Review of Optimization Basics

1.0 Introduction

At the end of our last discussion (see notes 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.

We saw that the method of optimizing generator costs that we have learned in previous coursework (e.g., EE 303), requires convex cost curves, and so to be guaranteed that our optimization method will result in a global minimum, we must use convex approximations for these cases, or else we have to apply so-called non-convex programming methods, which are significantly more complex and computational.

In these notes, we wish to review some basic convex programming concepts. It will be useful to do so as we lead into our basics of economics.

2.0 Problem Statement

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

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

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

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

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

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

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

3.0 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 1illustrates.

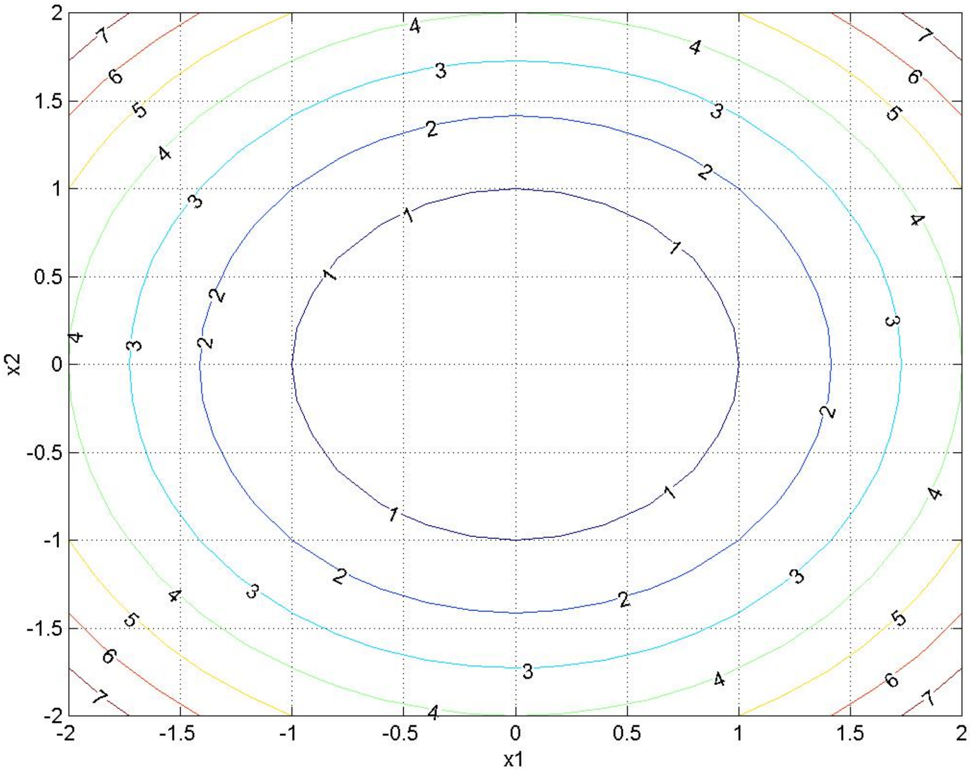

Fig. 1: Contour map for $f(x_1, x_2) = x_1^2 + x_2^2$

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.

