Review of Optimization Basics

1. Introduction

Electricity markets throughout the US are said to have a two-settlement structure. The reason for this is that the structure includes two different markets:

- The real time or balancing market
- The day-ahead market

These are two kinds of tools which are used in these two markets, and both are optimization tools.

- The security-constrained unit commitment (SCUC), which is used in the day-ahead market;
- The security constrained economic dispatch (SCED), which is used in both real time and day-ahead markets.

Although the SCED is most generally a nonlinear optimization problem, most electricity markets solve it by linearizing it and applying linear programming.

On the other hand, SCUC is most generally a mixed integer nonlinear optimization problem. Again, most electricity markets solve it by linearizing it so that it becomes a mixed integer linear optimization problem.

Thus, to fully understand electricity markets, one needs to also these two kinds of optimization problems.

Therefore, in these notes, we wish to review some basic convex programming concepts. This material in these notes should be read together with other resources. Some good additional resources include Appendix 3A of [1].

2. Convexity

At the end of our discussion on "Cost Curves," we considered generator types that had non-convex cost curves. These included all steam plants that have multiple valve points. It also includes combined cycle units.

Let's recall what we mean by convexity. There are formal mathematical definitions of convexity, but we will try to remain as simple as possible.

<u>Convexity #1</u>: A function f(x) is convex in an interval if its second derivative is positive on that interval.

The quadratic function $f(x)=x^2$ is convex, for example, since $f'(x)=2x$ and $f''(x)=2$.

This second derivative test is that the rate of change of slope should increase with x. One can see that the function $f(x)=x^2$ satisfies this property as shown in Fig. 1a below.

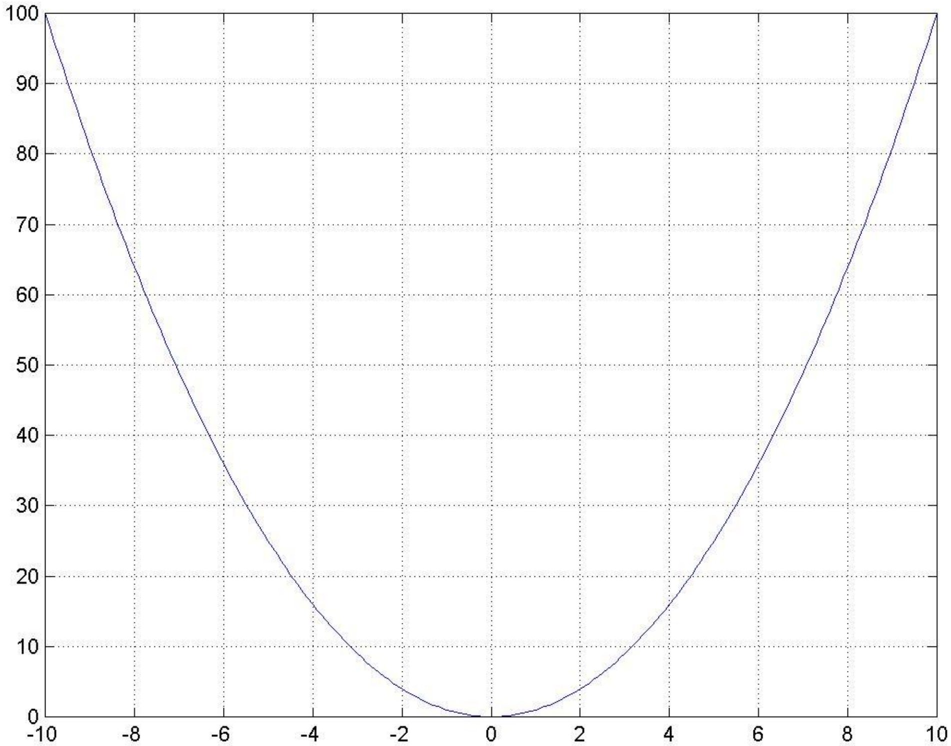


Fig. 1a

The second derivative test, although sufficient, is not necessary, as a second derivative need not even exist.

We may observe this in Fig. 1b [2] where we see some functions for which the first and higher derivatives do not exist.

We also observe the line segment drawn across the function f(x) on the left of Fig. 1b. This line segment indicates another kind of test for convexity, as follows.

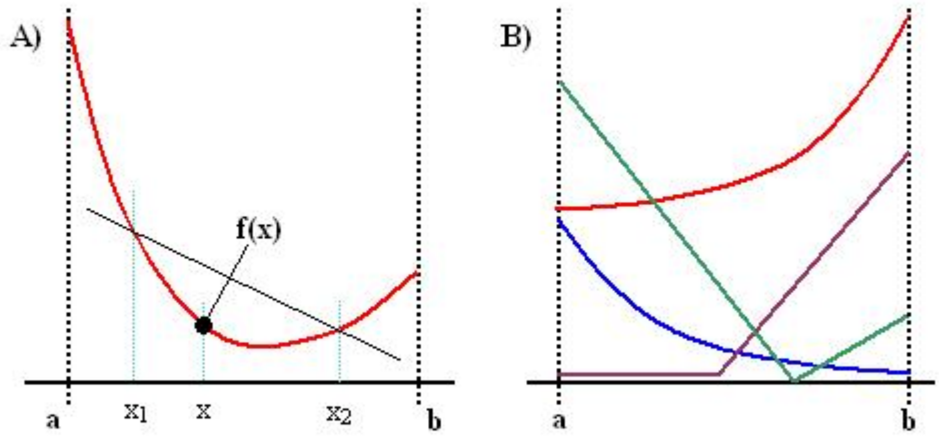
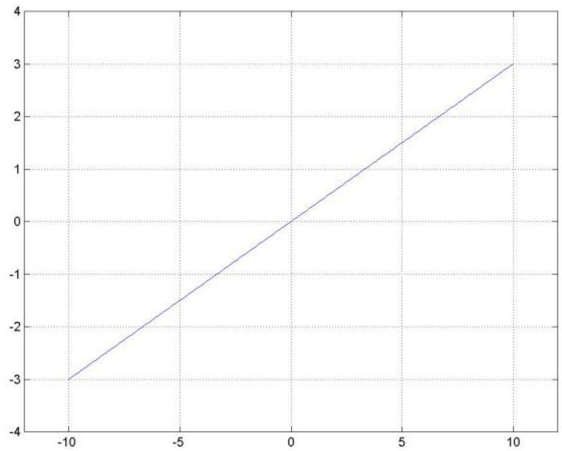


Fig. 1b

<u>Convexity #2</u>: A function is convex if a line drawn between any two points on the function remains on or above the function in the interval between the two points.

