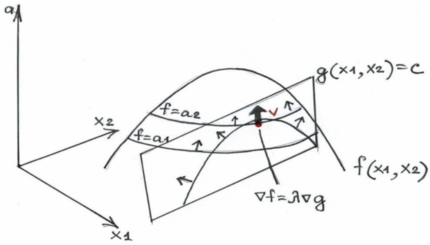


Remark

Basically the "level curves" are given by the intersections between planes or hyperplanes parallel to the horizontal plane.

Lagrange Method

.



In the drawing, the constraint is a plane that is cut through our hillside. If you imagine now to climb on the cut hillside as high as you can, you can reach just the point marked with letter v(vertical of the cut hillside).

Lagrange Method

.

If you consider the tangents to the level curves and the tangent to the cuthillside (set between function and constraint) just at point v the tangents are parallel. If the tangents are parallel, then the curves are also parallel.

We saw that, gradients are always perpendicular to level curves. So if these two curves are parallel, their gradients must also be parallel.

If you remember the definition of two parallel vectors which you can find here:

Two vectors v_1 and v_2 are parallel if

$$v_1 = \lambda v_2 \quad \textit{with } \lambda \textit{ constant.}$$

Conclusion:

At the maximun of the constraint intercected with the surface we have:

$$\nabla f(x_1, x_2) = \lambda \nabla g(x_1, x_2)$$

Following this idea we can write:

$$\nabla f(x_1, x_2) - \lambda \nabla g(x_1, x_2) = 0$$

This is the condition that must hold when we have reached the maximum of the function subject to the constraint.

$$g(x_1, x_2) = c$$

Lagrange Method

Now we can write a single equation that will capture this idea:

$$\nabla L = \nabla f(x_1, x_2) - \lambda \nabla g(x_1, x_2) = 0$$

In this sense, function L can be written after a partial integration in this way:

$$L(x_1, x_2, \lambda) = f(x_1, x_2) - \lambda(g(x_1, x_2) - c)$$

To understand this last equation better, we can consider its gradient explicitely.

Lagrange Method

Explicit gradient expression:

$$\nabla L = \begin{bmatrix} \frac{\partial f(x_1,x_2)}{\partial x_1} - \lambda \frac{\partial g(x_1,x_2)}{\partial x_1} \\ \\ \frac{\partial f(x_1,x_2)}{\partial x_2} - \lambda \frac{\partial g(x_1,x_2)}{\partial x_2} \\ \\ g(x_1, x_2) - c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

At this point, we have three equations in three unknowns to be solved. If the surface has an intersection with the constraint or constraints, the algebraic system admits a solution.

If the constraint is a compact one, then thanks to the Weierstrass Theorem we can directly check if the critical points are maxima or minima.

If no, the use of their bordered Hessian Matrix determines, whether those points are maxima, minima, or saddle points.

LEUPHANA
UNIVERSITY OF LÜNEBURG

Some final comments

Lagrange multipliers method is an algorithm that finds where the gradient of function points in the same direction as the gradients of its constraints. The intersection of function and constraints must be verified.

In terms of Lagrangian that optimization can be seen just a hill-climbing.

Maxima, Minima

.

Weierstrass Theorem:
A continuous function defined in a compact set has maxima and minima.

Remark

If the constraint is a compact set (closed and limited set), then the critical point can be checked directly by substitution into $f(x, y)$.

Maxima, Minima

.

Bordered Hessian Matrix is defined so:

$$H(L) = \begin{bmatrix} \frac{\partial L}{\partial \lambda^2} & \frac{\partial^2 L}{\partial \lambda \partial v} \\ \left(\frac{\partial^2 L}{\partial \lambda \partial v}\right)^T & \frac{\partial^2 L}{\partial v^2} \end{bmatrix} = \begin{bmatrix} 0 & -\frac{\partial g}{\partial v} \\ -\left(\frac{\partial g}{\partial v}\right)^T & \frac{\partial^2 L}{\partial v^2} \end{bmatrix}$$

where v is vector $v = (x, y)$.

Theorem:
Local minimum if $(-1)^m det(H(L(x_1^*, x_2^*, \lambda^*))) > 0$
Local maximum if $(-1)^m det(H(L(x_1^*, x_2^*, \lambda^*))) < 0$

m: number of constraints

LEUPHANA UNIVERSITY OF LÜNEBURG

Exam exercises:

$f(x, y) = x^2(t) + 4y^2(t) \qquad g(x, y) : x^2(t) + y^2(t) = 1$

$L(x_1, x_2, \lambda) = x^2(t) + 4y^2(t) - \lambda(x^2(t) + y^2(t) - 1)$

The equations to solve for Lagrange multipliers are:

$$2x - \lambda 2x = 0$$
$$8y - \lambda 2y = 0$$
$$x^2(t) + y^2(t) = 1$$

Solving these yields:

$i)\ x = 0$, $y = \pm 1$, $\lambda = 4$

$ii)\ x = \pm 1$, $y = 0$, $\lambda = 1$

The constraint is a compact set, thus the critical points can be changed check.

