Lagrange-multipliers

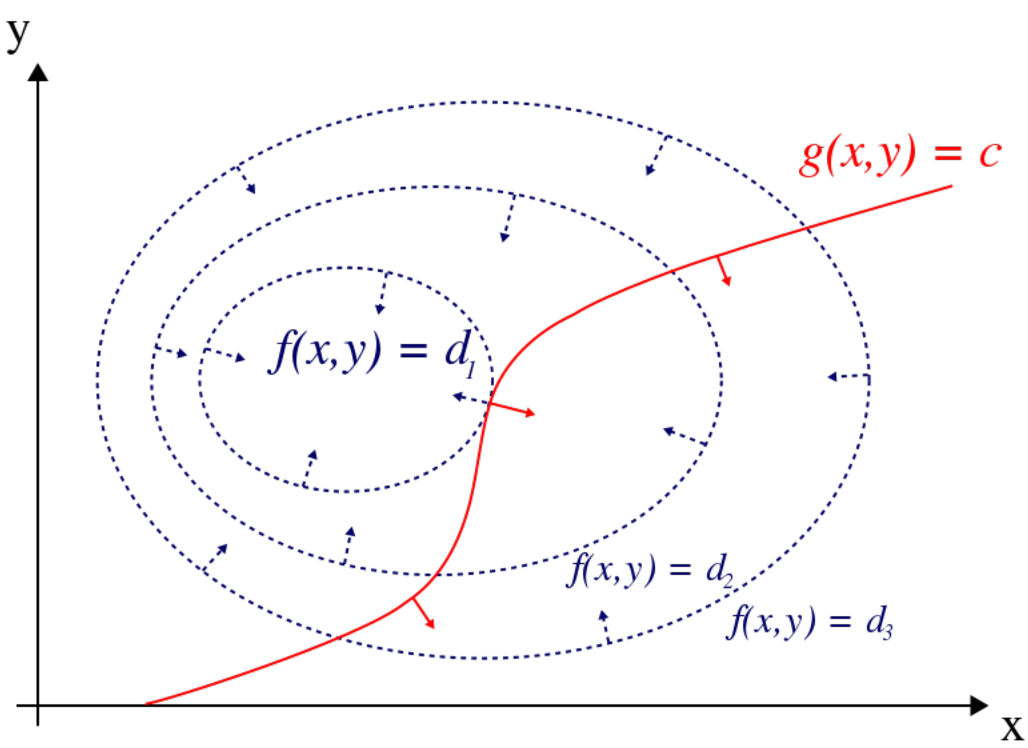
September 29, 2020

1 Introduction

The goal of the application of Lagrangian multiplier is to solve constrained minimization/maximization problems.

1.1 2D problem

In 2D, the task is to minimize the function $f(x, y)$ subject to $g(x, y) = c$. The problem can be visualized as follows:



(Image is from Wikipedia.)

In the image, different d_i values form different level sets. The curves $f(x, y) = d_i$ shows the locations where the value of the function f equals to d_i. Similarly, $g(x, y) = c$ is the curve, where the constraint is valid. When the curve g is tangent of a level set $f(x, y) = d_i$, the minimum/maximum or inflexion is reached. Formally, the normal of the curves can be obtained by calculation the gradient of the curves. They should be parallel:

$$\left[\begin{array}{c} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial x} \end{array} \right] = \lambda \left[\begin{array}{c} \frac{\partial g}{\partial x} \\ \frac{\partial g}{\partial x} \end{array} \right]$$

The idea of the famous Italian-French mathematician Joseph-Louis Lagrange is that the parallelity of the tangents and the constraint can be written in a cost function as

$$J = f(x, y) - \lambda(g(x, y) - c)$$

If one calculates the derivatives w.r.t. x and y, the paralellity is yielded:

$$\frac{\partial J}{\partial x} = \frac{\partial f}{\partial x} - \lambda \frac{\partial g}{\partial x} = 0 \rightarrow \frac{\partial f}{\partial x} = \lambda \frac{\partial g}{\partial x}$$

$$\frac{\partial J}{\partial y} = \frac{\partial f}{\partial y} - \lambda \frac{\partial g}{\partial y} = 0 \rightarrow \frac{\partial f}{\partial y} = \lambda \frac{\partial g}{\partial y}$$

The derivative w.r.t. λ gives back the constraint:

$$\frac{\partial J}{\partial \lambda} = g(x, y) - c = 0 \rightarrow g(x, y) = c$$

Therefore, a novel parameter is introducted and the cost function is modified. The new parameter λ is named Lagrange-multiplier in the literature.

