Chapter 2

Optimality Conditions

2.1 Global and Local Minima for Unconstrained Problems

When a minimization problem does not have any constraints, the problem is to find the minimum of the objective function. We distinguish between two kinds of minima. The *global* minimum is the minimum of the function over the entire domain of interest, while a *local* minimum is the minimum over a smaller subdomain. The design point where the objective function f reaches a minimum is called the *minimizer* and is denoted by an asterisk as $\mathbf{x}^*$. The global minimizer is the point x^* which satisfies

$$f(\mathbf{x}^*) \leq f(\mathbf{x})\,. \tag{2.1.1}$$

A point x^* is a local minimizer if you for some $r > 0$

$$f(\mathbf{x}^*) \leq f(\mathbf{x})\,. \qquad \text{if } ||x - x^*|| < r\,. \tag{2.1.2}$$

That is, x^* is a local minimizer, if you can find a sphere around it in which it is the minimizer.

2.2 Taylor series Expansion

To find conditions for a point to be a minimizer, we use the Taylor series expansion around a presumed minimizer x^*

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \sum_{i=1}^{n}(x_i - x_i^*)\frac{\partial f}{\partial x_i}(\mathbf{x}^*) + \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{n}(x_i - x_i^*)(x_k - x_k^*)\frac{\partial^2 f}{\partial x_i \partial x_k}(\mathbf{x}^*) + \text{higher order terms} \tag{2.2.1}$$

When $\mathbf{x}$ is very close to $\mathbf{x}^*$, we can neglect even the second derivative terms. Furthermore, if $\mathbf{x}$ and $\mathbf{x}^*$ are identical except for the jth component, then Eq. 2.2.1 becomes

$$f(x) \approx f(\mathbf{x}^*) + (x_j - x_j^*)\frac{\partial f}{\partial x_j}(\mathbf{x}^*)\,. \tag{2.2.2}$$

Then, for Eq. 2.1.2 to hold for both positive and negative values of $(x_j - x_j^*)$, we must have the familiar first order condition

$$\frac{\partial f}{\partial x_j}(\mathbf{x}^*) = 0\,, \qquad , j = 1, \ldots, n \tag{2.2.3}$$

Equation 2.2.3 is not applicable only to a minimum; the same condition is derived in the same manner for a maximum. It is called a *stationarity* condition, and a point satisfying it is called a stationary point.

To obtain conditions specific to a minimizer, we need to use the second derivative terms in the Taylor expansion. However, to facilitate working with these terms, we would like to rewrite this expansion in matrix notation, which is more compact. We define the gradient vector to be the vector whose components are the first derivatives, and denote it as ∇f. That is,

$$\nabla f = \begin{Bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \cdot \\ \cdot \\ \cdot \frac{\partial f}{\partial x_n} \end{Bmatrix} \tag{2.2.4}$$

Similarly, we define the *Hessian* (after the German mathematician) Otto Ludwig Hesse (1811-1874)) matrix H, to be the symmetric matrix of second derivatives. That is, h_{ij}, the element of the matrix H at the ith row and jth column is $\frac{\partial^2}{\partial x_i x_j}$,

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 x_n} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 x_n} \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \frac{\partial^2 f}{\partial x_n x_1} & \frac{\partial^2 f}{\partial x_n x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}. \tag{2.2.5}$$

Finally, we define $\Delta \mathbf{x} = \mathbf{x} - \mathbf{x}^*$, and then we can write the Taylor expansion, Eq. 2.2.1 as

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \Delta \mathbf{x}^T \nabla f(\mathbf{x}^*) + \frac{1}{2} \Delta \mathbf{x}^T H(\mathbf{x}^*) \Delta \mathbf{x} + \text{higher order terms} \tag{2.2.6}$$

For a minimum, the gradient vector is zero, so that if

$$\Delta \mathbf{x}^T H(\mathbf{x}^*) \Delta \mathbf{x} > 0 \tag{2.2.7}$$

then for small enough $\Delta \mathbf{x}$, the higher order terms can be neglected, and we will be assured that $\mathbf{x}^*$ is at least a local minimizer of f. A symmetric matrix which satisfies such positivity condition for any vector $\Delta \mathbf{x}$ is called *positive definite*. It can be shown that a matrix is positive definite if and only if all of its eigenvalues are positive. In one dimension, the condition of positive definiteness reduces to the condition that the second derivative is positive, indicating a positive curvature or a *convex* function. In two dimensions, a positive definite Hessian means that the curvature is positive in any direction, or again that f is convex.

