Lagrange Multipliers

(Com S 477/577 Notes)

Yan-Bin Jia

Nov 3, 2020

1 Introduction

We turn now to the study of minimization with constraints. More specifically, we will tackle the following problem:

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & h_1(\boldsymbol{x}) = 0 \\ & \vdots \\ & h_m(\boldsymbol{x}) = 0 \end{aligned}$$

where $\boldsymbol{x} \in \Omega \subset \mathbb{R}^n$, and the functions f, $h_1, \dots, h_m$ are continuous, and usually assumed to be in C^2 (i.e., with continuous second partial derivatives). When f and h_j's are linear, the problem is a linear programming one solvable using the simplex algorithm [3]. Hence we would like to focus on the case that these functions are nonlinear.

In order to gain some intuition, let us look at the case where $n = 2$ and $m = 1$. The problem becomes

$$\begin{aligned} \text{minimize} \quad & f(x, y) \\ \text{subject to} \quad & h(x, y) = 0, \qquad x, y \in \mathbb{R}. \end{aligned}$$

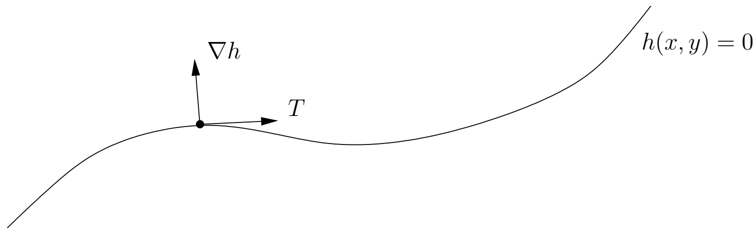
The constraint $h(x, y) = 0$ defines a curve. Differentiate the equation with respect to x:

$$\frac{\partial h}{\partial x} + \frac{\partial h}{\partial y}\frac{dy}{dx} = 0.$$

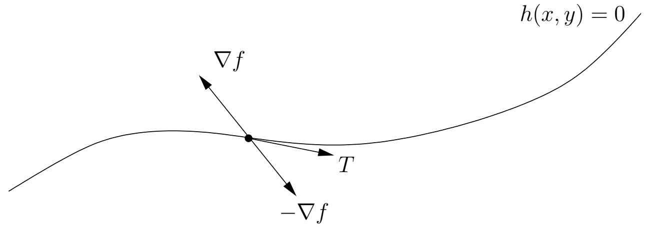
The tangent of the curve is $T(x, y) = \left(1, \frac{dy}{dx}\right)^T$, written as a column vector. And the gradient of the curve is $\nabla h = \left(\frac{\partial h}{\partial x}, \frac{\partial h}{\partial y}\right)$. So the above equation states that

$$\nabla h \, T = 0;$$

namely, the tangent of the curve must be normal to the gradient all the time. Suppose we are at a point on the curve. To stay on the curve, any motion must be along the tangent T.



In order to increase or decrease $f(x, y)$, motion along the constraint curve must have a component along the gradient of f, that is, $\nabla f\, T \neq 0$.

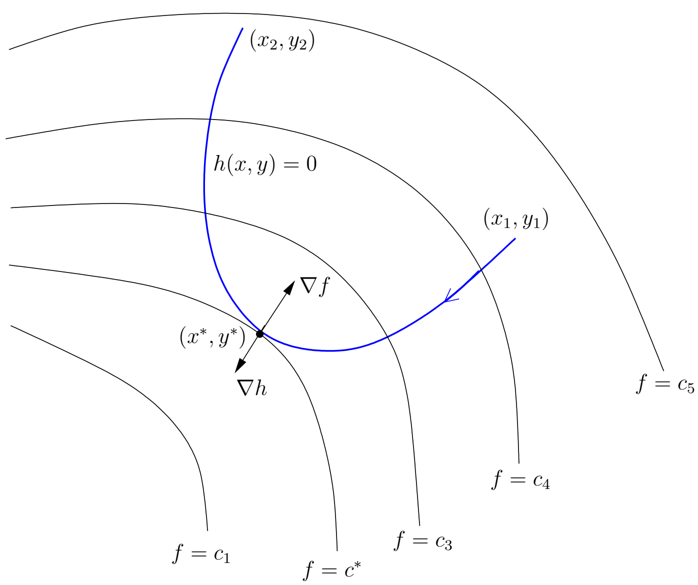


At an extremum of f, a differential motion should not yield a component of motion along ∇f. Thus T is orthogonal to ∇f; in other words, the condition

$$\nabla f\, T = 0$$

must hold. Now T is orthogonal to both gradients ∇f and ∇h at an extrema. This means that ∇f and ∇h must be parallel. Phrased differently, there exists some $\lambda \in \mathbb{R}$ such that

$$\nabla f + \lambda \nabla h = \mathbf{0}. \tag{1}$$



The figure above explains condition (1) by superposing the curve $h(x, y) = 0$ onto the family of *level curves* of $f(x, y)$, that is, the collection of curves $f(x, y) = c$, where c is any real number in the range of f. In the figure, $c_5 > c_4 > c_3 > c^* > c_1$. The tangent of a level curve is always orthogonal to the gradient ∇f. Otherwise moving along the curve would result in an increase or decrease of the value of f. Imagine a point moving on the curve $h(x,y) = 0$ from (x_1, y_1) to (x_2, y_2). Initially, the motion has a component along the negative gradient direction $-\nabla f$, resulting in the decrease of the value of f. This component becomes smaller and smaller. When the moving point reaches (x^*, y^*), the motion is orthogonal to the gradient. From that point on, the motion starts having a component along the gradient direction ∇f so the value of f increases. Thus at (x^*, y^*), f achieves a local minimum, which is c^*. The motion is in the tangential direction of the curve $h(x,y) = 0$, which is orthogonal to the gradient ∇h. Therefore, at the point (x^*, y^*) the two gradients ∇f and ∇h must be collinear. This is what equation (1) says. It is clear that the two curves $f(x,y) = c^*$ and $h(x,y) = 0$ are tangent at (x^*, y^*).

Suppose we find the set S of points satisfying the following equations:

$$\begin{aligned} h(x,y) &= 0, \\ (\nabla f)^T + \lambda (\nabla h)^T &= \mathbf{0}, \qquad \text{for some } \lambda. \end{aligned}$$

Then S contains the extremal points of f subject to the constraints $h(x,y) = 0$. The above two equations constitute a nonlinear system in the variables x, y, λ. It can be solved using numerical techniques, for example, Newton's method.

2 Lagrangian

It is convenient to introduce the *Lagrangian* associated with the constrained problem, defined as

$$\ell(x,y,\lambda) = f(x,y) + \lambda h(x,y).$$

Note that

$$\nabla \ell = \begin{pmatrix} \frac{\partial f}{\partial x} + \lambda \frac{\partial h}{\partial x} \\ \frac{\partial f}{\partial y} + \lambda \frac{\partial h}{\partial y} \\ h \end{pmatrix}^T = \left(\nabla f + \lambda \nabla h, h \right).$$

Thus, setting $\nabla \ell = 0$ yields the same system of nonlinear equations we derived earlier.

The value λ is known as the *Lagrange multiplier*. The approach of constructing the Lagrangians and setting its gradient to zero is known as the method of Lagrange multipliers. Here we are not minimizing the Lagrangian, but merely finding its stationary point (x, y, λ).

EXAMPLE 1 Find the extremal values of the function $f(x,y) = xy$ subject to the constraint

$$h(x,y) = \frac{x^2}{8} + \frac{y^2}{2} - 1 = 0.$$

We first construct the Lagrangian and find its gradient:

$$\ell(x,y,\lambda) = xy + \lambda \left(\frac{x^2}{8} + \frac{y^2}{2} - 1 \right),$$

$$\nabla \ell(x, y\lambda) = \begin{pmatrix} y + \dfrac{\lambda x}{4} \\ x + \lambda y \\ \dfrac{x^2}{8} + \dfrac{y^2}{2} - 1 \end{pmatrix}^T = \mathbf{0}.$$

The above leads to three equations

$$\begin{aligned} y + \frac{\lambda x}{4} &= 0, && (2) \\ x + \lambda y &= 0, && (3) \\ x^2 + 4y^2 &= 8. && (4) \end{aligned}$$