We see that all of the functions in Fig. 1b satisfy Convexity #2 definition, and this is a definition that is workable within this class.

Question: Is a linear function convex?

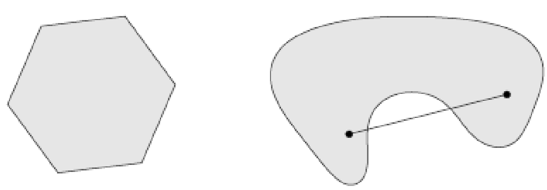


Answer is "yes" since a line drawn between any two points on the function remains ***<u>on</u>*** the function.

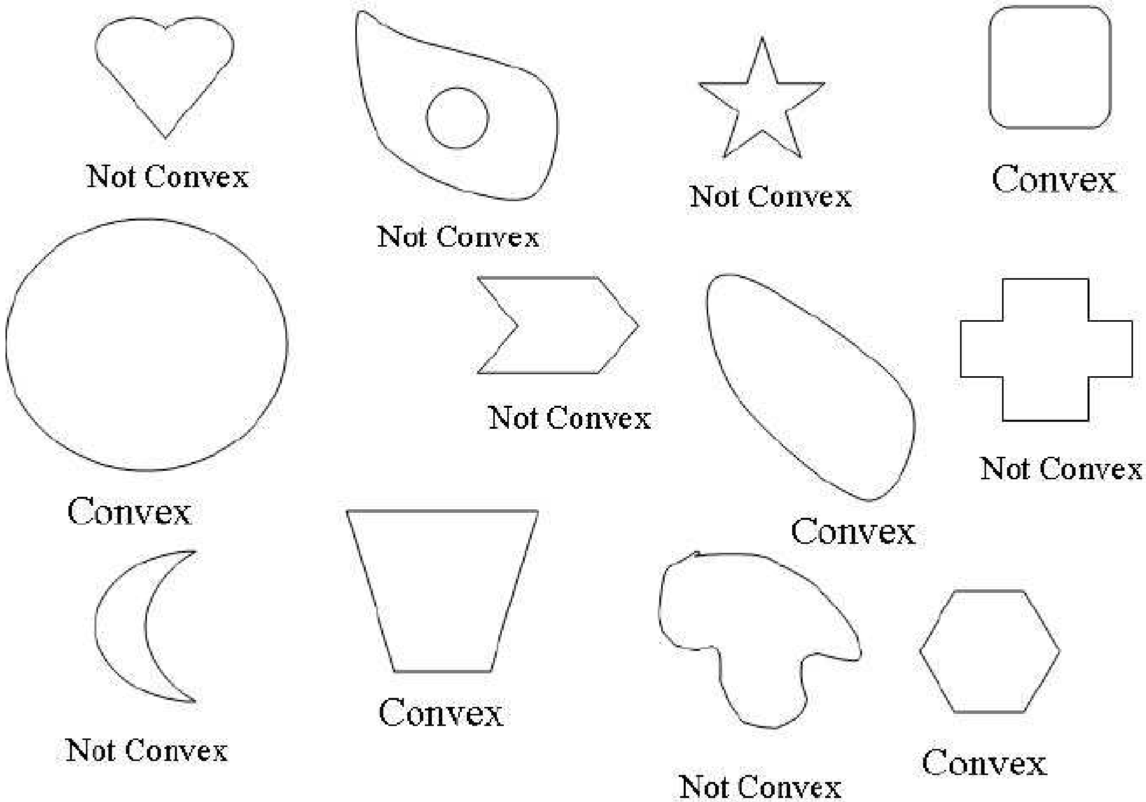
This leads to another important definition.

<u>Convex set</u>: A set C is convex if a line segment between any two points in C lies in C.

Which of the two sets below are convex [3]?



Here are some more examples of convex and non-convex sets [4].



3. Global vs. local optimum

We will use certain techniques to solve optimization problems. These techniques will result in solutions. An important quality of those solutions is whether they are local optimal or global optimal.

Example: Solve the following problem:

$$\text{Minimize } f(x)=x^2$$

We know how to do this as follows:

$$f'(x)=2x=0 \Rightarrow x^*=0.$$

The solution we just found is a local optimum. It is also the global optimum.

Example: Solve the following optimization problem.

$$\text{Minimize } f(x)=x^3-17x^2+80x-100$$

Applying the same procedure, we obtain:

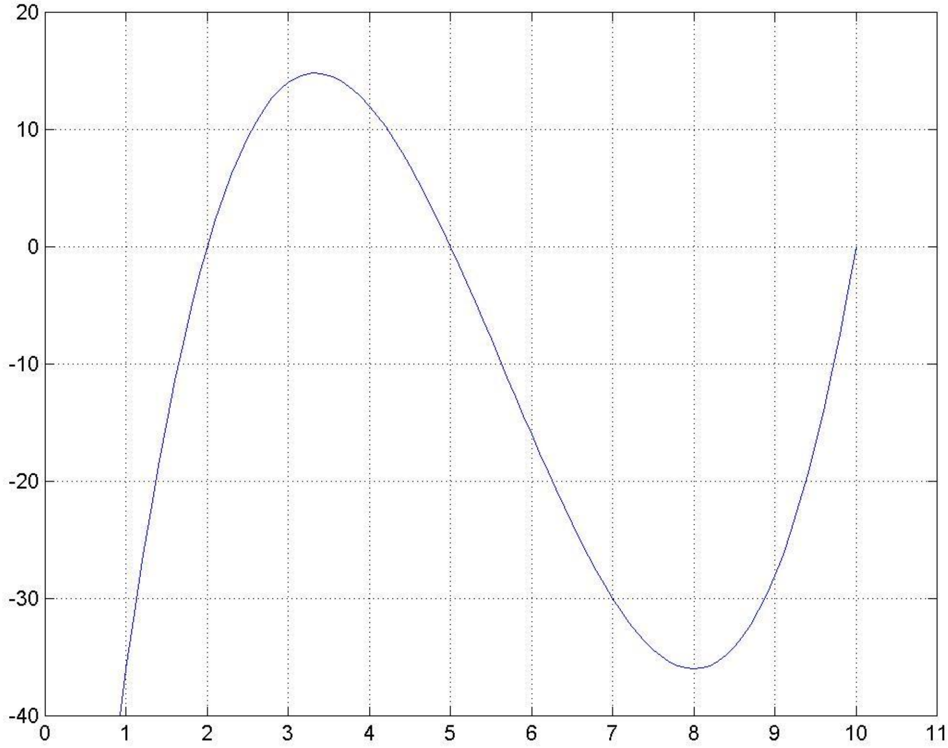
$$f'(x)=3x^2-34x+80=0$$

Solving the above results in x=3.33 and x=8.