The condition that H is positive definite is a *sufficient* condition for a local minimizer, but it is not a *necessary* condition. It can be shown that the necessary condition is a slightly milder version of Eq. 2.2.7

$$\Delta \mathbf{x}^T H(\mathbf{x}^*) \Delta \mathbf{x} \geq 0\,. \tag{2.2.8}$$

With this condition, the second-order terms in the Taylor series can be zero, and then the higher-order terms determine whether the point is a minimizer or not. In one dimension, the condition is that the second derivative, or the curvature, is non-negative, and a zero second derivative implies that the minimum is decided on the basis of higher derivatives. For example, $f = x^4$ has zero first and second derivatives at $x = 0$, and this point is a minimum. However, $f = -x^4$ satisfies the same conditions, but has a maximum at that point. Finally, $f = x^3$ satisfies this condition, and it has an inflection point. A matrix H which satisfies Eq. 2.2.8 for every vector $\Delta\mathbf{x}$ is called *positive semi-definite*, and all of its eigenvalues are non-negative (but some may be zero).

Since minimizing f is equivalent to maximizing $-f$, it follows that at a stationary point if $-H$ is positive definite, we have sufficient conditions for a local maximizer. Such a matrix is called *negative definite* and all of its eigenvalues are negative. Similarly, if $-H$ is positive semi-definite, we call H negative semi-definite, and the matrix has eigenvalues which are all non-positive.

In one dimension the second derivative can be positive, negative, or zero. corresponding to a minimum, maximum, or uncertainty about the nature of the stationary point. In higher dimensions we have the 4 possibilities that we have already listed, plus the possibility that the some of the eigenvalues of the matrix are positive and some are negative. A matrix with both positive and negative eigenvalues is called *indefinite*. In two dimensions, a function with an indefinite matrix will have one positive eigenvalue and one negative eigenvalue. These two eigenvalues will correspond to the extreme values of the curvature of the function in different directions. The two directions will be perpendicular to each other, and with one direction having positive curvature and the other having negative curvature, the function will locally look like a saddle. Therefore, a stationary point with an indefinite Hessian is called a *saddle point*.

To summarize, at a stationary point we need to inspect the Hessian. If it is positive definite we have a minimum. If it is positive semi-definite we may have a minimum, an inflection point or a saddle point. If the Hessian is indefinite we have a saddle point. If it is negative definite we must have a maximum, while if it is negative semi-definite we may have a maximum or an inflection point or a saddle point.

Example 2.2.1

Find the stationary point of the quadratic function $f = x_1^2 + x_2^2 + rx_1x_2 - 2x_1$. Determine the nature of the stationary point for all possible values of r.

The stationary point is found by setting the first derivatives to zero

$$\frac{\partial f}{\partial x_1} = 2x_1 + rx_2 - 2 = 0 \tag{2.2.9}$$

$$\frac{\partial f}{\partial x_2} = rx_1 + 2x_2 = 0 \tag{2.2.10}$$

For any value of r beside $r = 2$, the solution is

$$x_1 = \frac{2}{4 - r^2}, \qquad x_2 = \frac{-2r}{4 - r^2}. \tag{2.2.11}$$

The Hessian is

$$H = \begin{bmatrix} 2 & r \\ r & 2 \end{bmatrix}. \tag{2.2.12}$$

The eigenvalues μ of H are found be setting the determinant of $H - \mu I$, where I is the identity matrix

$$|H - \mu I| = \begin{vmatrix} 2-\mu & r \\ r & 2-\mu \end{vmatrix} = (2-\mu)^2 - r^2 = 0. \tag{2.2.13}$$

