

## Copyright Notice

Staff and students of the University of Cambridge are reminded that copyright subsists in this extract and the work from which it was taken. This Digital Copy has been made under the terms of a CLA Licence which allows you to:

- access and download a copy;
- print out a copy.

This Digital Copy and any digital or printed copy supplied to or made by you under the terms of this Licence are for use in connection with this Course of Study. You may retain such copies after the end of the course, but strictly for your own personal use.

All copies (including electronic copies) shall include this Copyright Notice and shall be destroyed and/or deleted if and when required by the University of Cambridge.

Except as provided for by copyright law, no further copying, storage or distribution (including by e-mail) is permitted without the consent of the copyright holder.

The author (which term includes artists and other visual creators) has moral rights in the work and neither staff nor students may cause, or permit, the distortion, mutilation or other modification of the work, or any derogatory treatment of it, which would be prejudicial to the honour or reputation of the author.

Course of Study: 3F8 Inference

Title: Pattern Recognition and Machine Learning

Name of Author: Christopher Bishop

Name of Publisher: Springer

Pages used: 707-710

## Appendix E. Lagrange Multipliers

*Lagrange multipliers*, also sometimes called *undetermined multipliers*, are used to find the stationary points of a function of several variables subject to one or more constraints.

Consider the problem of finding the maximum of a function  $f(x_1, x_2)$  subject to a constraint relating  $x_1$  and  $x_2$ , which we write in the form

$$g(x_1, x_2) = 0. \quad (\text{E.1})$$

One approach would be to solve the constraint equation (E.1) and thus express  $x_2$  as a function of  $x_1$  in the form  $x_2 = h(x_1)$ . This can then be substituted into  $f(x_1, x_2)$  to give a function of  $x_1$  alone of the form  $f(x_1, h(x_1))$ . The maximum with respect to  $x_1$  could then be found by differentiation in the usual way, to give the stationary value  $x_1^*$ , with the corresponding value of  $x_2$  given by  $x_2^* = h(x_1^*)$ .

One problem with this approach is that it may be difficult to find an analytic solution of the constraint equation that allows  $x_2$  to be expressed as an explicit function of  $x_1$ . Also, this approach treats  $x_1$  and  $x_2$  differently and so spoils the natural symmetry between these variables.

A more elegant, and often simpler, approach is based on the introduction of a parameter  $\lambda$  called a Lagrange multiplier. We shall motivate this technique from a geometrical perspective. Consider a  $D$ -dimensional variable  $\mathbf{x}$  with components  $x_1, \dots, x_D$ . The constraint equation  $g(\mathbf{x}) = 0$  then represents a  $(D-1)$ -dimensional surface in  $\mathbf{x}$ -space as indicated in Figure E.1.

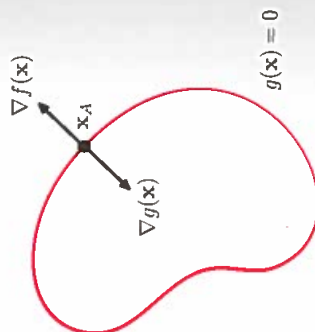
We first note that at any point on the constraint surface the gradient  $\nabla g(\mathbf{x})$  of the constraint function will be orthogonal to the surface. To see this, consider a point  $\mathbf{x}$  that lies on the constraint surface, and consider a nearby point  $\mathbf{x} + \epsilon$  that also lies on the surface. If we make a Taylor expansion around  $\mathbf{x}$ , we have

$$g(\mathbf{x} + \epsilon) \simeq g(\mathbf{x}) + \epsilon^T \nabla g(\mathbf{x}). \quad (\text{E.2})$$

Because both  $\mathbf{x}$  and  $\mathbf{x} + \epsilon$  lie on the constraint surface, we have  $g(\mathbf{x}) = g(\mathbf{x} + \epsilon)$  and hence  $\epsilon^T \nabla g(\mathbf{x}) \simeq 0$ . In the limit  $\|\epsilon\| \rightarrow 0$  we have  $\epsilon^T \nabla g(\mathbf{x}) = 0$ , and because  $\epsilon$  is

Figure E.1

A geometrical picture of the technique of Lagrange multipliers in which we seek to maximize a function  $f(\mathbf{x})$ , subject to the constraint  $g(\mathbf{x}) = 0$ . If  $\mathbf{x}$  is  $D$  dimensional, the constraint  $g(\mathbf{x}) = 0$  corresponds to a subspace of dimensionality  $D - 1$ , indicated by the red curve. The problem can be solved by optimizing the Lagrangian function  $L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda g(\mathbf{x})$ .



then parallel to the constraint surface  $g(\mathbf{x}) = 0$ , we see that the vector  $\nabla g$  is normal to the surface.

Next we seek a point  $\mathbf{x}^*$  on the constraint surface such that  $f(\mathbf{x})$  is maximized. Such a point must have the property that the vector  $\nabla f(\mathbf{x})$  is also orthogonal to the constraint surface, as illustrated in Figure E.1, because otherwise we could increase the value of  $f(\mathbf{x})$  by moving a short distance along the constraint surface. Thus  $\nabla f$  and  $\nabla g$  are parallel (or anti-parallel) vectors, and so there must exist a parameter  $\lambda$  such that

$$\nabla f + \lambda \nabla g = 0 \quad (\text{E.3})$$

where  $\lambda \neq 0$  is known as a *Lagrange multiplier*. Note that  $\lambda$  can have either sign.

At this point, it is convenient to introduce the *Lagrangian* function defined by

$$L(\mathbf{x}, \lambda) \equiv f(\mathbf{x}) + \lambda g(\mathbf{x}). \quad (\text{E.4})$$

The constrained stationarity condition (E.3) is obtained by setting  $\nabla_{\mathbf{x}} L = 0$ . Furthermore, the condition  $\partial L / \partial \lambda = 0$  leads to the constraint equation  $g(\mathbf{x}) = 0$ .

Thus to find the maximum of a function  $f(\mathbf{x})$  subject to the constraint  $g(\mathbf{x}) = 0$ , we define the Lagrangian function given by (E.4) and we then find the stationary point of  $L(\mathbf{x}, \lambda)$  with respect to both  $\mathbf{x}$  and  $\lambda$ . For a  $D$ -dimensional vector  $\mathbf{x}$ , this gives  $D + 1$  equations that determine both the stationary point  $\mathbf{x}^*$  and the value of  $\lambda$ . If we are only interested in  $\mathbf{x}^*$ , then we can eliminate  $\lambda$  from the stationarity equations without needing to find its value (hence the term 'undetermined multiplier').

As a simple example, suppose we wish to find the stationary point of the function  $f(x_1, x_2) = 1 - x_1^2 - x_2^2$  subject to the constraint  $g(x_1, x_2) = x_1 + x_2 - 1 = 0$ , as illustrated in Figure E.2. The corresponding Lagrangian function is given by

$$L(\mathbf{x}, \lambda) = 1 - x_1^2 - x_2^2 + \lambda(x_1 + x_2 - 1). \quad (\text{E.5})$$

The conditions for this Lagrangian to be stationary with respect to  $x_1$ ,  $x_2$ , and  $\lambda$  give the following coupled equations:

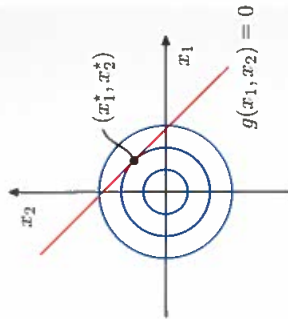
$$-2x_1 + \lambda = 0 \quad (\text{E.6})$$

$$-2x_2 + \lambda = 0 \quad (\text{E.7})$$

$$x_1 + x_2 - 1 = 0. \quad (\text{E.8})$$

Figure E.2

A simple example of the use of Lagrange multipliers in which the aim is to maximize  $f(x_1, x_2) = 1 - x_1^2 - x_2^2$  subject to the constraint  $g(x_1, x_2) = 0$  where  $g(x_1, x_2) = x_1 + x_2 - 1$ . The circles show contours of the function  $f(x_1, x_2)$ , and the diagonal line shows the constraint surface  $g(x_1, x_2) = 0$ .