Solving these yields:

$$i)\ x = 0 \qquad\qquad ii)\ x = \pm 1$$
$$y = \pm 1 \qquad\qquad y = 0$$
$$\lambda = 4 \qquad\qquad \lambda = 1$$

The constraint is a compact set, thus the critical points can be checked directly:

$f(0, \pm 1) = +4 \rightarrow \textit{Maximum}$
$f(\pm 1, 0) = +1 \rightarrow \textit{Minimum}$

Maxima, Minima

The border HL matrix is:

$$H(L) = \begin{bmatrix} 0 & -\frac{\partial g}{\partial v} \\ -\left(\frac{\partial g}{\partial v}\right)^T & \frac{\partial^2 L}{\partial v^2} \end{bmatrix} = \begin{bmatrix} 0 & -2x & -2y \\ -2x & 2-2\lambda & 0 \\ -2y & 0 & 8-2\lambda \end{bmatrix}$$

At the rst pair of points:

$$HL(0, \pm 1, 4) = \begin{bmatrix} 0 & 0 & \pm 2 \\ 0 & -6 & 0 \\ \pm 2 & 0 & 0 \end{bmatrix}$$

$-det(HL(0, \pm 1, 4)) =$
$-1(-1)(-6)(\pm 2)^2 = -24 < 0$
This point is a local maximum

At the second pair of points:

$$HL(\pm 1, 0, 1) = \begin{bmatrix} 0 & \pm 2 & 0 \\ \pm 2 & 0 & 0 \\ 0 & 0 & 6 \end{bmatrix}$$

$-det(HL(\pm 1, 0, 1)) =$
$-1(-1)(6)(\pm 2)^2 = 24 > 0$ This point is a local minimum

Maxima, Minima

.

Recall: Hessian positive definite:
$det(HL(x_{10}, x_{20}, \lambda_0)) > 0$ all Eigenvalues positive

Recall: Hessian negative definite:
$det(HL(x_{10}, x_{20}, \lambda_0)) < 0$ all Eigenvalues negative

In case with at least one change of sign in the eigenvalues, then we have a saddle point!

Ex.1

$$f(x,y) = x^2(t) + y^2(t)$$

$$g(x,y) : -x(t) - y(t) = 1$$

Ex.2

$$f(x,y) = x^2(t) + y^2(t)$$

$$g(x,y) : x(t) = 1$$

Ex.3

$$f(x,y) = x(t)y(t)$$

$$g(x,y) : x^2(t) + 4y^2(t) - 1 = 0$$

Ex.4

$$f(x,y) = x(t) + y^2(t)$$

$$g(x,y) : x^2(t) + 4y^2(t) - 1 = 0$$

- Optimization by Numerical Methods
- Software

Motivation

If the optimization problem consists of high order of polynomial or complex nonlinear functions or in general for non convex optimization problems, the symbolic solutions are not always suitable.

Once the derivative or gradients are calculated, then we can calculate the solution of the equations using numerical methods.

- Newton method
- Newton-Raphson method

These methods are very important in case of non convex optimization problems in which an analytical solution could be difficult to be found.

Remark

These methods are iterative methods and the solution depends strongly on the initial value of the iteration.

Solving Nonlinear Equations

.

$$C = \int_0^{t_f} f(x)dt \qquad \frac{d}{dx}C = f(x) = 0$$

Newton Method

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \qquad f'(x_i) \neq 0$$

Interpretation: $f'(x_i) = \frac{f(x_i)}{x_i - x_{i+1}}$

In Matlab: fzero
i:index of the iteration, i=0,1,

Newton method (Example and software)

Assume that the following function should be optimized with respect to x:

$$J = \frac{1}{3}x^3 - \frac{1}{2}x^2 + cos(x + 0.15)$$

Once ist derivative is calculated we have:

$$x^2 - x - sin(x + 0, 15)$$

and the optimum is calculated imposing

$$x^2 - x - sin(x + 0.15) = 0$$

```
%x0 is the initial value, d is the required accuracy
close all
clear all
fun = @(x)x^2 - x - sin(x + 0.15);
derf = @(x)2 * x - 1 - cos(x + 0.15);
it = 0; x0 = 1.5;

d = feval(fun, x0)/feval(derf, x0);
while abs(d) > 1e - 3
x1 = x0 - d;
it = it + 1;
x0 = x1;
= feval(fun, x0)/feval(derf, x0);
end
res = x0;
```

Newton-Raphson method

.

The Newton-Raphson method is the extention to the Newton method to solve systems of equations in a numerical way.

This method is used in case we have a multivariable optimization problem which generates a system of nonlinear equations to be solved.

Basically is the same methods. Instead of the derivative we have a Jacobian associated to the nonlinear system

In Matlab: fsolve