week02-02–lagrange

January 29, 2024

1 George McNinch Math 87 - Spring 2024

2 Week 2: Lagrange multpliers

(multivariable optimization, continued)

2.1 Some motivation

Let's recall our television manufacturing example for a moment. The model we solved was interesting but most likely unrealistic. A manufacturer, for instance, probably has a limited capacity and can only produce a certain amount of TVs in total per year.

Let's look at an example where that capacity is 10,000 TVs per year.

So given the *constraint*, $s + t \le 10,000$, how many of each model should they produce now??

Well, first notice that the *unconstrained optimum* with $s = 4,735$ and $t = 7,043$ – found *previously* – does not satisfy the constraint (since $s + t = 4735 + 7043 > 11700$).

Now let's recall that in the single variable case, trying to find a *constrained optimum* amounts to optimizing a function on a closed interval – so you proceed with the usual procedure for optimization but must also check boundary points (endpoints).

If we proceed in this fashion, we'd check the "boundary" conditions corresponding to $t = 0$ and $s = 0$.

Well, setting $t = 0$ we see that the profit function is given by

$$P(s,0) = 144s - 0.01s^2 - 400,000.$$

Since this is a function only of s we can use 1-dimensional optimization techniques:

$$\frac{\partial P}{\partial S}(S,0) = 144 - 0.02s = 0 \implies s = 7200.$$

On this boundary, we need to consider $(0,0)$ and $(10,000,0)$ (the boundary of the boundary…) as well as $(7200,0)$.

We can treat the boundary with $s = 0$ similarly.

But the boundary condition with $s + t = 0$ will be more of a pain.

The method of Lagrange multipliers gives a more systematic way to proceed!

2.2 Lagrange multipliers

Consider a function $f(x, y)$ of two variables. We are interested here in finding maximal or minimal values of f subject to a *constraint*. The sort of constraint we have in mind is a restriction on the possible pairs (x, y) – so we have a second function $g(x, y)$ and we want to maximize (or minimize) f subject to the condition that $g(x, y) = c$ for some fixed quantity c.

We introduce a "new" function – now of *three* variables - known as the **Lagrangian**. It is given by the formula

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda \cdot (g(x, y) - c)$$

(We use the Greek letter λ for our third variable in part because it plays a different role for us than the variables x, y).

We can calculate the *partial derivatives* of this Lagrangian; they are:

$$\frac{\partial \mathcal{L}}{\partial x} = \frac{\partial f}{\partial x} - \lambda \frac{\partial g}{\partial x}$$

$$\frac{\partial \mathcal{L}}{\partial y} = \frac{\partial f}{\partial y} - \lambda \frac{\partial g}{\partial y}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = -(g(x, y) - c)$$

If we seek critical points of the Lagrangian, we therefore must require

$$0 = \frac{\partial \mathcal{L}}{\partial x} = \frac{\partial f}{\partial x} - \lambda \frac{\partial g}{\partial x}$$

and similarly for y, so that

$$\frac{\partial f}{\partial x} = \lambda \frac{\partial g}{\partial x} \quad \text{and} \quad \frac{\partial f}{\partial y} = \lambda \frac{\partial g}{\partial y}.$$

Recall that the *gradient* ∇f of f is given by $\nabla f = \dfrac{\partial f}{\partial x}\mathbf{i} + \dfrac{\partial f}{\partial y}\mathbf{j}$ – or $\nabla f = \left(\dfrac{\partial f}{\partial x}, \dfrac{\partial f}{\partial y}\right)$ – so these conditions amount to:

$$\nabla f = \lambda \cdot \nabla \mathcal{L}$$

Moreover, we find that

$$0 = \frac{\partial \mathcal{L}}{\partial \lambda} = c - g(x, y).$$

Summarizing, we record

Proposition The condition that (x_0, y_0, λ_0) is a critical point of F is equivalent to two requirements:

(a). (x_0, y_0) must be on the level curve $g(x, y) = c$, and

(b). the gradient vectors must satisfy $\nabla f|_{(x_0,y_0)} = \lambda_0 \nabla g|_{(x_0,y_0)}$.

The reason for interest in the critical points of $\mathcal{L}$ is the following:

Proposition Optimal values for f along the level curve $g(x, y) = c$ will be found among the critical points of F.

Indeed, suppose (x_0, y_0) is a point on the level curve at which f takes its max (or min) value (on the level surface). We need to argue that the gradient vector $\nabla f|_{(x_0,y_0)}$ is "parallel" to the gradient vector $\nabla g|_{(x_0,y_0)}$ – i.e. that **(b)** above holds.

Well, we can of course write $\nabla f|_{(x_0,y_0)} = \mathbf{v} + \mu \nabla g|_{(x_0,y_0)}$ for a vector $\mathbf{v}$ perpendicular to $\nabla g|_{(x_0,y_0)}$ (and for some scalar μ). To confirm that **(b)** holds, we must argue that $\mathbf{v}$ is zero.