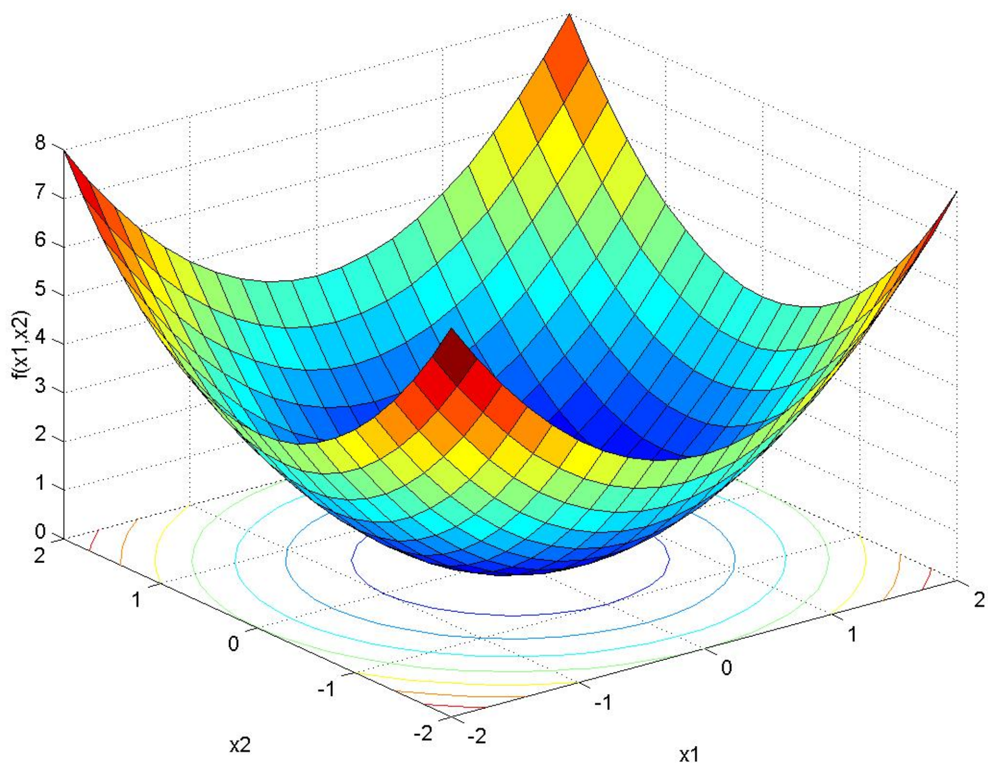

Fig. 2: 3-D illustration of $f(x_1,x_2) = x_1^2 + x_2^2$

4.0 Understanding the 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.

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

$$\min f(x_1, x_2) = x_1^2 + x_2^2$$
$$\text{s.t.} \quad h(x_1, x_2) = 2x_1x_2 = 3$$

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:

$$2x_1x_2 = 3 \Rightarrow x_2 = \frac{3}{2x_1}$$

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.

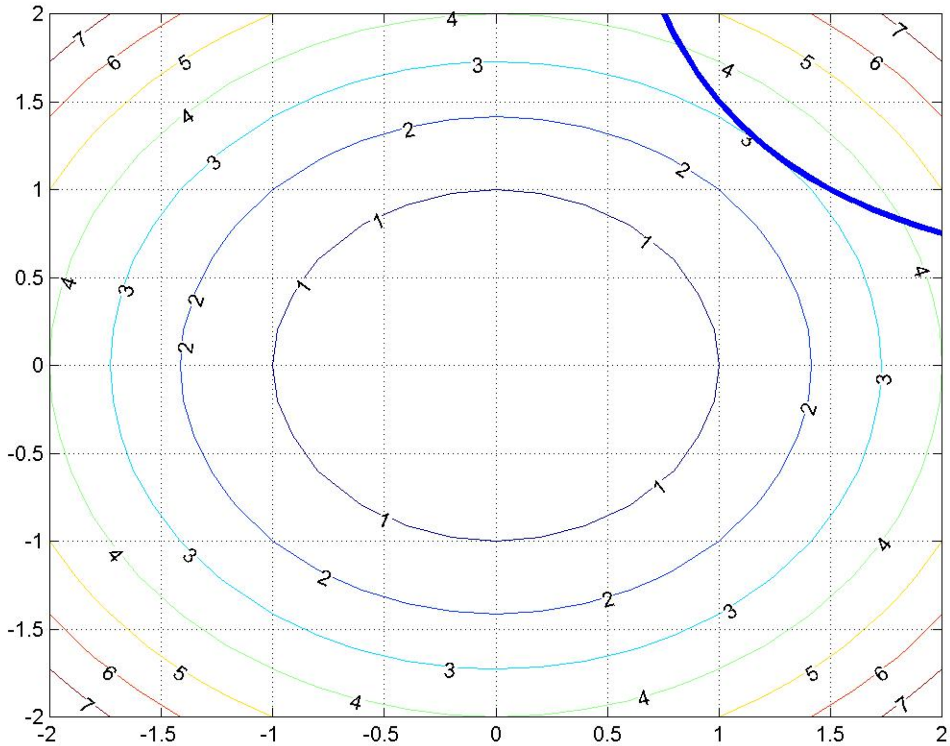

Fig. 3: Contour plots with equality constraint

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, however, let's look closely at Fig. 3, at the intersection of the equality constraint and the contour *f=3*. Notice that

- 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 to one another at the solution point.