1.2 Applying a Lagrange multiplier in arbitrary dimensions

The application of Lagrangien multipliers is not limited to the 2D case, it works in arbitrary dimensions. The cost function is then as follows:

$$J = f(\boldsymbol{x}) - \lambda(g(\boldsymbol{x}) - c)$$

where vector $\boldsymbol{x}$ consists of the parameters to be optimized.
The parallelity is valid if the following term is true:

$$\frac{\partial J}{\partial \boldsymbol{x}} = \nabla J - \lambda \nabla g = 0 \rightarrow \nabla J = \lambda \nabla g$$

Derivative w.r.t. multiplier λ gives the contraint similarly to the 2S case:

$$\frac{\partial J}{\partial \lambda} = g(\boldsymbol{x}) - c = 0 \rightarrow g(\boldsymbol{x}) = c$$

2 Application: optimal solution for a homogeneous linear system of equations.

The task is to solve the problem

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$$

For the overdetermined case, the norm of $\boldsymbol{A}\boldsymbol{x}$ has to be minimized. Thus the algebraic problem is to minimize the term:

$$\arg\min_{\boldsymbol{x}} \boldsymbol{x}^T \boldsymbol{A}^T \boldsymbol{A}\boldsymbol{x}$$

It is trivial that $\boldsymbol{x} = 0$ solves exactly the problem, however, one wants to find the non-trivial solution. Therefore, a constraint has to be introduced: let the length of vector $\boldsymbol{x}$be unit:

$$\boldsymbol{x}^T \boldsymbol{x} = 1$$

TIn this case, the minimization problem becomes

$$\arg\min \left\{ \boldsymbol{x}^{\boldsymbol{T}} \boldsymbol{A}^T \boldsymbol{A}\, \boldsymbol{x} - \lambda \left(\boldsymbol{x}^T \boldsymbol{x} - 1 \right) \right\}$$

The solution is obtained by derivating the cost function $J = \boldsymbol{x}^{\boldsymbol{T}} \boldsymbol{A}^T \boldsymbol{A}\, \boldsymbol{x} - \lambda \left(\boldsymbol{x}^T \boldsymbol{x} - 1 \right)$

$$\frac{\partial J}{\partial \boldsymbol{x}} = 2\boldsymbol{A}^T \boldsymbol{A}\boldsymbol{x} - \boldsymbol{2}\lambda \boldsymbol{x} = 0$$

This yields

$$\boldsymbol{A}^T \boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$$

This is a standard eigenvalue problem, the minima/maxima occur when $\boldsymbol{x}$ is one of the eigenvector of matrix $\boldsymbol{A}^T \boldsymbol{A}$.

The eigenvectors are substituted back to the original cost function:

$$\boldsymbol{x}^T \boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}^T\ \boldsymbol{x} = \lambda$$

as $\boldsymbol{x}^T \boldsymbol{x} = 1$. Thereofere, the cost equals to the eigenvalue. Remark that the symmetrix matrix $\boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{A}$ always has non-negative real eigenvalues.

The optimal value for the minimization is the eigenvector of $\boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{A}$ corresponding to the smallest eigenvalue. If the problem is to find the maximum of the least-squares norm of $\boldsymbol{A}\boldsymbol{x}$, then the optimum is the eigenvector of $\boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{A}$ corresponding to the largest eigenvalue.

3 Application in computer vision: plane fitting

A plane is given in implicit form as

$$ax + by + cz + d = 0$$

Thic can be written in homogeneous form as

$$\begin{bmatrix} x & y & z & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = 0$$

If points in the plane are given, each one forms an equation as $ax_i + by_i + cz_i + d = 0$. N points yields the following linear system of equations:

$$\begin{bmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_N & y_N & z_N & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = 0$$

The optimal solution is the eigenvector corresponding to the smallest eigenvalue of matrix

$$\begin{bmatrix} x_1 & x_2 & \dots & x_N \\ y_1 & y_2 & \dots & y_N \\ z_1 & z_2 & \dots & z_N \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_N & y_N & z_N & 1 \end{bmatrix} = \begin{bmatrix} \sum x_i^2 & \sum x_i y_i & \sum x_i z_i & \sum x_i \\ \sum x_i y_i & \sum y_i^2 & \sum y_i z & \sum y_i \\ \sum x_i z_i & \sum y_i z_i & \sum z_i^2 & \sum z_i \\ \sum x_i & \sum y_i & \sum z_i & N \end{bmatrix}$$

4 Application in geometry: optimal line/plane fitting

For optimal line/plane fitting, the distance between a line and a point should be determined. The line is given by its implicit form

$$ax + by + c = 0$$

If plane-point distance has to be calculated, the plane can also be given by an implicit form as

$$ax + by + cz + d = 0$$

In general, an N-dimensional (hyper)plane is given as

$$\boldsymbol{l}^T \boldsymbol{x} + m = 0$$

were $\boldsymbol{l^T} = [a, b]$ for lines, $\boldsymbol{l}^T = [a, b, c]$ for planes, ets.