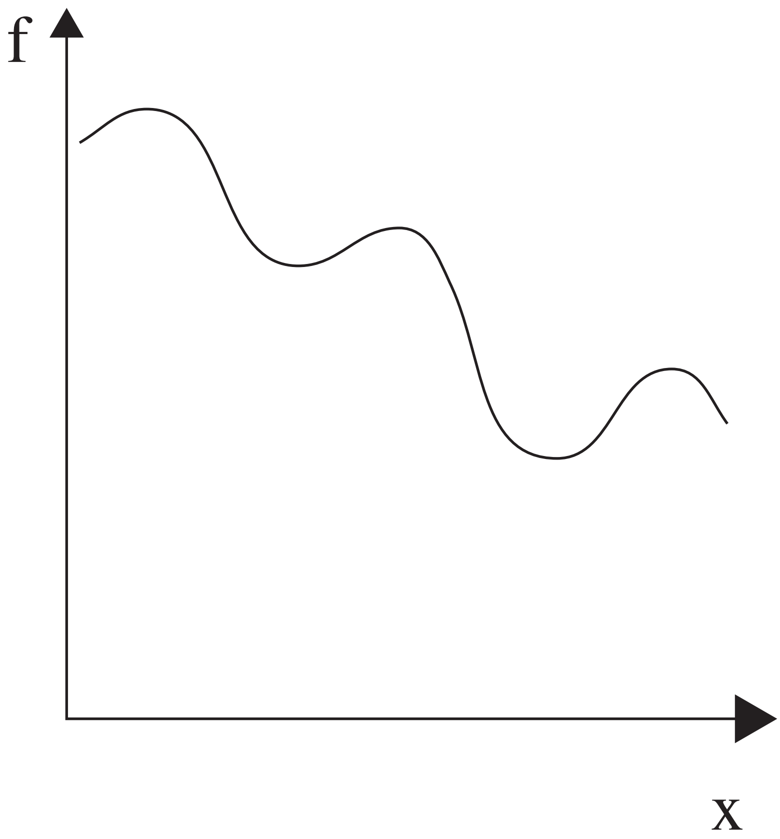

Figure 2.1: Function with several local minima

So that the two eigenvalues are $\mu_{1,2} = 2 \pm r$. For $r < 2$ we have two positive eigenvalues so that H is positive definite and the stationary point is a minimum. For $r > 2$, we have one positive and one negative eigenvalue, so that H is indefinite and the stationary point is a saddle point. For $r = 2$ one of the eigenvalues is zero so that the matrix is positive semidefinite. However, we do not have a stationary point for $r = 2$. ●●●

2.3 Condition for global minimum

Up to now we discussed the condition for a local minimizer. A function can have several local minima, as shown in the Fig. 2.1, and have a positive definite Hessian at each one of the minima. Therefore, when we find a stationary point where the Hessian is positive definite, we usually cannot tell whether the point is a local minimizer or a global one. However, if the function is convex everywhere it can be shown that any local minimizer is also a global minimizer. If f is twice differentiable everywhere, then convexity just means that the Hessian is positive semidefinite everywhere. However, it is useful to define convex functions even if they do not posses second derivatives. For example, the function $f = |x|$ is convex looking and has a global minimum at $x = 0$. Similarly, we can replace a smooth convex function by short straight segments, see Fig. 2.2, we would still have a convex looking function. Therefore, the definition of convex function is generalized to include these cases, and it can be shown that the proper definition is the following: A function $f(\mathbf{x})$ is defined to be convex if for every two points $\mathbf{x}_1$ and $\mathbf{x}_2$ and every scalar $0 < \alpha < 1$

$$f[\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2] \leq \alpha f(\mathbf{x}_1) + (1-\alpha)f(\mathbf{x}_2)\,. \tag{2.3.1}$$

Equation 2.3.1 simply states that if we connect any two points on the surface of the function by a straight segment, the segment will lie above the function. If we can have strict inequality in Eq. /refeq:convex, then the function f is *strictly convex*.

A local minimum of a convex function can be shown to be also the global minimum, or the only minimum of the function. If the function is strictly convex, the minimizer is a single point, while a more general convex function can have the minimum be attained in entire region of design space.

It is usually not convenient to check on the convexity of a function from the definition of a convex function. If the function is C^2 (twice continuously differentiable), then convexity is assured if the Hessian of the function is positive semi-definite everywhere, and strict convexity is assured if the function is positive definite everywhere. One implication of this result is that the minimum of a quadratic function is the global minimum, because if a quadratic function is positive definite at one point, it is positive definite everywhere (the Hessian is a constant matrix).

2.4 Necessary conditions for constrained local minimizer

We will consider first the case when we have only a single equality constraint

$$\begin{aligned} &\textbf{minimize} && f(\mathbf{x})\,, \\ &\textbf{such that} && h(\mathbf{x}) = 0\,, \end{aligned} \tag{2.4.1}$$

$$\tag{2.4.2}$$

In this case, if we want to check whether a candidate point $\mathbf{x}^*$ is a minimizer, we can no longer compare it with any neighboring point. Instead we have to limit ourselves only to points that satisfy the equality constraint. That is, instead of Eq. 2.1.2 the condition should be

$$\begin{aligned} f(\mathbf{x}^*) \leq f(\mathbf{x})\,. \qquad &\text{if } \;||x - x^*|| < r\,, \\ &\text{and} \qquad h(x) = 0\,, \end{aligned} \tag{2.4.3}$$

$$\tag{2.4.4}$$

We can obtain a first order optimality condition by considering a point $\mathbf{x}$ which is infinitesimally close to $\mathbf{x}^*$, that is $\mathbf{x} = \mathbf{x}^* + d\mathbf{x}$. Then

$$df = f(\mathbf{x}) - f(\mathbf{x}^*) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i\,. \tag{2.4.5}$$