Eliminate x from (2) and (3):

$$y - \frac{\lambda^2}{4} y = 0,$$

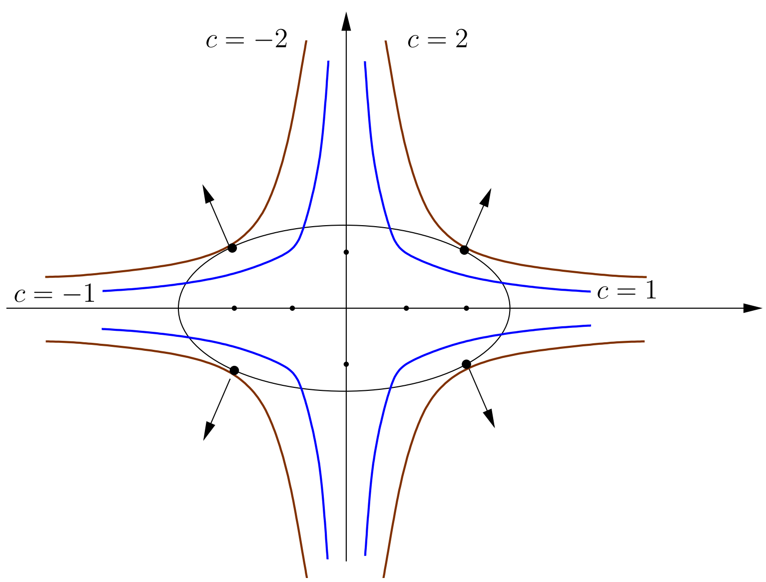
which, since $y \neq 0$ (otherwise $x = 0$ would hold by (2)), leads to

$$\lambda^2 = 4 \qquad \text{and} \qquad \lambda = \pm 2.$$

Now $x = \pm 2y$ by (3). Substituting this equation into (4) gives us

$$y = \pm 1 \qquad \text{and} \qquad x = \pm 2.$$

So there are four extremal points of f subject to the constraint h: $(2, 1), (-2, -1), (2, -1)$, and $(-2, -1)$. The maximum value 2 is achieved at the first two points while the minimum value -2 is achieved at the last two points.



Graphically, the constraint h defines an ellipse. The level contours of f are the hyperbolas $xy = c$, with $|c|$ increasing as the curves move out from the origin.

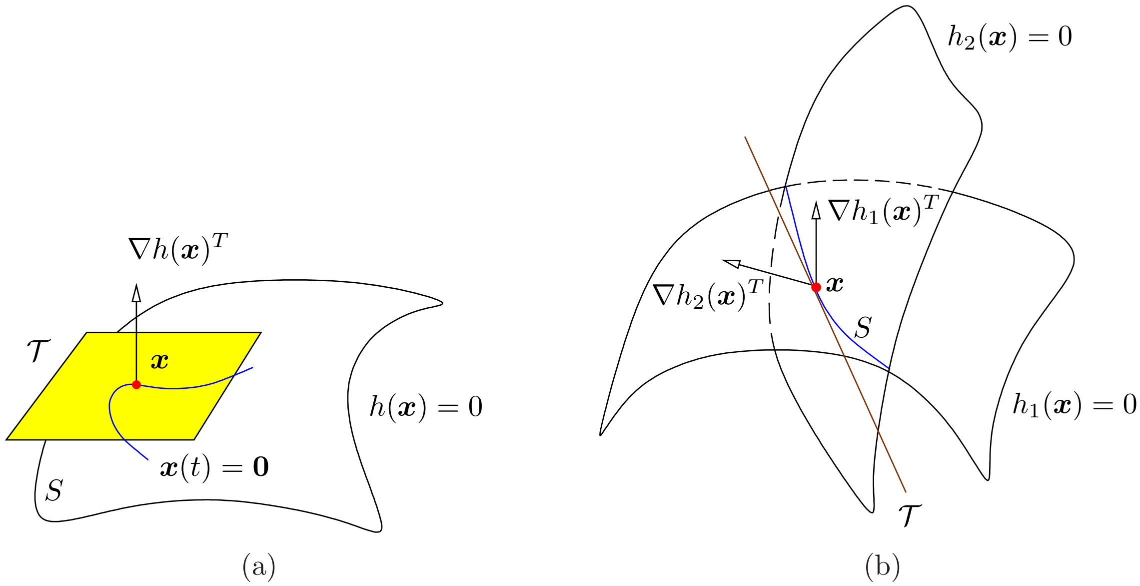


Figure 1: Tangent space at $\boldsymbol{x}$ on a surface defined by (a) one constraint $h(\boldsymbol{x}) = 0$ (shown as a tangent plane) and (b) two constraints $h_1(\boldsymbol{x}) = 0$ and $h_2(\boldsymbol{x}) = 0$ (shown as a tangent line).