We have two problems to address:

- Which is the best solution?
- Is the best solution the global solution?

We can immediately see the answers to our questions by plotting the original function f(x), as shown below.

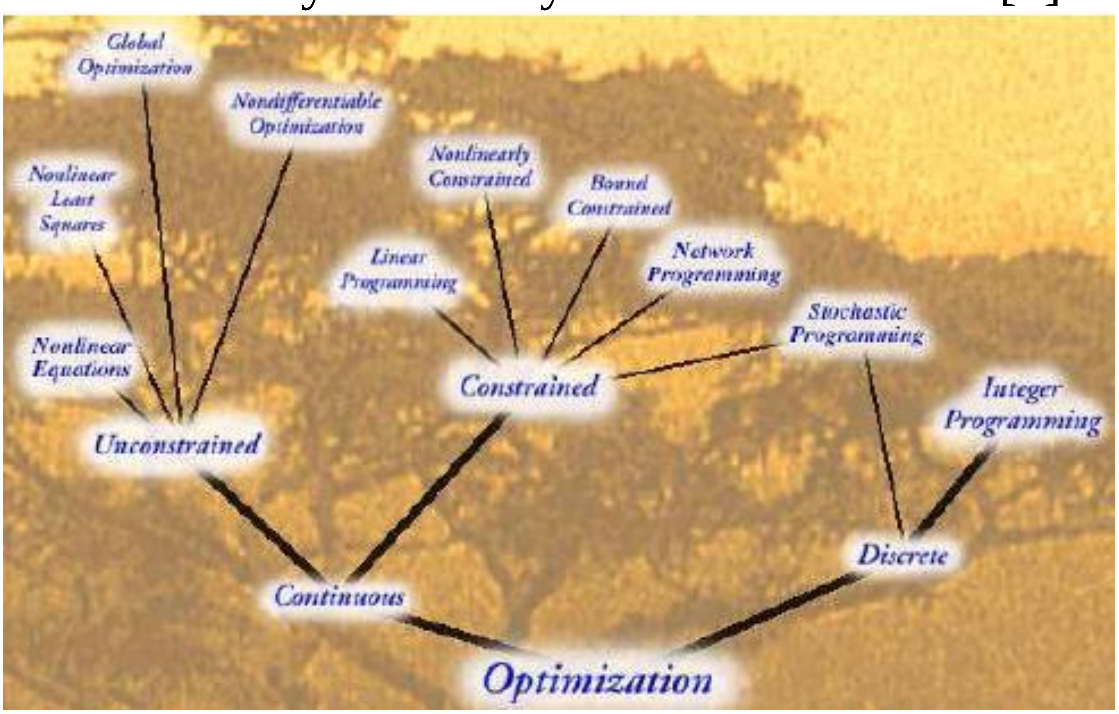


And so we see that x=8 is the best solution, but it is not the global solution. It appears that this particular problem has no global solution, i.e., the smaller we make x, the more we can minimize f. It is unbounded.

This shows that when minimizing a function, if we want to be sure that we can get a global solution via differentiation, we need to impose some requirements on our objective function. We will also need to impose some requirements on the set of possible values that the solution x* may take.

4. Convex optimization

Optimization problems can be classified in a number of different ways. One way is illustrated below [5].



In this figure, the branch of optimization is divided into continuous and discrete. We observe that the continuous types are then divided into unconstrained and constrained, with the constrained being divided into several types including linear programming.

In this class, we will be interested in the linear programming problems and in the discrete integer programming problems.

Another way to classify optimization problems is by whether they are convex or not.

Consider the following problem

$$\min \ f(\underline{x})$$

$$\text{s.t. } \underline{\mathrm{x}} \in S$$

This problem says:

minimize f($\underline{\mathrm{x}}$) subject to the requirement that the point $\underline{\mathrm{x}}$ must lie in a region defined by S.

<u>Definition</u>: If f($\underline{\mathrm{x}}$) is a convex function, and if S is a convex set, then the above problem is a convex programming problem.

The desirable quality of a convex programming problem is that any *locally optimal solution* is also a *globally optimal solution*.

Thus, if we have a method of finding a locally optimal solution, then that method also finds for us the globally optimum solution.

The undesirable quality of a nonconvex programming problem is that any method which finds a locally optimal solution does not necessarily find the globally optimum solution.

These are very important qualities of optimization problems, and they motivate us to provide another classification of optimization problems:
- Convex programming problems
- Nonconvex programming problems.

Of interest to us is that
  - linear programs are convex programming problems.
  - mixed integer programs are nonconvex programming problems.

We focus in these notes on some fundamental concepts related to convex programming.

We will address nonconvex programming problems when we look at the unit commitment problem.

5. Problem Statement

The general problem that we want to solve is the two-variable equality-constrained minimization problem, as follows:

$$\begin{aligned} &\min \ f(x_1, x_2) \\ &s.t. \ \ h(x_1, x_2) = c \end{aligned} \tag{1}$$

Problem (1) is the 2-dimensional characterization of a similar n-dimensional problem:

$$\begin{aligned} &\min \ f(\underline{x}) \\ &\text{s.t.} \ \ h(\underline{x}) = c \end{aligned} \tag{2}$$

And problem (2) is *n*-dimensional, single-constraint characterization of a similar n-dimensional, multi-constraint problem:

$$\begin{aligned} &\min \ f(\underline{x}) \\ &\text{s.t.} \ \ \underline{h}(\underline{x}) = \underline{c} \end{aligned} \tag{3}$$

Whatever we can conclude about (1) will also apply to (2) and (3).

6. Contour maps

To facilitate discussion about our two-dimensional problem (1), we need to fully understand what a contour map is. A contour map is a 2-dimensional plane, i.e., a coordinate system in two variables, say x_1 and x_2, that illustrates curves (or contours) of constant functional value $f(x_1, x_2)$.

Example 1: Draw the contour map for $f(x_1, x_2) = x_1^2 + x_2^2$.

Solution: The below Matlab code does it:

```
[X,Y] = meshgrid(-2.0:.2:2.0,-2.0:.2:2.0);
Z = X.^2+Y.^2;
[c,h]=contour(X,Y,Z);
clabel(c,h);
grid;
xlabel('x1');
ylabel('x2');
```

Figure 1 illustrates.