But if $\mathbf{v}$ is non-zero, then walking along the level curve $g(x, y) = c$ "in the direction of $\mathbf{v}$" will lead to points at which the function f will take larger values, contrary to the assumption that on the level curve, f has a maximum at (x_0, y_0). ■

2.3 Maxima (or minima) and an interpretation of λ

Haivng found the critical points of F, one can find the extrema for f simple by evaluating and comparing the results (*well, this works at least if there are just a finite list of critical points*).

If M^* denotes the max (or min) value of f subject to the constraint $g(x, y) = c$, we can view F^* as a function $F^*(c)$ of c. (Different values of c result in different extrema for f on the curve $g(x, y) = c$...)

We now claim that if (x_0, y_0, λ_0) is a critical point for F for which $F^* = f(x_0, y_0)$, then

$$\frac{dF^*}{dc} = \lambda_0.$$

We are first going to consider this observation in the case of an *application*. Then we'll explain why this observation is true in general.

2.4 Televisions, again

Let's return to the television manufacturing problem. We consider the constraint

$$g(s, t) = s + t = 10,000$$

i.e. "the manufacturer produces exactly 10,000 televisions".

Consider the *Lagrangian* function $\mathcal{L}(s, t, \lambda) = P(s, t) - \lambda(s + t - 10,000)$.

Looking for critical points of $\mathcal{L}$, we find that:

$$\begin{cases} \dfrac{\partial P}{\partial s} - \lambda \dfrac{\partial g}{\partial s} & = & 144 - 0.02s - 0.007t - \lambda & = 0 \\ \dfrac{\partial P}{\partial t} - \lambda \dfrac{\partial g}{\partial t} & = & 174 - 0.007s - 0.002t - \lambda & = 0 \\ g(s,t) - c & = & -10000 + s + t & = 0 \end{cases}$$

This leads to 3 linear equations in 3 unknowns, which we can easily solve bny hand. Or we can use `python` and `numpy` as follows:

[12]:
```
import numpy as np

## coefficient matrix
M=np.array([[0.02,.007,1],[.007,.02,1],[1,1,0]])

b=np.array([144,174,10000])

np.linalg.solve(M,b)
```

[12]:
```
array([3846.15384615, 6153.84615385,   24.        ])
```

we find that $s \approx 3846$, $t \approx 6154$ and $\lambda = 24$.

Now, remember that we want to maximize the profit function P subject to the constraint $s + t = 10,000$ where $s \geq 0$ and $t \geq 0$. On this "closed interval", the function P will assume a maximum and a minimum value. We've found the critical point of P on this "interval" – namely $(3846, 6154)$. The *endpoints* of the interval are $(0, 10000)$ and $(10000, 0)$.

Let's compare the values of P at these three points of interest:

[17]:
```
def p(s,t):
    return -400000 + 144*s + 174*t - 0.01*s**2 - 0.01*t**2 - .007*s*t

ss = [ 0,
       3846,
       10000
     ]
tt = [ 10000,
       6154,
       0
     ]

list(map(p,ss,tt))
```

[17]:
```
[340000.0, 532307.692, 40000.0]
```

This shows that the *maximum constrained profit* is

$$P^* = P(3846, 6154) \approx 532308.$$

Q: What choice of s, t give the *minimum* constrained profit?

(Could you have guessed that *a priori*?)

2.5 Shadow prices and an interpretation for the "multiplier" λ

Let's carry out *sensitivity analysis* on the value of the *constraint*.

Recall that our constraint is $g(s, t) = s + t = 10000$. So we instead set

$$g(s, t) = s + t = c.$$

In this case, we must instead solve the system of equations $Mx = b$ where

$$M = \begin{pmatrix} 0.02 & 0.007 & 1 \\ 0.007 & 0.02 & 1 \\ 1 & 1 & 0 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} 144 \\ 174 \\ c \end{pmatrix}$$

Since the vector b has an "unknown" coefficient, it isn't clear how to use `numpy`'s `linalg.solve` method. We can circumvent this by *inverting* the coefficient matrix.

[12]:
```
import numpy as np
import sympy as sp

import pprint as pp

## coefficient matrix
M=np.array([[0.02,0.007,1],[0.007,0.02,1],[1,1,0]])

## inverse of coefficient matrix
Mi = np.linalg.inv(M)

print(M)
print(Mi)
```

```
[[0.02  0.007 1.   ]
 [0.007 0.02  1.   ]
 [1.    1.    0.   ]]
[[ 3.84615385e+01 -3.84615385e+01  5.00000000e-01]
 [-3.84615385e+01  3.84615385e+01  5.00000000e-01]
 [ 5.00000000e-01  5.00000000e-01 -1.35000000e-02]]
```

We find the result by the computation $\mathbf{x} = M^{-1}\mathbf{b}$.