3 General Formulation

Now we would like to generalize to the case with multiple constraints. Let $\boldsymbol{h} = (h_1, \ldots, h_m)^T$ be a function from $\mathbb{R}^n$ to $\mathbb{R}^m$. Consider the constrained optimization problem below.

$$\begin{array}{rl} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}. \end{array} \tag{5}$$

Each function h_i defines a surface $S_i = \{\, \boldsymbol{x} \mid h_i(\boldsymbol{x}) = 0 \,\}$ of generally $n-1$ dimensions in the space $\mathbb{R}^n$. This surface is smooth provided $h_i(\boldsymbol{x}) \in C^1$. The m constraints $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}$ together defines a surface S, which is the intersection of the surfaces $S_1, \ldots, S_m$; namely,

$$\{\, \boldsymbol{x} \mid \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0} \,\} = \{\, \boldsymbol{x} \mid h_1(\boldsymbol{x}) = 0 \,\} \cap \cdots \cap \{\, \boldsymbol{x} \mid h_m(\boldsymbol{x}) = 0 \,\}.$$

The surface S is of generally $n - m$ dimensions. In Figure 1(a), S is defined by one constraint; in (b), the surface is defined by two constraints. The constrained optimization (5) is carried out over the surface S.

Consider a point $\boldsymbol{p}$ on the surface S, and all the curves on $\mathcal{S}$ passing through $\boldsymbol{p}$. The *tangent space* $\mathcal{T}$ at $\boldsymbol{p}$ consists of the derivatives of these curves at $\boldsymbol{p}$. Let $\boldsymbol{x}(t)$ be one such curve, and let $\boldsymbol{x} = \boldsymbol{p}$ by a slight abuse of notation. Since $\boldsymbol{x}(t)$ is also on S_i, every tangent vector to S at $\boldsymbol{x}$ is automatically a tangent vector to S_i. This implies that the tangent space $\mathcal{T}$ to S at $\boldsymbol{x}$ is a subspace of the tangent space $\mathcal{T}_i$ to S_i at the same point, for $1 \le i \le m$. Meanwhile, the gradient ∇h_i is orthogonal to $\mathcal{T}_i$ because

$$h_i'(\boldsymbol{x}) = \nabla h_i(\boldsymbol{x}) \cdot \boldsymbol{x}'(t) = 0.$$

Therefore, ∇h_i must be orthogonal to the subspace $\mathcal{T}$ of $\mathcal{T}_i$; namely, $\nabla h_i \in \mathcal{T}^\perp$. The point $\boldsymbol{x}$ is a *regular point* of the surface S if the gradient vectors $\nabla h_1(\boldsymbol{x})$, …, $\nabla h_m(\boldsymbol{x})$ are linearly independent.

Theorem 1 *At a regular point* $\boldsymbol{x}$ *of the surface* S *defined by* $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}$*, the tangent space is the same as*

$$\{\, \boldsymbol{y} \mid \nabla \boldsymbol{h}(\boldsymbol{x})\, \boldsymbol{y} = \mathbf{0} \,\},$$

where

$$\nabla \boldsymbol{h} = \begin{pmatrix} \nabla h_1 \\ \vdots \\ \nabla h_m \end{pmatrix}.$$

The rows of the matrix $\nabla \boldsymbol{h}(\boldsymbol{x})$ are the gradient vectors $\nabla h_j(\boldsymbol{x})$, $j = 1, \ldots, m$. The theorem says that the tangent space $\mathcal{T}$ at $\boldsymbol{x}$ is equal to the null space of $\nabla \boldsymbol{h}(\boldsymbol{x})$. Thus, its orthogonal complement $\mathcal{T}^{\perp}$ must equal the row space of $\nabla \boldsymbol{h}(\boldsymbol{x})$, which are spanned by the vectors $\nabla h_j(\boldsymbol{x})$, $1 \le j \le m$

Example 2. Suppose $h(x_1, x_2) = x_1$. Then $h(\boldsymbol{x}) = 0$ yields the x_2 axis. And $\nabla h = (1, 0)$ at all the points on the axis. So every $\boldsymbol{x} \in \mathbb{R}^2$ is regular. The tangent space is also the x_2 axis and has dimension 1.