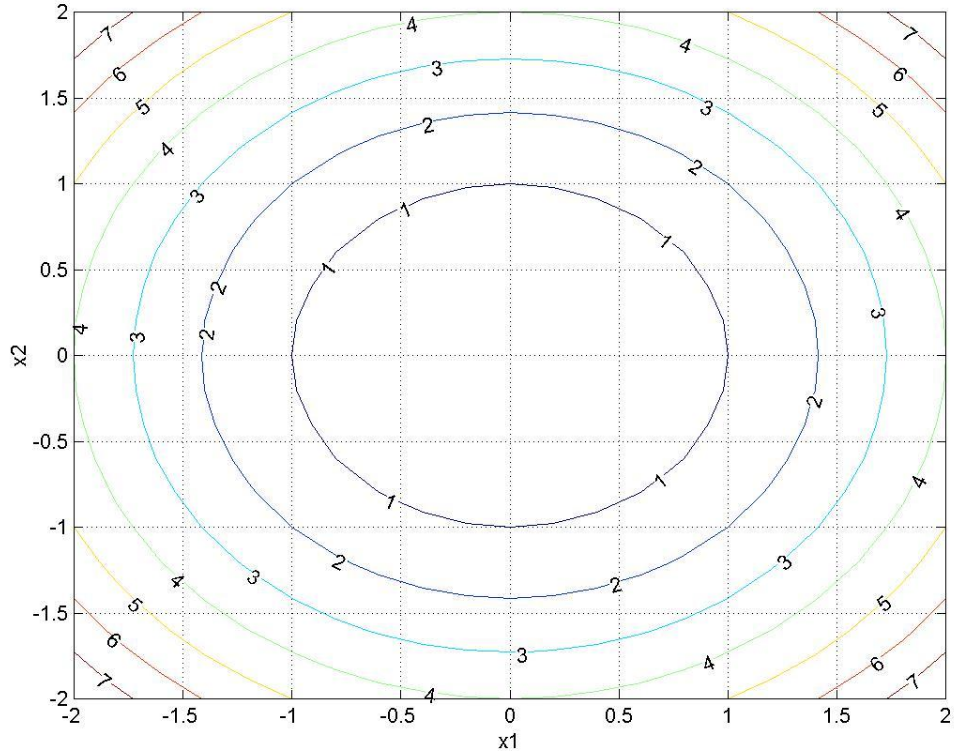


Fig. 1: Contour map for $f(x_1, x_2) = x_1^2 + x_2^2$

The numbers on each contour show the value of $f(x_1,x_2)$ for that contour, and so the contours show $f(x_1,x_2)=1$, $f(x_1,x_2)=2$, $f(x_1,x_2)=3$, $f(x_1,x_2)=4$, $f(x_1,x_2)=5$, $f(x_1,x_2)=6$, and $f(x_1,x_2)=7$.

We could show similar information with a 3-D figure, where the third axis provides values of $f(x_1,x_2)$, as shown in Fig. 2. I used the following commands to get Fig. 2.

```
[X,Y] = meshgrid(-2.0:.2:2.0,-2.0:.2:2.0);
Z = X.^2+Y.^2;
surfc(X,Y,Z)
xlabel('x1')
ylabel('x2')
zlabel('f(x1,x2)')
```

Figure 2 also shows the contours, where we see that each contour of fixed value *f* is the projection onto the *x1-x2* plane of a horizontal slice made of the 3-D figure at a value *f* above the *x1-x2* plane.

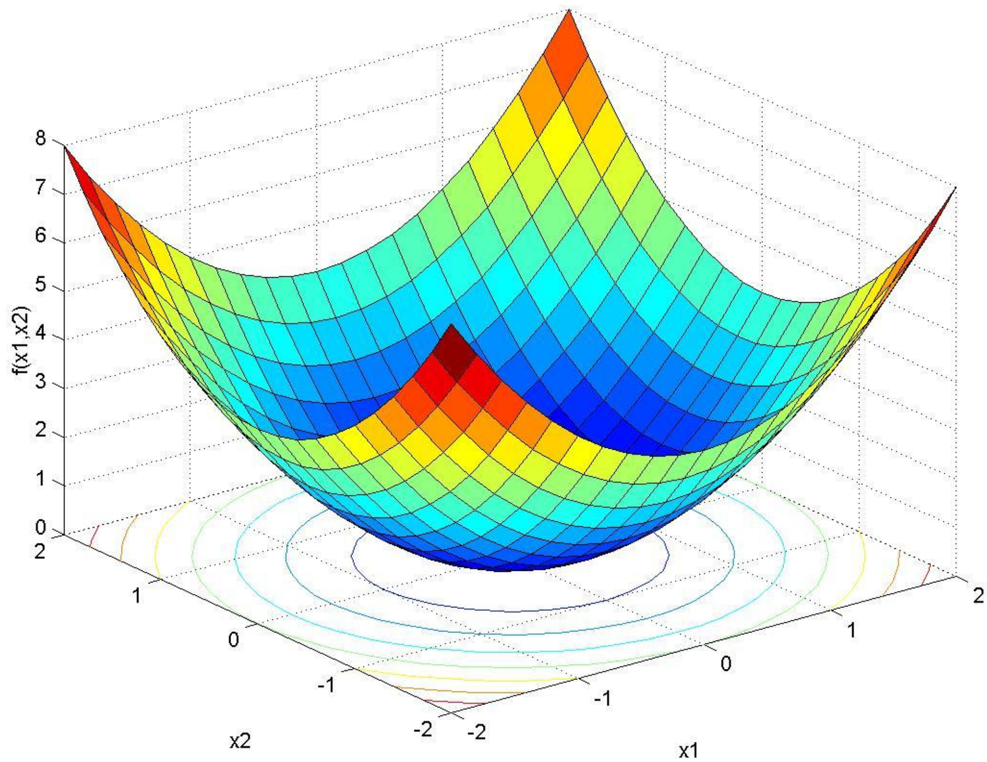


Fig. 2: 3-D illustration of $f(x_1, x_2) = x_1^2 + x_2^2$

7. Understanding problem 1 solution procedure

We desire to solve Problem 1. Let's consider it by following up on the example which we began in the previous section.

Example 2: Use graphical analysis to solve the following specific instance of Problem 1.

$$\min \ f(x_1, x_2) = x_1^2 + x_2^2$$
$$s.t. \ \ h(x_1,x_2) = 2x_1 x_2 = 3$$

To visualize the solution to this problem, let's express the equality constraint where x_2 is the dependent variable and x_1 is the independent variable, according to:

$$2x_1 x_2 = 3 \Rightarrow x_2 = \frac{3}{2x_1}$$

This is a function that we can plot on our x_1, x_2 Cartesian plane, and we will do so by superimposing it over the contour plot of $f(x_1,x_2)$, as in Fig. 3.

One can immediately identify the answer from Fig. 3, because of two requirements of our problem:

- $f(x_1,x_2)$ **must** be minimized, and so we would like the solution to be as close to the origin as possible;
- the solution **must** be on the thick line in the right-hand corner of the plot, since this line represents the equality constraint.