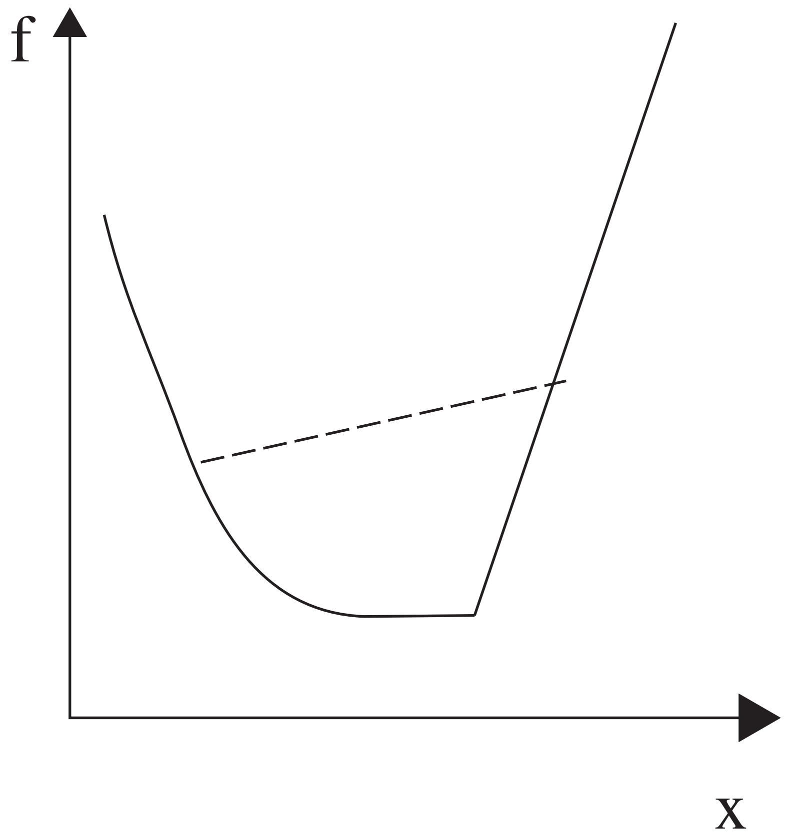

Figure 2.2: Convex function

Similarly, with $h(\mathbf{x}) - h(\mathbf{x}^*) = dh = 0$ we have

$$dh = \sum_{i=1}^{n} \frac{\partial h}{\partial x_i} dx_i = 0\,. \tag{2.4.6}$$

This equation indicates that we cannot choose the components of $d\mathbf{x}$ independently. Rather, we can express one of them, say the jth one, in terms of the other $n-1$, provided that $\partial h/\partial x_j \neq 0$. That is

$$dx_j = -\frac{1}{\partial h/\partial x_j} \sum_{i \neq j} \frac{\partial h}{\partial x_i} dx_i\,. \tag{2.4.7}$$

We will therefore assume that we can choose all the components of $d\mathbf{x}$ arbitrarily except for the jth one.

We now calculate $df + \lambda dh$ where λ is any number. Because $dh = 0$ we have

$$df = df + \lambda dh = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} + \lambda \frac{\partial h}{\partial x_i} \right) dx_i \tag{2.4.8}$$

We now choose λ to eliminate the contribution of the dependent component dx_j, that is

$$\frac{\partial f}{\partial x_j} + \lambda \frac{\partial h}{\partial x_j} = 0\,. \tag{2.4.9}$$

Then in Eq. 2.4.8 if we choose the $dx_k \neq 0$ and $dx_i = 0$ for $i \neq k$ and $i \neq j$ then

$$df = \left(\frac{\partial f}{\partial x_k} + \lambda \frac{\partial h}{\partial x_k} \right) dx_k\,. \tag{2.4.10}$$

Then, because dx_k can be either positive or negative, to insure that $df \geq 0$ we must have

$$\frac{\partial f}{\partial x_k} + \lambda \frac{\partial h}{\partial x_k} = 0 \tag{2.4.11}$$

This applies to every $k \neq j$, but for j we have Eq. 2.4.9, which is also the same. These equations are usually made to look the same as the conditions for an unconstrained minimization by defining a new function L called the *Lagrangian*

$$L = f + \lambda h\,, \tag{2.4.12}$$

and call λ a Lagrange multiplier. Then the condition for optimality is that we can find a Lagrange multiplier that satisfies the following equation

$$\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} + \lambda \frac{\partial h}{\partial x_i} = 0\,, \qquad , i = 1, \ldots, n\,. \tag{2.4.13}$$