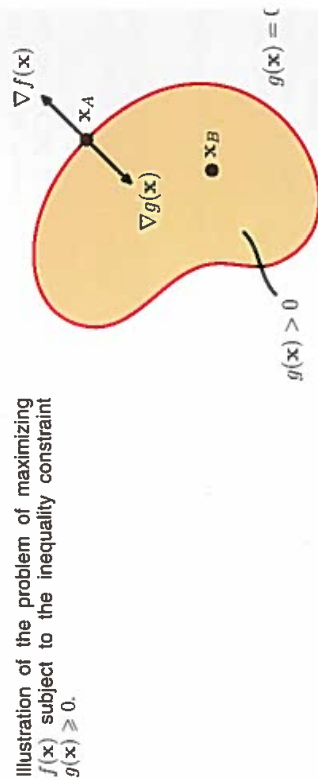
Solution of these equations then gives the stationary point as  $(x_1^*, x_2^*) = (\frac{1}{2}, \frac{1}{2})$ , and the corresponding value for the Lagrange multiplier is  $\lambda = 1$ .

So far, we have considered the problem of maximizing a function subject to an equality constraint of the form  $g(\mathbf{x}) = 0$ . We now consider the problem of maximizing  $f(\mathbf{x})$  subject to an inequality constraint of the form  $g(\mathbf{x}) \geq 0$ , as illustrated in Figure E.3.

There are now two kinds of solution possible, according to whether the constrained stationary point lies in the region where  $g(\mathbf{x}) > 0$ , in which case the constraint is *inactive*, or whether it lies on the boundary  $g(\mathbf{x}) = 0$ , in which case the constraint is said to be *active*. In the former case, the function  $g(\mathbf{x})$  plays no role and so the stationary condition is simply  $\nabla f(\mathbf{x}) = 0$ . This again corresponds to a stationary point of the Lagrange function (E.4) but this time with  $\lambda = 0$ . The latter case, where the solution lies on the boundary, is analogous to the equality constraint discussed previously and corresponds to a stationary point of the Lagrange function (E.4) with  $\lambda \neq 0$ . Now, however, the sign of the Lagrange multiplier is crucial, because the function  $f(\mathbf{x})$  will only be at a maximum if its gradient is oriented away from the region  $g(\mathbf{x}) > 0$ , as illustrated in Figure E.3. We therefore have  $\nabla f(\mathbf{x}) = -\lambda \nabla g(\mathbf{x})$  for some value of  $\lambda > 0$ .

For either of these two cases, the product  $\lambda g(\mathbf{x}) = 0$ . Thus the solution to the

Figure E.3



problem of maximizing  $f(\mathbf{x})$  subject to  $g(\mathbf{x}) \geq 0$  is obtained by optimizing the Lagrange function (E.4) with respect to  $\mathbf{x}$  and  $\lambda$  subject to the conditions

$$g(\mathbf{x}) \geq 0 \quad (\text{E.9})$$

$$\lambda \geq 0 \quad (\text{E.10})$$

$$\lambda g(\mathbf{x}) = 0 \quad (\text{E.11})$$

These are known as the *Karush-Kuhn-Tucker* (KKT) conditions (Karush, 1939; Kuhn and Tucker, 1951).

Note that if we wish to minimize (rather than maximize) the function  $f(\mathbf{x})$  subject to an inequality constraint  $g(\mathbf{x}) \geq 0$ , then we minimize the Lagrangian function  $L(\mathbf{x}, \lambda) = f(\mathbf{x}) - \lambda g(\mathbf{x})$  with respect to  $\mathbf{x}$ , again subject to  $\lambda \geq 0$ .

Finally, it is straightforward to extend the technique of Lagrange multipliers to the case of multiple equality and inequality constraints. Suppose we wish to maximize  $f(\mathbf{x})$  subject to  $g_j(\mathbf{x}) = 0$  for  $j = 1, \dots, J$ , and  $h_k(\mathbf{x}) \geq 0$  for  $k = 1, \dots, K$ . We then introduce Lagrange multipliers  $\{\lambda_j\}$  and  $\{\mu_k\}$ , and then optimize the Lagrangian function given by

$$L(\mathbf{x}, \{\lambda_j\}, \{\mu_k\}) = f(\mathbf{x}) + \sum_{j=1}^J \lambda_j g_j(\mathbf{x}) + \sum_{k=1}^K \mu_k h_k(\mathbf{x}) \quad (\text{E.12})$$

subject to  $\mu_k \geq 0$  and  $\mu_k h_k(\mathbf{x}) = 0$  for  $k = 1, \dots, K$ . Extensions to constrained functional derivatives are similarly straightforward. For a more detailed discussion of the technique of Lagrange multipliers, see Nocedal and Wright (1999).