Optimization: Part 2

Abhinav Anand, IIMB

2018/06/14

Background

The problems seeking to maximize profits or minimize costs often feature nontrivial constraints which the optimal needs to satisfy. The solution must lie at the intersection of constraints (for equality constraints) or on one side of the constraint surface (for inequality constrints).

The problem in a general form is:

$$\max f(x) : x \in \mathbb{R}^n, x \geq 0$$

$$g_1(x_1, \ldots, x_n) \leq b_1, \ldots, g_k(x_1, \ldots, x_n) \leq b_k$$

$$h_1(x_1, \ldots, x_n) = c_1, \ldots, h_m(x_1, \ldots, x_n) = c_m$$

The objective function f is real valued, i.e., $f : \mathbb{R}^n \to \mathbb{R}$; $g(\cdot)$ are functional forms of the *inequality* constraints while $h(\cdot)$ are functional forms for the *equality* constraints.

Equality Constraints

Consider the case when $x = (x_1, x_2)$ and there is a single equality constraint $h_1(x) = c_1$.

$$\max f(x) : x \in \mathbb{R}^2, x \geq 0$$

$$h_1(x) = c_1$$

To make the illustration more concrete, consider $f(x) = x_1 x_2$ and $h_1(x) = c_1 : x_1 + x_2 = 5$.

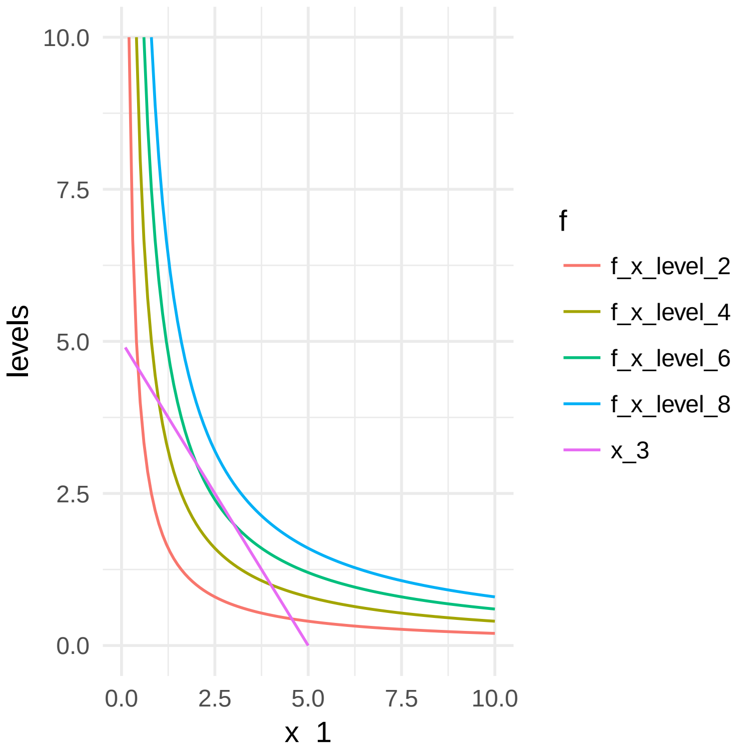
```r
x_1 <- seq(0.1, 10, 0.1)
x_2 <- seq(0.1, 10, 0.1)

f_x_level_2 <- 2/x_2 #level sets
f_x_level_4 <- 4/x_2
f_x_level_6 <- 6/x_2
f_x_level_8 <- 8/x_2

x_3 <- 5 - x_2 #constraint set

data_obj_l <- cbind(x_1,
                    f_x_level_2,
                    f_x_level_4,
                    f_x_level_6,
                    f_x_level_8,
                    x_3
                    ) %>%
  dplyr::as_tibble() %>%
  tidyr::gather(.,
                f_x_level_2:x_3,
                key = "f",
                value = "levels")

ggplot(data_obj_l, aes(x_1, levels, color = f)) +
  geom_line() +
  scale_y_continuous(limits = c(0, 10)) +
  scale_x_continuous(limits = c(0, 10)) +
  theme_minimal()
```



Geometrically we need to find the highest valued level set for $f(x) = x_1 x_2$ that satisfies $x_1 + x_2 = 5, x_1, x_2 \geq 0$. The key observation is the following: at the optimal, the levels sets and the constraint set must be tangent—just touching (intersecting) each other at exactly one point. (Why must this be so? What happens if the plots cross over? Can we improve the objective function then?)

The Lagrangian

If the curves are tangents to each other at the optimal point then it must be so that the tangents to the curves at the optimal are in the same direction.