If instead $h(x_1, x_2) = x_1^2$ then $h(\boldsymbol{x}) = 0$ still defines the x_2 axis. On this axis, $\nabla h = (2x_1, 0) = (0, 0)$. Thus, no point is regular. The dimension of T, which is the x_2 axis, is still one, but the dimension of the space $\{\, \boldsymbol{y} \mid \nabla h \cdot \boldsymbol{y} = 0 \,\}$ is two.

Lemma 2 *Let* $\boldsymbol{x}^*$ *be a local extremum of* f *subject to the constraints* $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}$*. Then for all* $\boldsymbol{y}$ *in the tangent space of the constraint surface at* $\boldsymbol{x}^*$*,*

$$\nabla f(\boldsymbol{x}^*) \boldsymbol{y} = 0.$$

The above lemma implies that $\nabla f(\boldsymbol{x}^*)$ lies in $\mathcal{T}^{\perp}$. We would like to claim that, at the extremal point $\boldsymbol{x}^*$, ∇f is a linear combination of ∇h_i', $1 \le i \le m$. This is only valid if these gradients span $\mathcal{T}^{\perp}$, which is true when $\boldsymbol{x}^*$ is regular. The next theorem states that the Lagrange multiplier method is a necessary condition for the existence of an extremum point.

Theorem 3 (First-Order Necessary Conditions) *Let* $\boldsymbol{x}^*$ *be a local extremum point of* f *subject to the constraints* $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}$*. Assume further that* $\boldsymbol{x}^*$ *is a regular point of these constraints. Then there is a* $\boldsymbol{\lambda} \in \mathbb{R}^m$ *such that*

$$\nabla f(\boldsymbol{x}^*) + \boldsymbol{\lambda}^T \nabla \boldsymbol{h}(\boldsymbol{x}^*) = \mathbf{0}.$$

The first order necessary conditions together with the constraints

$$\boldsymbol{h}(\boldsymbol{x}^*) = \mathbf{0}$$

give a total of $n + m$ equations in $n + m$ variables $\boldsymbol{x}^*$ and $\boldsymbol{\lambda}$. Thus a unique solution can be determined at least locally.

Example 3. We seek to construct a cardboard box of maximum volume, given a fixed area of cardboard. Denoting the dimensions of the box by x, y, z, the problem can be expressed as

$$\begin{array}{rl} \text{maximize} & xyz \\ \text{subject to} & xy + yz + xz = \frac{c}{2}, \end{array}$$

where $c > 0$ is the given area of cardboard. Consider the Lagrangian $xyz + \lambda(xy + yz + xz - \frac{c}{2})$. The first-order necessary conditions are easily found to be

$$yz + \lambda(y + z) = 0, \tag{6}$$
$$xz + \lambda(x + z) = 0, \tag{7}$$
$$xy + \lambda(x + y) = 0, \tag{8}$$

together with the original constraint. Before solving the above equations, we note that their sum is

$$(xy + yz + xz) + 2\lambda(x + y + z) = 0,$$

which, under the original constraint, becomes

$$c/2 + 2\lambda(x + y + z) = 0.$$

Hence, it is clear that $\lambda \neq 0$. Neither of x, y, z can be zero since if either is zero, all must be so according to (6)–(8).

To solve the equations (6)–(8), multiply (6) by x and (7) by y, and then subtract the two to obtain

$$\lambda(x - y)z = 0.$$

Operate similarly on the second and the third to obtain

$$\lambda(y - z)x = 0.$$

Since no variables can be zero, it follows that

$$x = y = z = \sqrt{\frac{c}{6}}$$

is the unique solution to the necessary conditions. The box must be a cube.