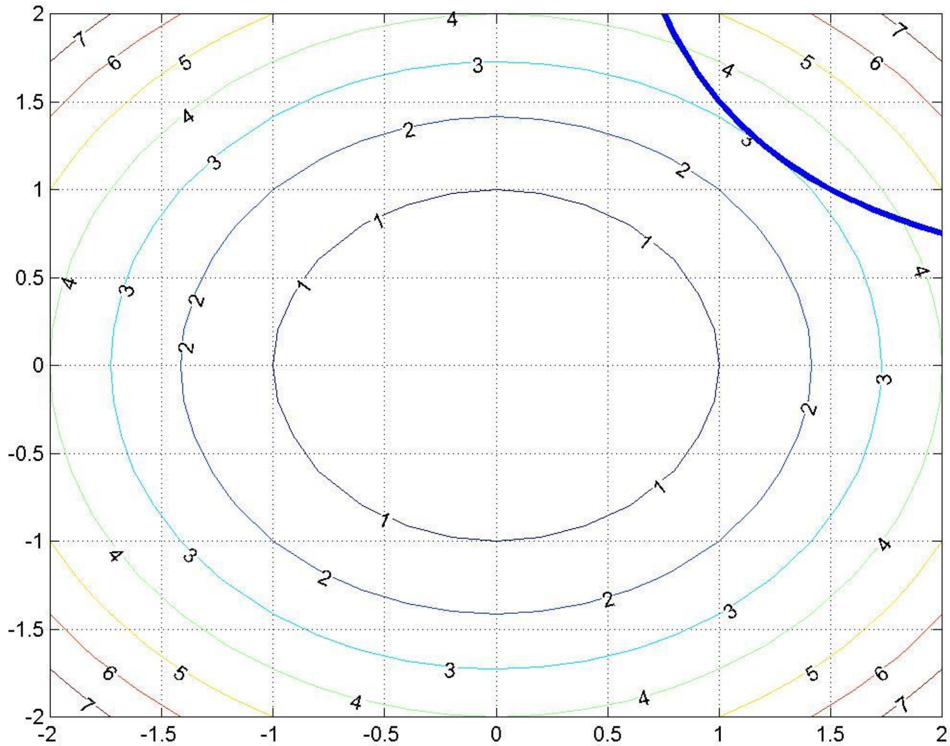


Fig. 3: Contour plots with equality constraint

From the plot, we see that the solution is about $(x_1, x_2)_0 \approx (1.25, 1.25)$, as this point results in the smallest possible value of *f* that is still on the equality constraint. The value of *f* at this point is 3. We will check this analytically in the next section.

Before we do that, note in Fig. 3 the intersection of the equality constraint and the contour *f=3*. Notice

- any contour *f<3* does not intersect the equality constraint;
- any contour *f>3* intersects the equality constraint at two points.

This means that the *f=3* contour and the equality constraint *just touch* each other at the point identified as the problem solution, about $(x_1, x_2)_0 \approx (1.25, 1.25)$. The notion of "just touching" implies

> *The two curves are tangent[1] to one another at the solution point.*

The notion of tangency is equivalent to another one:

> *The normal (gradient) vectors[2] of the two curves, at the solution (tangent) point, are parallel.*

From multivariable calculus, we know we can express a normal vector to a curve as the gradient of the function characterizing that curve.

The gradient operator is ∇. It operates on a function by taking first derivatives with respect to each variable. For example,

$$\nabla\left\{x_1^2 + x_2^2\right\} = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$$

$$\nabla\left\{2x_1x_2 - 3\right\} = \begin{bmatrix} 2x_2 \\ 2x_1 \end{bmatrix}$$

And then, we can evaluate those derivatives at some certain value of $\underline{x}$, a point which, notationally, we refer to as $\underline{x}_0=(x_1, x_2)_0$. Gradients have magnitude and angle.

---

[1] The word "tangent" comes from the Latin word *tangere*, which means "to touch."
[2] Normal vectors are *perpendicular* (2-space) and more generally, *orthogonal* (n-space), to the tangent.

The functions of the two curves are $f(x_1,x_2)$ and $h(x_1,x_2)$. If the two normal vectors are to be parallel to one another at the point $\underline{x}_0$, then

$$\nabla f(x_1,x_2)_0 = \lambda \nabla \left(h(x_1,x_2)_0 - c \right) \tag{4a}$$

Alternatively,

$$\nabla f(x_1,x_2)_0 = -\lambda \nabla \left(c - h(x_1,x_2)_0 \right) \tag{4b}$$

The reason for parameter λ is as follows.

- Recall that gradient gives both magnitude and direction;
- Yet, the only thing we know is that the two gradients have the same direction - we do *not* know that they also have the same magnitude;
- And so we insert λ as a "multiplier" to account for the fact that the two gradients may not have the same magnitudes.

Because it was Joseph Louis Lagrange (1736-1813) who first thought of the "calculus of variations," as it was called then (and still is by mathematicians), we call λ the "Lagrange multiplier." We will see later that Lagrange multipliers are very important in relation to locational marginal prices.

Now from (4a), we move the right side to the left:

$$\nabla f(x_1,x_2)_0 - \lambda \left(\nabla h(x_1,x_2)_0 - c \right) = 0 \tag{5a}$$

Alternatively, from (4b):

$$\nabla f(x_1,x_2)_0 + \lambda \left(c - \nabla h(x_1,x_2)_0 \right) = 0 \tag{5b}$$

Since the gradient operation is precisely the same operation on f as it is on h (taking first derivatives with respect to x_1 and x_2), we can write (5a) as

$$\nabla\left\{f(x_1,x_2)_0-\lambda\left(h(x_1,x_2)_0-c\right)\right\}=0 \tag{6a}$$

Alternatively,

$$\nabla\left\{f(x_1,x_2)_0+\lambda\left(c-h(x_1,x_2)_0\right)\right\}=0 \tag{6b}$$

And so we observe that the solution, i.e., the value of (x_1, x_2) that identifies a feasible point corresponding to a minimum value of f, will satisfy the partial derivative equations associated with (6), according to

$$\nabla\left\{f(x_1,x_2)-\lambda\left(h(x_1,x_2)-c\right)\right\}_0=0 \tag{7a}$$

Alternatively,

$$\nabla\left\{f(x_1,x_2)+\lambda\left(c-h(x_1,x_2)\right)\right\}_0=0 \tag{7b}$$

Expressing (7a) in terms of partial derivatives yields:

$$\begin{bmatrix} \dfrac{\partial}{\partial x_1}\left(f(x_1,x_2)-\lambda\left(h(x_1,x_2)-c\right)\right) \\ \dfrac{\partial}{\partial x_2}\left(f(x_1,x_2)-\lambda\left(h(x_1,x_2)-c\right)\right) \end{bmatrix}_0=\begin{bmatrix}0\\0\end{bmatrix} \tag{8a}$$

Alternatively, using (7b):

$$\begin{bmatrix} \dfrac{\partial}{\partial x_1}\left(f(x_1,x_2)+\lambda\left(c-h(x_1,x_2)\right)\right) \\ \dfrac{\partial}{\partial x_2}\left(f(x_1,x_2)+\lambda\left(c-h(x_1,x_2)\right)\right) \end{bmatrix}_0=\begin{bmatrix}0\\0\end{bmatrix} \tag{8b}$$

Notice in (7) and (8) that the “0”-subscript, indicating “evaluation at the solution,” has been shifted to the outside of the brackets, implying “evaluation at the solution” occurs after taking the derivatives.