As in the case of unconstrained minimization, Eq. 2.4.13 is only a stationarity condition, which applies equally well to minima and maxima. Equations 2.4.13 together with the condition $h = 0$ can be viewed as $n+1$ equations for the components of the minimizer $\mathbf{x}^*$ and the Lagrange multiplier λ. However, we usually calculate the minimizer by some graphical or numerical method, and then use Eq. 2.4.13 to check whether we are at the minimum. When we have only a single design variable, we can always find a Lagrange multiplier that will satisfy Eq. 2.4.13, so that it does not seem that the optimality condition is much of a condition. However, indeed, with a single design variable and a single constraint, we are not left with any freedom to optimize. If we have a point that satisfies the constraint, in most cases we will not be able to move the single design variable from this point without violating the constraint. When we have multiple design variables, then Eq. 2.4.13 will give us a different equation for each design variable. At the optimum point all of these equations should give us the same value of λ.

Another useful function of the Lagrangian is to estimate the effect of changing the data in a problem on the optimum design. That is, consider a situation where the objective function f and the constraint h are a function of a parameter p, then the minimizer $\mathbf{x}^*$ and the minimum value of the objective function f^* will depend on p,

$$f = f(\mathbf{x}, p)\,, \qquad h = h(\mathbf{x}, p)\,, \qquad \mathbf{x}^* = \mathbf{x}^*(p)\,, \qquad f^*(p) = f(\mathbf{x}^*, p)\,. \tag{2.4.14}$$

It can be shown that the derivative of f^* with respect to p is equal to the partial derivative of the Lagrangian with respect to p,

$$\frac{df^*}{dp} = \frac{\partial L}{\partial p}(\mathbf{x}^*, p)\,. \tag{2.4.15}$$

Example 2.4.1

For the soft-drink can design, Example1.2.1, assume that l, the design variable which decides the aspect ratio constraint on the can, is an outside parameter rather than a design variable. Further assume that the aspect ratio condition is imposed as an equality constraint, with $H = 2D$ for $l = 0$ and $H = 2.2D$ for $l = 1$. Check that for $l = 0$, the maximum profit per ounce, p_o is achieved for $D = 2.04$ in. Calculate the derivative of p_o with respect to l, and use it to estimate the effect on the profit of changing l from 0 to 1. Check your answer.

The equality constraint dictates that $H = 4.08$ in. and $V_o = 7.409$ ounces. Because this design point is not at the limits of either the dimensions or the volume, we can reformulate the design problem as

$$\begin{aligned} &\textbf{minimize} && -p_o = \frac{C - P}{V_o} \\ &\textbf{such that} && h = 1 - \frac{H}{(2 + 0.2l)D} = 0\,. \end{aligned} \tag{2.4.16}$$

For convenience the equations for the cost, C, price, P, and volume in ounces V_o are copied from Example 1.2.1

$$C = 0.8V_o + 0.1S\,, \qquad P = 2.5V_o - 0.02V_o^2 + 5l\,. \tag{2.4.17}$$

where

$$V_o = 0.25\pi D^2 H/1.8\,, \qquad S = 2(0.25\pi D^2) + \pi DH\,, \tag{2.4.18}$$

For the candidate point, these equations give us $p_o = 1.11065$ cents per ounce. In order to check the optimality condition, Eq. 2.4.13, it may be tempting to substitute the expressions for the cost, price, volume and area into the objective function. However, differentiating the complicated expression that will result from the substitution is difficult. Instead, it is more convenient to differentiate the individual expressions as shown below. The Lagrangian function is given as

$$L = \frac{C - P}{V_o} + \lambda\left[1 - \frac{H}{(2 + 0.2l)D}\right]\,, \tag{2.4.19}$$

and the optimality condition, Eq. 2.4.13 is obtained by differentiating the Lagrangian with respect to the two design variables

$$\frac{\partial L}{\partial D} = \frac{1}{V_o}\left(\frac{\partial C}{\partial D} - \frac{\partial P}{\partial D}\right) - \frac{C-P}{V_o^2}\frac{\partial V_o}{\partial D} + \lambda\frac{H}{(2+0.2l)D^2}\,,$$
$$\frac{\partial L}{\partial H} = \frac{1}{V_o}\left(\frac{\partial C}{\partial H} - \frac{\partial P}{\partial H}\right) - \frac{C-P}{V_o^2}\frac{\partial V_o}{\partial H} - \frac{\lambda}{(2+0.2l)D}\,. \quad (2.4.20)$$