We can derive second-order conditions for constrained problems, assuming f and $\boldsymbol{h}$ are twice continuously differentiable.

Theorem 4 (Second-Order Necessary Conditions) *Suppose that $\boldsymbol{x}^*$ is a local minimum of f subject to $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$ and that $\boldsymbol{x}^*$ is a regular point of these constraints. Denote by F the Hessian of f, and by H_i the Hessian of h_i, for $1 \leq i \leq m$. Then there is a $\boldsymbol{\lambda} \in \mathbb{R}^m$ such that*

$$\nabla f(\boldsymbol{x}^*) + \boldsymbol{\lambda}^T \nabla \boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}.$$

The matrix

$$L(\boldsymbol{x}^*) = F(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i H_i(\boldsymbol{x}^*) \tag{9}$$

is positive semidefinite on the tangent space $\{\, \boldsymbol{y} \mid \nabla \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{y} = \boldsymbol{0} \,\}$.

Theorem 5 (Second-Order Sufficient Conditions) *Suppose there is a point $\boldsymbol{x}^*$ satisfying $\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}$, and a $\boldsymbol{\lambda}$ such that*

$$\nabla f(\boldsymbol{x}^*) + \boldsymbol{\lambda}^T \nabla \boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}.$$

Suppose also that the matrix $L(\boldsymbol{x}^*)$ defined in (9) is positive definite on the tangent space $\{\, \boldsymbol{y} \mid \nabla \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{y} = \boldsymbol{0} \,\}$. Then $\boldsymbol{x}^*$ is a strict local minimum of f subject to $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$.

EXAMPLE 4. consider the problem

$$\begin{array}{rl} \text{minimize} & x_1x_2 + x_2x_3 + x_1x_3 \\ \text{subject to} & x_1 + x_2 + x_3 = 3 \end{array}$$

The first order necessary conditions become

$$\begin{aligned} x_2 + x_3 + \lambda &= 0, \\ x_1 + x_3 + \lambda &= 0, \\ x_1 + x_2 + \lambda &= 0. \end{aligned}$$

We can solve these three equations together with the one constraint equation and obtain

$$x_1 = x_2 = x_3 = 1 \qquad \text{and} \qquad \lambda = -2.$$

Thus $x^* = (1, 1, 1)^T$.

Now we need to resort to the second-order sufficient conditions to determine if the problem achieves a local maximum or minimum at $x_1 = x_2 = x_3 = 1$. We find the matrix

$$\begin{aligned} L(\boldsymbol{x}^*) &= F(\boldsymbol{x}^*) + \lambda H(\boldsymbol{x}^*) \\ &= \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} + \lambda \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \end{aligned}$$

is neither positive nor negative definite. On the tangent space $M = \{\, \boldsymbol{y} \mid y_1 + y_2 + y_3 = 0 \,\}$, however, we note that

$$\begin{aligned} \boldsymbol{y}^T L \boldsymbol{y} &= y_1(y_2 + y_3) + y_2(y_1 + y_3) + y_3(y_1 + y_2) \\ &= -(y_1^2 + y_2^2 + y_3^2) \\ &< 0, \qquad\qquad \text{for all } \boldsymbol{y} \neq \mathbf{0}.. \end{aligned}$$

Thus L is negative definite on M and the solution 3 we found is at least a local maximum.

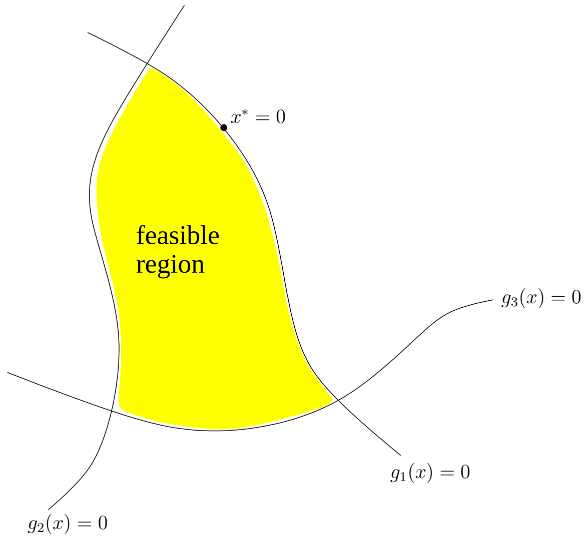
4 Inequality Constraints

Finally, we address a problem of the general form

$$\begin{array}{rl} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0} \\ & \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0} \end{array}$$

where $\boldsymbol{h} = (h_1, \ldots, h_m)^T$ and $\boldsymbol{g} = (g_1, \ldots, g_p)^T$.