But let’s think of it in a little different fashion. Let’s write down the partial derivatives without knowing the solution (and we certainly can do that). Then, eq. (7) (or (8)) provides equations that can be used to find the solution, by solving them simultaneously.

Of course, there is still one small issue. By (8) we see that we only have two equations, yet we have the unknowns x_1, x_2, and λ. We cannot solve two equations in three unknowns! What do we do????

This issue is resolved by recalling that we actually do have a third equation: $h(x_1, x_2)-c=0$ (or alternatively, c- $h(x_1, x_2)$). This is just our equality constraint. And so we see that we have 3 equations, 3 unknowns, and at least in principle, we can solve for our unknowns x_1, x_2, λ. To summarize, the 3 equations are:

$$\begin{bmatrix} \frac{\partial}{\partial x_1}\left(f(x_1,x_2)-\lambda\left(h(x_1,x_2)-c\right)\right) \\ \frac{\partial}{\partial x_2}\left(f(x_1,x_2)-\lambda\left|h(x_1,x_2)-c\right|\right) \\ h(x_1,x_2)-c \end{bmatrix}_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad (9a)$$

Alternatively,

$$\begin{bmatrix} \frac{\partial}{\partial x_1}\left(f(x_1,x_2) + \lambda\left(c - h(x_1,x_2)\right)\right) \\ \frac{\partial}{\partial x_2}\left(f(x_1,x_2) + \lambda\left(c - h(x_1,x_2)\right)\right) \\ c - h(x_1,x_2) \end{bmatrix}_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad (9b)$$

We make one more startling observation, and that is that the three equations are simply partial derivatives of the function $f(x_1,x_2) - \lambda\left(h(x_1,x_2) - c\right)$ (or alternatively $f(x_1,x_2) + \lambda\left(c - h(x_1,x_2)\right)$) with respect to each of our unknowns!!!! This is obviously true for the first two equations in (9), but it is not so obviously true for the last one. But to see it, observe:

$$\begin{aligned} &\frac{\partial}{\partial \lambda}\left(f(x_1,x_2) - \lambda\left(h(x_1,x_2) - c\right)\right) = 0 \\ &\Rightarrow -h(x_1,x_2) + c = 0 \Rightarrow h(x_1,x_2) = c \end{aligned} \qquad (10a)$$

Alternatively

$$\begin{aligned} &\frac{\partial}{\partial \lambda}\left(f(x_1,x_2) + \lambda\left(c - h(x_1,x_2)\right)\right) = 0 \\ &\Rightarrow c - h(x_1,x_2) = 0 \Rightarrow h(x_1,x_2) = c \end{aligned} \qquad (10b)$$

We are now in a position to formalize solution to our 2-dimensional, one constraint problem. Let's define the Lagrangian function as:

$$\mathcal{L}(x_1,x_2,\lambda) = f(x_1,x_2) - \lambda\left(h(x_1,x_2) - c\right) \qquad (11a)$$

$$\mathcal{L}F(x_1,x_2,\lambda) = f(x_1,x_2) + \lambda\left(c - h(x_1,x_2)\right) \qquad (11b)$$

Then the first-order conditions for solving this problem are

$$\nabla \mathcal{L}(x_1, x_2, \lambda) = 0 \tag{12}$$

or,

$$\begin{aligned} \frac{\partial}{\partial x_1} \mathcal{L}(x_1, x_2, \lambda) &= 0 \\ \frac{\partial}{\partial x_2} \mathcal{L}(x_1, x_2, \lambda) &= 0 \\ \frac{\partial}{\partial \lambda} \mathcal{L}(x_1, x_2, \lambda) &= 0 \end{aligned} \tag{13}$$

In slightly more compact notation, (13) becomes:

$$\begin{aligned} \frac{\partial}{\partial \underline{x}} \mathcal{L}(\underline{x}, \lambda) &= 0 \\ \frac{\partial}{\partial \lambda} \mathcal{L}(\underline{x}, \lambda) &= 0 \end{aligned} \tag{14}$$

where we have used $\underline{x}=(x_1, x_2)$.

The conditions expressed by (14) are the general conditions for finding the optimal solution to an *n*-dimensional problem having a single equality constraint. The first equation in (14) is a vector equation, comprised of *n* scalar equations, with each scalar equation consisting of a derivative with respect to one of the *n* variables x_i.

The second equation in (14) just returns the equality constraint. Now let's see how this works in practice.

Example 3: Use our first-order conditions to solve the following specific instance of Problem 1.

$$\min\ f(x_1, x_2) = x_1^2 + x_2^2$$
$$\text{s.t. } h(x_1, x_2) = 2x_1 x_2 = 3$$

The Lagrangian function is:

$$\mathcal{L}(x_1, x_2) = x_1^2 + x_2^2 - \lambda(2x_1 x_2 - 3)$$

Applying first-order conditions, we obtain:

$$\frac{\partial}{\partial x_1}\mathcal{L}(x_1, x_2, \lambda) = 2x_1 - 2\lambda x_2 = 0 \qquad (15)$$

$$\frac{\partial}{\partial x_2}\mathcal{L}(x_1, x_2, \lambda) = 2x_2 - 2\lambda x_1 = 0 \qquad (16)$$

$$\frac{\partial}{\partial \lambda}\mathcal{L}(x_1, x_2, \lambda) = -(2x_1 x_2 - 3) = 0 \qquad (17)$$

This is a set of 3 equations in 3 unknowns, and so we may solve them. Unfortunately, these are not linear equations, and so we cannot set up $\underline{Ax}=\underline{b}$ and then solve by $\underline{x}=\underline{A}^{-1}\underline{b}$. In general, we must use a nonlinear solver (such as Newton) to solve nonlinear equations. But this case happens to be simple enough to use substitution. The details of the substitution procedure are not really important for our purposes, but I will give them here nonetheless, just for completeness…