To evaluate the derivatives appearing in the optimality conditions we differentiate the equations for the cost

$$\frac{\partial C}{\partial D} = 0.8\frac{\partial V_o}{\partial D} + 0.1\frac{\partial S}{\partial D}\,, \qquad \frac{\partial C}{\partial H} = 0.8\frac{\partial V_o}{\partial H} + 0.1\frac{\partial S}{\partial H}\,, \quad (2.4.21)$$

and the price

$$\frac{\partial P}{\partial D} = (2.5 - 0.04V_o)\frac{\partial V_o}{\partial D}\,, \qquad \frac{\partial P}{\partial H} = (2.5 - 0.04V_o)\frac{\partial V_o}{\partial H}\,. \quad (2.4.22)$$

Finally, we differentiate the equations for V_o

$$\frac{\partial V_o}{\partial D} = 0.5\pi DH/1.8\,, \qquad \frac{\partial V_o}{\partial H} = 0.25\pi D^2/1.8\,, \quad (2.4.23)$$

and S

$$\frac{\partial S}{\partial D} = \pi D + \pi H\,, \qquad \frac{\partial S}{\partial H} = \pi D\,. \quad (2.4.24)$$

Now we can substitute the values of D and H at the candidate optimum into the above equations in reverse order. First we evaluate the derivatives of the volume and the area. For $D = 2.04$ in. and $H = 4.08$ in. we get

$$\frac{\partial V_o}{\partial D} = 7.263\ oz/in\,, \quad \frac{\partial V_o}{\partial H} = 1.816\ oz/in\,, \quad \frac{\partial S}{\partial D} = 19.23\ in\,, \quad \frac{\partial S}{\partial H} = 6.409\ in\,. \quad (2.4.25)$$

Next we evaluate the derivatives of the price from Eq. 2.4.22

$$\frac{\partial P}{\partial D} = 16.01\ cents/in\,, \qquad \frac{\partial P}{\partial H} = 4.001\ cents/in\,, \quad (2.4.26)$$

and the derivatives of the cost from Eq. 2.4.21

$$\frac{\partial C}{\partial D} = 7.733\ cents/in\,, \qquad \frac{\partial C}{\partial H} = 2.094\ cents/in\,. \quad (2.4.27)$$

Finally, we can substitute these values into the optimality conditions, Eq. 2.4.20

$$\frac{\partial L}{\partial D} = -0.02774 + 0.4902\lambda = 0\,, \quad (2.4.28)$$
$$\frac{\partial L}{\partial H} = 0.01469 - 0.2451\lambda = 0\,. \quad (2.4.29)$$

The first equation gives us a value of $\lambda = 0.0566$, while the other gives us $\lambda = 0.0599$. The difference reflects the fact that the candidate point, which we found by graphical optimization, is not exactly the optimum. However, the closeness of the two values indicates that we are close to the optimum design. For this problem we can check the exact optimum by substituting $H = 2D$ from the constraint into the objective function. After some algebra we get

$$p_o = 1.7 - \frac{0.9}{D} - 0.0174533D^2\,. \quad (2.4.30)$$

Setting the derivative of p_o to zero we get $D = 2.036$ in., so that $H = 4.072$, and $p_o = 1.11065$ cents per ounce. That is, the small change in the design variables did not change the objective function even in the sixth significant digit. This insensitivity of the value of the optimum to a small error in the design variables is extreme. However, in most problems the value of the optimum objective is not very sensitive to such small errors in the design variables.

Also, in most problems we do not have the luxury of solving the optimization problem analytically, and so we will continue with the slightly inaccurate candidate optimum, and take an average value of the Lagrange multiplier, $\lambda \approx 0.0572$ to estimate the effect of changing the aspect ratio constraint. From Eq. 2.4.15 we have

$$\frac{dp_o^*}{dl} = -\frac{df^*}{dl} = -\frac{\partial L}{\partial l} = \frac{1}{V_o}\frac{\partial P}{\partial l} - \lambda\frac{\partial h}{\partial l} = \tag{2.4.31}$$

$$\frac{5}{V_o} - \lambda\frac{0.2H}{(2+0.2l)^2 D} = 0.67\ cents, . \tag{2.4.32}$$

Note that the second term, involving the Lagrange multiplier, is only -0.0056 cents per ounce. That is, most of the change in profit comes from the 5 cents higher price that can be charged for the taller can, and the decrease in efficiency due to the higher aspect ratio has only a small effect.