The slope of the level set of f at x^* is:

$$-\frac{\frac{\partial f}{\partial x_1}(x^*)}{\frac{\partial f}{\partial x_2}(x^*)}$$

and that of the equality constraint is:

$$-\frac{\frac{\partial h}{\partial x_1}(x^*)}{\frac{\partial h}{\partial x_2}(x^*)}$$

Since they're equal

$$\frac{\frac{\partial f}{\partial x_1}(x^*)}{\frac{\partial f}{\partial x_2}(x^*)} = \frac{\frac{\partial h}{\partial x_1}(x^*)}{\frac{\partial h}{\partial x_2}(x^*)} = \lambda$$

This can be rearranged as:

$$\frac{\frac{\partial f}{\partial x_1}(x^*)}{\frac{\partial h}{\partial x_1}(x^*)} = \frac{\frac{\partial f}{\partial x_2}(x^*)}{\frac{\partial h}{\partial x_2}(x^*)} = \lambda$$

or,

$$\frac{\partial f}{\partial x_1}(x^*) - \lambda \frac{\partial h}{\partial x_1}(x^*) = 0$$
$$\frac{\partial f}{\partial x_2}(x^*) - \lambda \frac{\partial h}{\partial x_2}(x^*) = 0$$

There are three unknowns:(x_1^*, x_2^*, λ). There are two equations above, and there is a third equation—the constraint equation $h(x_1, x_2) = c_1$. Together, we can find x^* and λ.

The following function is formally referred to as the *Lagrangian*, and λ as the Lagrange multiplier:

$$\mathcal{L}(x_1, x_2, \lambda) := f(x_1, x_2) - \lambda \cdot (h(x_1, x_2) - c)$$

We consider the critical points of the Lagrangian: $\frac{\partial \mathcal{L}}{\partial x_1}(x^*), \frac{\partial \mathcal{L}}{\partial x_2}(x^*), \frac{\partial \mathcal{L}}{\partial \lambda}(x^*) = 0$.

Essentially by forming the Lagrangian, we are transforming a constrained optimization program featuring the objective function $f(\cdot)$ into an *unconstrained* optimization program featuring the Lagrangian $\mathcal{L}(\cdot)$. However, there is an extra variable λ, the Lagrange multiplier, that is introduced in the new program.

Constraint Qualification

In order for the slopes to be well-defined, $\frac{\partial h}{\partial x_1}(x^*) \neq 0$ and $\frac{\partial h}{\partial x_2}(x^*) \neq 0$. Since this is a restriction on the constraint set, it's called a *constraint qualification*.

Theorem: For the function $f : \mathbb{R}^n \to \mathbb{R}, f \in C^1$, if $x^* \in \mathbb{R}^n$ is a solution of

$$\max f(x) : x \geq 0, h(x) = c_1$$

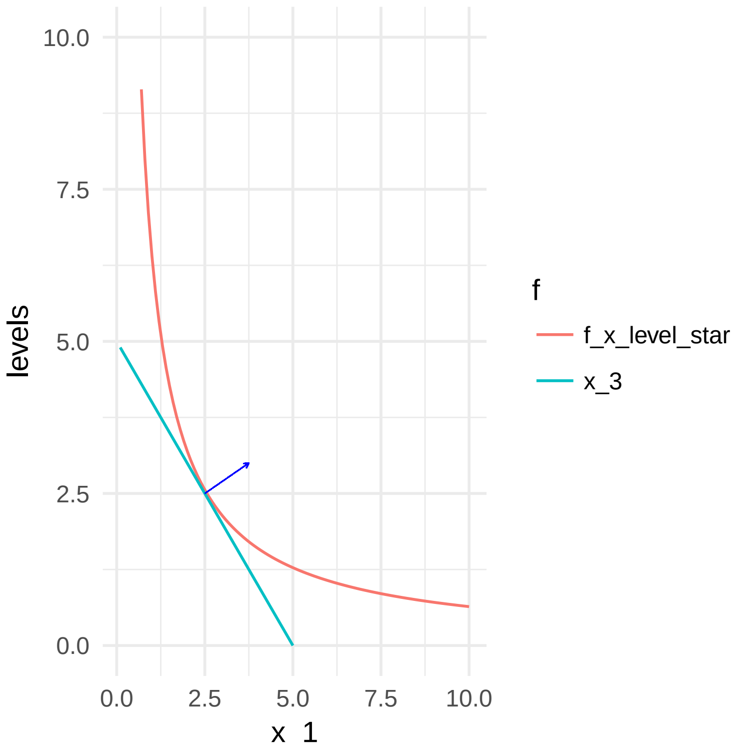
and x^* is *not* a critical point of h; then, there is a real number λ^* such that (x^*, λ^*) is a critical point of $\mathcal{L} = f(x) - \lambda \cdot (h(x) - c)$.

Level-Set Gradients and Lagrangians

```
f_x_level_star <- 6.4/x_2 #set manually

data_plot_grad <- cbind(x_1,
                        f_x_level_star,
                        x_3
                        ) %>%
  dplyr::as_tibble() %>%
  tidyr::gather(.,
                c(f_x_level_star, x_3),
                key = 'f',
                value = 'levels'
                )

ggplot(data_plot_grad, aes(x_1, levels, color = f)) +
  geom_line() +
  scale_y_continuous(limits = c(0, 10)) +
  geom_segment(aes(x = 2.5,
                   y = 2.5,
                   xend = 3.75, #set manually
                   yend = 3
                   ),
               color = "blue",
               arrow = arrow(length = unit(0.015, "npc")),
               size = 0.2
               ) +
  theme_minimal()
```



The gradient of f and h are respectively $[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}]$ and $[\frac{\partial h}{\partial x_1}, \frac{\partial f}{\partial x_2}]$. These point to the directions of maximum change and are orthogonal to the level sets of f, h.

Since the level sets and the constraints are tangent at the optimal implies that their respective gradients—orthogonal to the tangent—must again point in the same direction.

$$[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}] = \lambda[\frac{\partial h}{\partial x_1}, \frac{\partial h}{\partial x_1}]$$

This yields exactly the Lagrangian function for the optimization program.

Several Equality Constraints

Consider the following variation on the maximization problem where there are several equality constraints now:

$$\max f(x) : x \in \mathbb{R}^n, x \geq 0$$

$$C_h = \{h_1(x) = c_1, \ldots, h_m(x) = c_m\}$$

The generalization from one to many equality constraints is straightforward.

Constraint Qualification

In the case of one constraint, the qualification is:

$$[\frac{\partial h}{\partial x_1}(x^*), \ldots, \frac{\partial h}{\partial x_n}(x^*)] \neq (0, \ldots, 0)$$

Likewise, in the case of m equality constraints: $\{h_1(x) = c_1, \ldots, h_m(x) = c_m\}$, their *Jacobian* (first derivative matrix) must be *invertible* at the critical point x^*.

$$Dh(x^*) = \begin{bmatrix} \frac{\partial h_1}{\partial x_1}(x^*), \ldots, \frac{\partial h_1}{\partial x_n}(x^*) \\ \frac{\partial h_2}{\partial x_1}(x^*), \ldots, \frac{\partial h_2}{\partial x_n}(x^*) \\ \vdots \\ \frac{\partial h_m}{\partial x_1}(x^*), \ldots, \frac{\partial h_m}{\partial x_n}(x^*) \end{bmatrix}$$

For the constraint Jacobian at the critical point required to be invertible implies that the matrix above must have *full rank*, further equivalent to the condition that the determinant be non-zero at the critical point. More formally, it is said that $(h_1, \ldots, h_m)$ satisfy the *non-degenrate constraint qualification* (NCDQ) at x^* if matrix $Dh(x^*)$ is invertible at x^* (has full rank).

Theorem: Given $f, h_1, \ldots, h_m : \mathbb{R}^n \to \mathbb{R} \in C^1$, where f is the objective function and $h_i = c_i$ are equality constraints which x must satisfy; and that h_i satisfiy the NDCQ condition; then there are $\lambda^* = (\lambda_1^*, \ldots, \lambda_m^*)$ such that (x^*, λ^*) is the critical point of the Lagrangian $\mathcal{L}(x, \lambda)$.

Hence

$$\mathcal{L}(x, \lambda) := f(x) - \lambda_1 \cdot (h_1(x) - c_1) + \ldots + \lambda_m \cdot (h_m(x) - c_m)$$

and

$$\frac{\partial \mathcal{L}}{\partial x_1}(x^*, \lambda^*) = 0, \ldots, \frac{\partial \mathcal{L}}{\partial x_n}(x^*, \lambda^*) = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda_1}(x^*, \lambda^*) = 0, \ldots, \frac{\partial \mathcal{L}}{\partial \lambda_m}(x^*, \lambda^*) = 0$$

In total we get $n + m$ equations with $n + m$ unknowns: $(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_m)$.

Inequality Constraints

Consider now a maximization program in $\mathbb{R}^2$ with one inequality constraint:

$$\max f(x_1, x_2), (x_1, x_2) \geq 0, g(x_1, x_2) \leq b$$