4.1 Distance of a point and a line/plane

For computing the distance of point $\boldsymbol{x}_0$, the problem can be written by applying a Lagrangian multiplier λ:

$$J = (\boldsymbol{x} - \boldsymbol{x}_0)^T (\boldsymbol{x} - \boldsymbol{x}_0) + \lambda \left(\boldsymbol{l}^T \boldsymbol{x} + m \right)$$

The solution is given by derivating the problem:

$$\frac{\partial J}{\partial \boldsymbol{x}} = 2 (\boldsymbol{x} - \boldsymbol{x}_0) + \lambda \boldsymbol{l} = 0$$

Therefore,

$$\boldsymbol{x} = \frac{2\boldsymbol{x}_0 - \lambda \boldsymbol{l}}{2}$$

Multiplier λ is given from the constraint $\boldsymbol{l}^T \boldsymbol{x} + m = 0$:

$$\boldsymbol{l}^T \frac{2\boldsymbol{x}_0 - \lambda \boldsymbol{l}}{2} + m = 0$$

$$\boldsymbol{l}^T \boldsymbol{x}_0 + m = \frac{\lambda \boldsymbol{l}^T \boldsymbol{l}}{2}$$

$$\lambda = 2 \frac{\boldsymbol{l}^T \boldsymbol{x}_0 + m}{\boldsymbol{l}^T \boldsymbol{l}}$$

The distance vector itself is

$$\boldsymbol{x} - \boldsymbol{x}_0 = \frac{2\boldsymbol{x}_0 - 2\frac{\boldsymbol{l}^T \boldsymbol{x}_0 + m}{\boldsymbol{l}^T \boldsymbol{l}} \boldsymbol{l}}{2} - \boldsymbol{x}_0 =$$

$$\boldsymbol{x} - \boldsymbol{x}_0 = -\frac{\boldsymbol{l}^T \boldsymbol{x} + m}{\boldsymbol{l}^T \boldsymbol{l}} \boldsymbol{l}$$

Its length is the distance value:

$$d = \sqrt{(\boldsymbol{x} - \boldsymbol{x}_0)^T (\boldsymbol{x} - \boldsymbol{x}_0)} = \frac{\boldsymbol{l}^T \boldsymbol{x} + m}{\boldsymbol{l}^T \boldsymbol{l}} \sqrt{\boldsymbol{l}^T \boldsymbol{l}} =$$

$$\frac{\boldsymbol{l}^T \boldsymbol{x} + m}{\sqrt{\boldsymbol{l}^T \boldsymbol{l}}}$$

If $\boldsymbol{l}^T \boldsymbol{l} = 1$, then $d = \boldsymbol{l}^T \boldsymbol{x}_0 + m$, in other words, the implicit equation gives the distance itself.

Remark that this distance calculation is valid for arbitrary dimension.

4.2 Optimal line/plane fitting

If there are N data points for which the line/plane has to be fitted, the problem of finding the most dominant direction of the can be written in as

4.2.1 Translation

$$\arg\min J = \arg\min \sum_{i=1}^{N} d_i^2 = \arg\min \sum_{i=1}^{N} \left(\boldsymbol{l}^T \left(\boldsymbol{x}_i + \boldsymbol{t} \right) \right)^2$$

$$\frac{\partial J}{\partial \boldsymbol{t}} = 2 \sum_{i=1}^{N} \boldsymbol{l}^T \left(\boldsymbol{x}_i + \boldsymbol{t} \right) \boldsymbol{l} = \boldsymbol{l^T} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \left[\boldsymbol{x_i} + \boldsymbol{t} \right] \boldsymbol{l} = \boldsymbol{0}$$

$$\sum_{i=1}^{N} \left[\boldsymbol{x_i} + \boldsymbol{t} \right] = 0$$

$$\boldsymbol{t} = -\frac{\sum \boldsymbol{x}_i}{N}$$

Or, in other words, the optimal translation is the center of the gravity of the points.

4.3 Rotation

For the sake of simplicity, let the origin be the center of gravity. Then only the direction of the line should be determined:

$$\arg\min \sum_{i=1}^{N} d_i^2 = \arg\min \sum_{i=1}^{N} \left(\boldsymbol{l}^T \boldsymbol{x}_i \right)$$

subject to $\boldsymbol{l}^T \boldsymbol{l}$=1.

This is equivalent to a simple homogeneous linear problem:

$$\boldsymbol{X} \boldsymbol{l} = 0$$

where

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1^T \\ \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_N^T \end{bmatrix}$$

The solution is the eigenvektor of matrix $\boldsymbol{X}^T \boldsymbol{X}$ corresponding to the smallest eigenvalue. The obtained vector $\boldsymbol{l}$ is the normal of the line/plane.

5 Singular Value Decomposition

Given a matrix $\boldsymbol{A}$(size:$N \times M$), its singular value decompostion always exists, and can be written in the form

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\boldsymbol{T}}$$

where matrix $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal, thus

$$\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}_{N\times N},$$

$$\boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}_{M\times M}$$

and $\boldsymbol{S}$ is a diagonal matrix

$$\boldsymbol{S} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sigma_M \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

The columns of matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are the eigenvectors of $\boldsymbol{A}\boldsymbol{A}^T$ and $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}$, respectively, and the square roots of the eigenvalues of the matrices are stacked in σ_i. Remark that the eigenvalues of $\boldsymbol{A}\boldsymbol{A}^T$ and $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}$ are the same, and they are listed in descending order. All eigenvalues are non-negative real numbers.