To predict the change in profit per ounce we can use the derivative

$$p_o^*(l=1) = p_o^*(l=0) + \frac{dp_o^*}{dl} = 1.11 + 0.67 = 1.78\ cents/ounce\,. \tag{2.4.33}$$

To check this prediction, we can again use the aspect ratio constraint, $H = 2.2D$, and obtain the optimum can corresponding to this constraint. With some more algebra we get the following expression for the profit per ounce

$$p_o = 1.7 - \frac{0.883636}{D} - 0.0191986D^3 + \frac{5.20871}{D^3}\,. \tag{2.4.34}$$

This time we cannot find a real solution that maximizes p_o. The last term, corresponding to the 5 cents extra that a taller can commands, dominates. With this extra charge, it pays to make the can as small as possible, so that the additional 5 cents will have large effect on the profit per volume. We assume then that we will go to the smallest can possible, with $V_o = 5$ oz., and calculate the dimensions of $D = 1.733$ in., $H = 3.816$ corresponding to $p_o = 2.090$ cents per ounce. This is substantially higher than the predicted value of 1.7828 cents. The discrepancy is due to the fact that Eq. 2.4.33 is only a linear extrapolation. • • •

Next we consider the case of multiple equality constraints. The optimization problem is written as

$$\begin{aligned} &\textbf{minimize} && f(\mathbf{x})\,, \\ &\textbf{such that} && h_i(\mathbf{x}) = 0, \qquad i = 1,\ldots,n_e\,. \end{aligned} \tag{2.4.35}$$

For this case we define Lagrange multipliers for each constraint, λ_i, $i = 1,\ldots,n_e$, and a Lagrangian function L given as

$$L = f + \sum_{i=1}^{n_e} \lambda_i h_i\,. \tag{2.4.36}$$

Then, if the gradients of the constraints are linearly independent at a point $\mathbf{x}^*$, it can be shown that a necessary condition for an optimum is that the derivatives of the Lagrangian with respect to all the design variables are zero, that is

$$\frac{\partial f}{\partial x_j}(\mathbf{x}^*) + \sum_{i=1}^{n_e} \lambda_i \frac{\partial h_i}{\partial x_j} = 0\,, \quad j = 1,\ldots,n \tag{2.4.37}$$

This is again a stationarity condition, equally applicable to a minimum or a maximum. As in the case of a single constraint, the derivative of the Lagrangian with respect to a parameter is equal to the derivative of the optimum objective function with respect to that parameter. That is, if the objective function and constraints depend on a parameter p, $f = f(\mathbf{x}, p)$, and $h_i = h_i(\mathbf{x}, p)$, Eq. 2.4.15 applies.

When we have inequality constraints the treatment remains very similar. Consider the standard format

$$\begin{aligned} &\textbf{minimize} && f(\mathbf{x})\,, && \\ &\textbf{such that} && h_i(\mathbf{x}) = 0, && i = 1, \ldots, n_e\,, \qquad (2.4.38)\\ & && g_j(\mathbf{x}) \le 0, && j = 1, \ldots, n_g\,, \end{aligned}$$

(2.4.39)

where any upper and lower limits on the design variables are assumed here to be included in the inequality constraints. We now construct a Lagrangian as

$$L = f + \sum_{i=1}^{n_e} \lambda_i h_i + \sum_{i=n_e+1}^{ne+ng} \lambda_i g_{i-n_e}\,. \tag{2.4.40}$$

The conditions for a stationary point are again that the derivatives of the Lagrangian with respect to all the design variables are equal to zero. Again, if the objective function and constraints depend on a parameter p, then the derivative of the optimum objective function f^* with respect to that parameter, are given by Eq. 2.4.15. That is

$$\frac{df^*}{dp} = \frac{\partial f}{\partial p} + \sum_{i=1}^{n_e} \lambda_i \frac{\partial h_i}{\partial p} + \sum_{i=1}^{ng} \lambda_{i+n_e} \frac{\partial g_i}{\partial p}\,. \tag{2.4.41}$$

From Eq. 2.4.41 it is seen that the Lagrange multipliers measure the sensitivity of the objective function to changes in the constraints. For these reasons, the Lagrange multipliers are sometimes called 'shadow prices'. It should also be clear that when a constraint is not active, it should not contribute anything to the derivative of the optimum objective with respect to the parameter. The reason is that for small changes in p, the constraint will remain inactive, so that the solution of the optimization problem will not depend at all on this constraint. Furthermore, consider the case when the parameter p enters an inequality constraints in the form of

$$g_i(\mathbf{x}, p) = \bar{g}(\mathbf{x}) + p \le 0\,, \tag{2.4.42}$$

so that $\partial g_i / \partial p = 1$. In this case, it is clear that as p increases, it becomes more difficult to satisfy the constraint, so that the optimum value of the objective function could become worse but not improve. That is, $df^*/dp \ge 0$. From Eq. 2.4.41 this indicates that λ_i has to be positive.