The notion of tangency is equivalent to another one:

The normal vectors 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 $\underline{x}$, in which case, notationally, we say $\underline{x}_0=(x_1, x_2)_0$. Gradients have magnitude and angle.

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

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

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

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

or, 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 (5) as

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

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

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

Expressing (7) in terms of partial derivatives yields:

$$\begin{bmatrix} \dfrac{\partial}{\partial x_1}\Big(f(x_1,x_2)-\lambda\big(h(x_1,x_2)-c\big)\Big) \\ \dfrac{\partial}{\partial x_2}\Big(f(x_1,x_2)-\lambda\big(h(x_1,x_2)-c\big)\Big) \end{bmatrix}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad (8)$$

Notice in (7) and (8) that the “0”-subscript, which indicates “evaluation at the solution,” has been shifted to the outside of the brackets, indicating that the “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: $h(x_1, x_2)-c=0$. This is just our equality constraint. And so we see that we have three equations and three unknowns, and at least in principle, we can solve for our unknowns x_1, x_2, and λ. To summarize, the three equations are:

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

We make one more startling observation, and that is that the three equations are simply partial derivatives of the function $f(x_1,x_2)-\lambda\left(h(x_1,x_2)-c\right)$ 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:

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

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

$$F(x_1,x_2,\lambda)=f(x_1,x_2)-\lambda\left(h(x_1,x_2)-c\right) \qquad (11)$$

Then the first-order conditions for solving this problem are

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

or,

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

In slightly more compact notation, (13) becomes:

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

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.

$$\min f(x_1, x_2) = x_1^2 + x_2^2$$
$$\text{s.t.} \quad h(x_1,x_2) = 2x_1x_2 = 3$$

The Lagrangian function is:

$$F(x_1, x_2) = x_1^2 + x_2^2 - \lambda(2x_1x_2 - 3)$$

Applying first-order conditions, we obtain:

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

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

$$\frac{\partial}{\partial \lambda} F(x_1, x_2, \lambda) = -(2x_1x_2 - 3) = 0 \qquad (17)$$

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 $\underline{A}\underline{x}=\underline{b}$ 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 $\lambda = 1$, 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 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

$$f(x_1, x_2) = x_1^2 + x_2^2 = \left(\sqrt{\frac{3}{2}}\right)^2 + \left(\sqrt{\frac{3}{2}}\right)^2 = 3,$$

which is consistent with our observation from Fig. 3.

5.0 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.

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

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:

$$F(x_1, x_2, \lambda) = f(x_1, x_2) - \lambda_1\left(h_1(x_1, x_2) - c_1\right)$$
$$- \lambda_2\left(h_2(x_1, x_2) - c_2\right) - \ldots - \lambda_m\left(h_m(x_1, x_2) - c_m\right) \quad (18)$$

The first order conditions are:

$$\frac{\partial}{\partial \underline{x}} F(\underline{x}, \underline{\lambda}) = 0$$
$$\frac{\partial}{\partial \underline{\lambda}} F(\underline{x}, \underline{\lambda}) = 0 \quad (19)$$

6.0 One inequality constraint

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

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

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

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

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

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

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

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

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

$$\frac{\partial}{\partial \underline{x}} F(\underline{x}, \underline{\lambda}) = 0$$

$$\frac{\partial}{\partial \underline{\lambda}} F(\underline{x}, \underline{\lambda}) = 0$$

$$\mu_k h_k(\underline{x}) = 0$$

The last condition is called the complementary condition.

7.0 Multiple inequality constraints

This is for your homework assignment.