From (15), $2x_1 = 2\lambda x_2$, and then substitution into (16) yields $2x_2 - 2\lambda^2 x_2 = 0 \Rightarrow 1-\lambda^2 = 0 \Rightarrow \lambda^2 = 1 \Rightarrow \lambda = \pm 1$ Choosing $\lambda = 1$ (choosing $\lambda = -1$ results in complex solutions), and since (15) gives $2x_1 = 2\lambda x_2$, we have that $2x_1 = 2x_2$, and substitution into (17) results in $\left(2x_2^2 - 3\right) = 0 \Rightarrow 2x_2^2 = 3 \Rightarrow x_2^2 = \frac{3}{2} \Rightarrow x_2 = \pm\sqrt{\frac{3}{2}}$, and since $2x_1 = 2x_2$, $x_1 = \pm\sqrt{\frac{3}{2}}$. From Fig. 3, we see that the desired solution[3] is $x_1 = x_2 = \sqrt{\frac{3}{2}} = 1.2247$, which results in a minimum value of $f(x_1, x_2)$ given by $f(x_1, x_2) = x_1^2 + x_2^2 = \left(\sqrt{\frac{3}{2}}\right)^2 + \left(\sqrt{\frac{3}{2}}\right)^2 = 3$, consistent with our observation from Fig. 3.

8. Multiple equality constraints

We can extend our *n*-dimensional slightly by considering that it may have multiple equality constraints. In this case, we have (3), repeated here for convenience.

$$\begin{aligned} &\min\ f(\underline{x}) \\ &\text{s.t.}\ \ \underline{\mathrm{h}}(\underline{x}) = \underline{c} \end{aligned} \tag{3}$$

[3] The negative values of x_1, x_2 originate from the fact that the equality constraint $2x_1x_2-3=0 \Rightarrow x_2=3/(2x_1)$, mathematically, may also have negative values of x_1, resulting in a curve in the lower left hand quadrant.

Consider that we have m equality constraints. Then we may apply the exact same approach as before, i.e., we formulate the Lagrangian and then apply first-order conditions, except in this case we will have m Lagrange multipliers, as follows:

$$\begin{aligned}\mathcal{L}(x_1, x_2, \lambda) = f(x_1, x_2) - \lambda_1\left(h_1(x_1, x_2) - c_1\right) \\ - \lambda_2\left(h_2(x_1, x_2) - c_2\right) - \ldots - \lambda_m\left(h_m(x_1, x_2) - c_m\right)\end{aligned} \qquad (18a)$$

or, alternatively

$$\begin{aligned}\mathcal{L}(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda_1\left(c_1 - h_1(x_1, x_2)\right) \\ + \lambda_2\left(c_2 - h_2(x_1, x_2)\right) + \ldots + \lambda_m\left(c_m - h_m(x_1, x_2)\right)\end{aligned} \qquad (18b)$$

The first order conditions are:

$$\begin{aligned}\frac{\partial}{\partial \underline{x}} \mathcal{L}(\underline{x}, \underline{\lambda}) = 0 \\ \frac{\partial}{\partial \underline{\lambda}} \mathcal{L}(\underline{x}, \underline{\lambda}) = 0\end{aligned} \qquad (19)$$

9. One inequality constraint

The general form of our problem when we have 1 inequality constraint is:

$$\begin{aligned} &\min \ f(\underline{x}) \\ &\text{s.t.} \ \ \underline{h}(\underline{x}) = \underline{c} \\ &\qquad g(\underline{x}) \geq b \end{aligned} \qquad (20)$$

An effective strategy for solving this problem is to first solve it by ignoring the inequality constraint, i.e., solve

$$\begin{aligned} &\min\ f(\underline{x}) \\ &\text{s.t.}\ \ \underline{h}(\underline{x}) = \underline{c} \end{aligned} \tag{21}$$

by writing our first-order conditions. Then check the solution against the inequality constraint. There are two possible outcomes at this point:

- If the inequality constraint is satisfied, then the problem is solved.
- If the inequality constraint is violated, then we know the inequality constraint must be *binding*. This means that the inequality constraint will be enforced with equality, i.e.,

$$g(\underline{x}) = b \tag{22}$$

If this is the case, then we include (22) as an equality constraint in our optimization problem, resulting in the following equality-constrained problem:

$$\begin{aligned} &\min\ f(\underline{x}) \\ &\text{s.t.}\ \ \underline{h}(\underline{x}) = \underline{c} \\ &\qquad g(\underline{x}) = b \end{aligned} \tag{23}$$

The Lagrangian for this problem is

$$\begin{aligned} \mathcal{L}(\underline{x}, \underline{\lambda}, \mu) = {} & f(\underline{x}) - \lambda_1\big(h_1(x_1, x_2) - c_1\big) - \lambda_2\big(h_2(x_1, x_2) - c_2\big) \\ & - \ldots - \lambda_m\big(h_m(x_1, x_2) - c_m\big) - \mu\big(g(\underline{x}) - b\big) \end{aligned} \tag{24a}$$

Alternatively,

$$\boldsymbol{\mathcal{L}}(\underline{x}, \underline{\lambda}, \mu) = f(\underline{x}) + \lambda_1\left(c_1 - h_1(x_1, x_2)\right) + \lambda_2\left(c_2 - h_2(x_1, x_2)\right) + \ldots + \lambda_m\left(c_m - h_m(x_1, x_2)\right) + \mu\left(b - g(\underline{x})\right) \tag{24b}$$

And the first-order conditions for solving this problem are:

$$\begin{aligned} \frac{\partial}{\partial \underline{x}} \boldsymbol{\mathcal{L}}(\underline{x}, \underline{\lambda}, \mu) &= 0 \\ \frac{\partial}{\partial \underline{\lambda}} \boldsymbol{\mathcal{L}}(\underline{x}, \underline{\lambda}, \mu) &= 0 \\ \frac{\partial}{\partial \mu} \boldsymbol{\mathcal{L}}(\underline{x}, \underline{\lambda}, \mu) &= 0 \end{aligned} \tag{25}$$

The procedure that we just described, where we first solved the problem without the inequality constraint, then tested for violation, and then resolved if a violation existed, is equivalently stated as

$$\begin{aligned} \frac{\partial}{\partial \underline{x}} \boldsymbol{\mathcal{L}}(\underline{x}, \underline{\lambda}) &= 0 \\ \frac{\partial}{\partial \underline{\lambda}} \boldsymbol{\mathcal{L}}(\underline{x}, \underline{\lambda}) &= 0 \\ \mu_k g(\underline{x}) &= 0 \end{aligned} \tag{26}$$

The last is called the complementary condition, and the overall conditions expressed as (26) are called the Kurash-Kuhn-Tucker (KKT) conditions.