Altogether, for a problem with inequality constraints, we have two other conditions besides the vanishing of the derivatives of the Lagrangian. The Lagrange multipliers associated with inactive constraints have to be zero, and the Lagrange multipliers associated with active inequality constraints have to be positive. These conditions are summarized as

$$\begin{aligned} \frac{\partial f}{\partial x_k} + \sum_{i=1}^{n_e} \lambda_i \frac{\partial h_i}{\partial x_k} + \sum_{i=1}^{ng} \lambda_{i+n_e} \frac{\partial g_i}{\partial x_k} &= 0, \qquad ,k = 1, \ldots, n \\ g_i \lambda_{i+n_e} = 0\,, \qquad & i = 1, \ldots, n_g\,, \\ \lambda_{i+n_e} \ge 0\,, \qquad & i = 1, \ldots, n_g\,. \end{aligned} \tag{2.4.43}$$

The second equation, which stipulates that the product of the inequality constraint times its Lagrange multiplier is equal to zero is a convenient way of requiring zero Lagrange multipliers for inactive constraints, as inactive constraints have values different from zero. These conditions, which need to be satisfied at the minimizer $\mathbf{x}^*$ are called the Kuhn-Tucker conditions.

The Kuhn-Tucker conditions are necessary but not sufficient for a minimum. To get sufficient conditions, we again need to examine the second derivatives. However, now we need to consider only directions that are tangent to the active constraints. That is, we define the *tangent space* T as

$$T(\mathbf{x}^*) = \{\mathbf{y}; \mathbf{y}^T \nabla g_{aj}(\mathbf{x}^*) = 0,\ \mathbf{y}^T \nabla h_i(\mathbf{x}^*) = 0,\ i = 1, n_e\}\,, \tag{2.4.44}$$

where ∇f is the gradient of f, that is the vector of first derivatives of f, and where g_{aj} indicates the active inequality constraints at the minimizer, $\mathbf{x}^*$. That is T is the space of all vectors which are tangent to the active constraints, that is vectors that are perpendicular to the gradient of these constraints. A necessary second-order condition for a minimum is that

$$\mathbf{y}^T \mathbf{H}_L(\mathbf{x}^*)\mathbf{y} \geq 0\,, \quad \text{for every } \mathbf{y} in T\,, \tag{2.4.45}$$

where H_L is the Hessian matrix of the Lagrangian. A strict equality in the equation is a sufficient condition for a minimum, when the point satisfies the Kuhn-Tucker conditions.

2.5 Conditions for global optimality

We have seen that the condition for global optimality of unconstrained minimum requiress the convexity of the objective function. A similar condition for a global minimum requires the convexity of the inequality constraints. However, more rigorously, the condition is that the feasible domain is convex, and so we start by defining a convex domain.

A set of points S is convex, if for any pair of points $\mathbf{x}_1$ and $\mathbf{x}_2$ that belong to the set, the line segment connecting them also belongs to the set. That is,

$$\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2 \text{ belong to } S\,, \tag{2.5.1}$$

It can be shown that the feasible domain is convex if all the inequality constraint function are convex and all the equality constraints are linear.

This condition, however, means not only that the local minimizer is a global minimizer, but also that there is only one local minimum. This condition, therefore, is quite strong, and is rarely satisfied. For most engineering problems, it is impossible to prove that a point is a global minimum, and instead we have to keep looking for local minima, and be satisfied with probabilistic estimates of the chance that we have found the local minimum which is also the global one.

2.6 Exercises

Find all the stationary points for the following functions and characterize them as minima, maxima, saddle points or inflection points:

1. $f = 2x_1^2 + x_1x_2 + 3x_2^2 + 6x_1$
2. $f = x_1^2 + 4x_1x_2 + 3x_2^2 + 5$
3. $f = \cos x_1 \sin x_2$

4. For the soft-drink can problem in Chapter 1, the optimum design for a standard-size can ($l = 0$) was $D = 2.581$ in., $H = 5.162$ in., and the profit per can was 15.77 cents. The two active constraints for that problem were the aspect ratio and the volume constraints. Ignoring the other constraints we can formulate the problem as

$$\begin{aligned} &\textbf{minimize} && -p_c = C - P \\ &\textbf{such that} && g_1 = 1 - \frac{H}{(2 + 0.2l)D} \leq 0\,, \\ & && g_2 = \frac{V_o}{15} - 1 \leq 0\,. \end{aligned} \tag{2.6.1}$$

Find the Lagrange multipliers associated with this solution, and use them to estimate the profit per can for the tall can design ($l = 1$). Check how well the estimate agrees with the actual optimum at $l = 1$.