A fundamental concept that provides a great deal of insight as well as simplifies the required theoretical development is that of an *active constraint*. An inequality constraint $g_i(\boldsymbol{x}) \leq 0$ is said to be *active* at a feasible point $\boldsymbol{x}$ if $g_i(\boldsymbol{x}) = 0$ and *inactive* at $\boldsymbol{x}$ if $g_i(\boldsymbol{x}) < 0$. By convention we refer to any equality constraint $h_i(\boldsymbol{x}) = 0$ as active at any feasible point. The constraints active at a feasible point $\boldsymbol{x}$ restrict the domain of feasibility in neighborhoods of $\boldsymbol{x}$. Therefore, in studying



the properties of a local minimum point, it is clear that attention can be restricted to the active constraints. This is illustrated in the figure below where local properties satisfied by the solution $\boldsymbol{x}^*$ obviously do not depend on the inactive constraints g_2 and g_3.

Assume that the functions f, $\boldsymbol{h} = (h_1, \ldots, h_m)^T$, $\boldsymbol{g} = (g_1, \ldots, g_p)^T$ are twice continuously differentiable. Let $\boldsymbol{x}^*$ be a point satisfying the constraints

$$\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0} \quad \text{and} \quad \boldsymbol{g}(\boldsymbol{x}^*) \leq \boldsymbol{0},$$

and let $J = \{\, j \mid g_j(\boldsymbol{x}^*) = 0 \,\}$. Then $\boldsymbol{x}^*$ is said to be a *regular point* of the above constraints if the gradient vectors $\nabla h_i(\boldsymbol{x}^*)$, $\nabla g_j(\boldsymbol{x}^*)$, $1 \leq i \leq m$, $j \in J$ are linearly independent. Now suppose this regular point $\boldsymbol{x}^*$ is also a relative minimum point for the original problem (1). Then it can be shown that there exists a vector $\boldsymbol{\lambda} \in \mathbb{R}^m$ and a vector $\boldsymbol{\mu} \in \mathbb{R}^p$ with $\boldsymbol{\mu} \geq \boldsymbol{0}$ such that

$$\begin{aligned} \nabla f(\boldsymbol{x}^*) + \boldsymbol{\lambda}^T \nabla \boldsymbol{h}(\boldsymbol{x}^*) + \boldsymbol{\mu}^T \nabla \boldsymbol{g}(\boldsymbol{x}^*) &= \boldsymbol{0}; \\ \boldsymbol{\mu}^T \boldsymbol{g}(\boldsymbol{x}^*) &= 0. \end{aligned}$$

Since $\boldsymbol{\mu} \geq \boldsymbol{0}$ and $\boldsymbol{g}(\boldsymbol{x}^*) \leq \boldsymbol{0}$, the second constraint above is equivalent to the statement that $\mu_j \neq 0$, for each $1 \leq j \leq p$, only if g_j is active.

To find a solution, we enumerate various combinations of *active* constraints, that is, constraints where equalities are attained at $\boldsymbol{x}^*$, and check the signs of the resulting Lagrange multipliers.

There are a number of distinct theories concerning this problem, based on various regularity conditions or constraint qualifications, which are directed toward obtaining definitive general statements of necessary and sufficient conditions. One can by no means pretend that all such results can be obtained as minor extensions of the theory for problems having equality constraints only. To date, however, their use has been limited to small-scale programming problems of two or three variables. We refer you to [3] for more on the subject.

References

[1] T. Abell. Lecture notes. The Robotics Institute, School of Computer Science, Carnegie Mellon University, 1992.

[2] M. Avriel. *Nonlinear Programming: Analysis and Methods*. Prentice-Hall, 1976; Dover Publications, 2003.

[3] D. G. Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley, 2nd edition, 1984.