10. Multiple inequality constraints

The method of the previous section for solving problems having a single equality constraint also generalizes to the case of a problem having multiple inequality constraints, except after each solution, one must check all of the remaining inequality constraints, and any that are binding are then brought in as equality constraints.

11. Terminology and basics of OR

Operations research (OR), or “optimization” is about decision-making. In light of this, we provide 2 basic definitions.

The *decision variables* are the variables in the problem, which, once known, determine the decision to be made.

The *objective function* is the function to be minimized or maximized. It is also sometimes known as the *cost function*.

The *constraints* are equality or inequality relations in terms of the decision variables which impose limitations or restrictions on what the solution may be, i.e., they *constrain* the solution.

Inequality constraints may be either *non-binding* or *binding*. A non-binding inequality constraint is one that does not influence the solution. A binding inequality constraint does restrict the solution, i.e., the objective function becomes "better" (greater if the problem is maximization or lesser if the problem is minimization) if a binding constraint is removed.

Optimization problems are often called *programs* or *programming* problems. Such terminology is not to be confused with use of the same terminology for a piece of source code (a program) or what you do when you write source code (programming). Use of the terminology here refers to an analytical statement of a decision problem. In fact, optimization problems are often referred to as *mathematical programs* and their solution procedures as *mathematical programming*. Such use of this terminology is indicated when one uses the term *linear programming* (LP), *nonlinear programming* (NLP), or *integer programming* (IP).

The general form of a mathematical programming problem is to find vector $\underline{x}$ in:

$$\begin{aligned} &\text{Min } f(\underline{x}) \\ &\qquad \text{subject to:} \qquad\qquad\qquad\qquad (1) \\ &\underline{g}(\underline{x}) \le \underline{b} \\ &\underline{h}(\underline{x}) = \underline{c} \end{aligned}$$

and: $\underline{x} \geq 0$

Here, f, $\underline{g}$, and $\underline{h}$ are given functions of the n decision variables $\underline{x}$. The condition $\underline{x} \geq \underline{0}$ can be satisfied by appropriate definition of decision variables.

The LaGrangian function of (1) is:

$$F\left(\underline{x}, \underline{\lambda}, \underline{\mu}\right) \equiv f\left(\underline{x}\right) - \underline{\lambda}^T \left[\underline{h}\left(\underline{x}\right) - \underline{c}\right] - \underline{\mu}^T \left[\underline{g}\left(\underline{x}\right) - \underline{b}\right] \qquad (2a)$$

or, alternatively,

$$F\left(\underline{x}, \underline{\lambda}, \underline{\mu}\right) \equiv f\left(\underline{x}\right) \underline{+ \lambda}^T \left[\underline{c} - \underline{h}\left(\underline{x}\right)\right] + \underline{\mu}^T \left[\underline{b} - \underline{g}\left(\underline{x}\right)\right] \qquad (2b)$$

where individual elements of $\underline{\lambda} = \left(\lambda_1, \lambda_2 \ldots, \lambda_m\right)$ and $\underline{\mu} = \left(\mu_1, \mu_2 \ldots, \mu_r\right)$ are called LaGrange multipliers.

The LaGrangian function is simply the summation of the objective function with the constraints. It is assumed that f, $\underline{h}$, and $\underline{g}$ are continuous and differentiable, that f is convex, and that the region in the space of decision variables defined by the inequality constraints is a convex region.

Given that $\underline{x}$ is a feasible point, the conditions for which the optimal solution occurs are:

$$\frac{\partial F}{\partial x_i} = 0 \qquad \forall\ i = 1, n \tag{3}$$

$$\frac{\partial F}{\partial \lambda_j} = 0 \qquad \forall\ j = 1, J \tag{4}$$

$$\mu_k\left[g_k\left(\underline{x}\right) - b_k\right] = 0 \qquad \forall\ k = 1, K \tag{5}$$

These conditions are known as the Karush-Kuhn-Tucker (KKT) conditions or, more simply, as the Kuhn-Tucker (KT) conditions. The KKT conditions state that for an optimal point

1) The derivatives of the LaGrangian with respect to all decision variables must be zero (3).
2) All equality constraints must be satisfied (4).
3) A multiplier μ_k cannot be zero when its corresponding constraint is binding (5).

Requirement 3, corresponding to eq. (5), is called the "complementary" condition. The complementary condition is important to understand. It says that if $\underline{x}$ occurs on the boundary of the k^{th} inequality constraint, then $g_k(\underline{x}) = b_k$. In this case eq. (5) allows μ_k to be non-zero. Once it is known that the k^{th} constraint is binding, then the k^{th} constraint can be moved to the vector of equality constraints; i.e. $g_k(\underline{x})$ can then be renamed as $h_{J+1}(\underline{x})$ and μ_k as λ_{J+1}, according to:

$$
\begin{aligned}
g_k(\underline{x}) &\rightarrow h_{J+1}(\underline{x}) \\
\mu_k &\rightarrow \lambda_{J+1}
\end{aligned}
\qquad (7)
$$

On the other hand, if the solution x does not occur on the boundary of the k^{th} inequality constraint, then (assuming $\underline{x}$ is an attainable point) $g_k(\underline{x}) - b_k < 0$. In this case, eq. (5) requires that $\mu_k = 0$ and the k^{th} constraint makes no contribution to the LaGrangian.

It is important to understand the significance of μ and λ. The optimal values of the LaGrangian Multipliers are in fact the rates of change of the optimum attainable objective value $f(\underline{x})$ with respect to changes in the right-hand-side elements of the constraints. Economists know these variables as shadow prices or marginal values. This information can be used not only to investigate changes to the original problem but also to accelerate repeat solutions. The marginal values λ_j or μ_k indicate how much the objective $f(\underline{x})$ would improve if a constraint b_j or c_k, respectively, were changed. One constraint often investigated for change is the maximum production of a plant.

[1] A. Wood and B. Wollenberg, "Power generation, operation, and control," second edition, Wiley, 1996.
[2] www.ebyte.it/library/docs/math09/AConvexInequality.html
[3] S. Boyd and L. Vandenberghe, "Convex optimization," Cambridge University Press, 2004.
[4] http://www.econ.iastate.edu/faculty/hallam/Mathematics/ConvexOpt.pdf
[5] Christoph Schnorr "Convex and Non-Convex Optimization," available at http://visiontrain.inrialpes.fr/pdf/LesHouches2006/ConvexTutorial.pdf.