In order to represent the vector $\mathbf{b} = \begin{pmatrix} 144 \\ 174 \\ c \end{pmatrix}$, we first use `sympy` to introduce a `symbol` for the unknown `c`.

[14]:
```
c = sp.Symbol('c')
b=np.array([144,174,c])

#res=np.matmul(Mi,b)  # which we can also write as follows
res = Mi @ b

print(f"s = {res[0]}")
print(f"t = {res[1]}")
print(f"lambda = {res[2]}")
```

```
s = 0.5*c - 1153.84615384615
t = 0.5*c + 1153.84615384615
lambda = 159.0 - 0.0135*c
```

So we see that the solution has

$$s = \frac{c}{2} - 1153.85 \qquad t = \frac{c}{2} + 1153.85 \qquad \text{and} \qquad \lambda = 159 - 0.0135c.$$

Thus $\frac{ds}{dc} = \frac{dt}{dc} = \frac{1}{2}$. We can compute the sensitivity

$$S(s, c = 10000) \approx 1.3$$

of s as a function of c, and

$$S(t, c = 10000) \approx 0.8$$

pf t as a function of t. In particular, an increase in the maximum production level of TVs (i.e. an increase in c) will result in an increase in the optimal production levels s and t.

What about the sensitivity $S(P, c) = \frac{dP}{dc} \cdot \frac{c}{P}$? For this, we first need to compute $\frac{dP}{dc}$.

To compute $\frac{dP}{dc}$ we can use the several-variable chain rule

$$\frac{dP}{dc} = \frac{\partial P}{\partial s}\frac{ds}{dc} + \frac{\partial P}{\partial t}\frac{dt}{dc},$$

or just rewrite P as a function of c. Interestingly, observe that *at the critical point* on the curve $g(s,t) = c$, the partial derivatives of P satisfy

$$\frac{\partial P}{\partial s} = \lambda \frac{\partial g}{\partial s} = \lambda \quad \text{and} \quad \frac{\partial P}{\partial t} = \lambda \frac{\partial g}{\partial t} = \lambda$$

(since $g(s,t) = s + t$).

Since $\frac{ds}{dc} = \frac{dt}{dc} = 1/2$, conclude that

$$\frac{dP}{dc} = \lambda.$$

This confirms the formula we mentioned previously. In fact, as we pointed out, this formula holds in general. See the discussion at the end of this notebook, below.

We can now compute the *sensitivity* of the profit to the parameter c:

$$S(P, c = 10000)\frac{dP}{dc} \cdot \frac{c}{P} \approx 24 \cdot \frac{10000}{532308} \approx 0.45.$$

So a 1% increase in c yields a 0.45 increase in P.

Thus, the Lagrange multiplier,   actually has a *concrete* meaning here as a *shadow price*. If you are allowed to produce more TV's it tells you how much that change affects your profit. Therefore, if the cost of making the change in production level is less than the additional profit, you probably should go for it!

3 Interpreting λ, revisited

Lets return to our general setting: f is a function $f(x, y)$ of two variables that we want to optimize, and our constraint is given by an equation $g(x, y) = c$.

Recall that the Lagrangian $\mathcal{L}$ is given by

$$\mathcal{L} = f(x, y) - \lambda(g(x, y) - c).$$

Let's suppose that - as we've described above - for a given c, the maximum value of f is determined by the critical point $(x^*(c), y^*(c), \lambda^*(c))$ of the Lagrangian $\mathcal{L}$.

Now, let's view the Lagrangian as a function of the *four* variables

$$\mathcal{L}(x, y, \lambda, c).$$

We first note that

$$\frac{\partial \mathcal{L}}{\partial c} = \frac{\partial}{\partial c}\left[f(x, y) - \lambda(g(x, y) - c)\right] = \lambda.$$

Since $(x^*(c), y^*(c), \lambda^*(c))$ is a critical point of the Lagrangian, we know that $g(x^*(c), y^*(c)) - c = 0$, so that

$$F^*(c) = \mathcal{L}(x^*(c), y^*(c), \lambda^*(c), c).$$

Now use the multi-variate chain rule we see that

$$\frac{dF^*}{dc} = \frac{\partial \mathcal{L}}{\partial x}\frac{dx^*}{dc} + \frac{\partial \mathcal{L}}{\partial y}\frac{dy^*}{dc} + \frac{\partial \mathcal{L}}{\partial \lambda}\frac{d\lambda^*}{dc} + \frac{\partial \mathcal{L}}{\partial c}$$

We now notice that the partial derivatives $\frac{\partial \mathcal{L}}{\partial x}$, $\frac{\partial \mathcal{L}}{\partial y}$, $\frac{\partial \mathcal{L}}{\partial \lambda}$ are all zero when evaluated at $(x^*(c), y^*(c), \lambda^*(c)$.

This confirms that

$$\frac{dF^*}{dc} = \frac{\partial \mathcal{L}}{\partial c} = \lambda$$

as we asserted.

[ ]: