Constrained Optimization

Guillaume Rabault

This post collects constrained optimzation results for reference.

Kuhn-Tucker

The presentation below is an adaptation of *Functional Analysis, Calculus of Variations and Optimal Control*, by Francis Clarke, Springer Verlag 2013 (Chapter 9).

We start with the following data:

- S is a convex set of $\mathbb{R}^N$
- $f : S \to \mathbb{R}$
- f is concave
- $g_i : S \to \mathbb{R}$, $i = 1, 2, \ldots, m$
- each g_i is convex
- $h_i : S \to \mathbb{R}$, $i = 1, 2, \ldots, n$
- each h_i is affine

Program P:

$$\max_x f(x)$$

$$\text{s.t.} :$$

$$x \in S,$$

$$h(x) = 0,$$

$$g(x) \leq 0.$$

Theorem (Kuhn-Tucker): *Let x_* be a solution of P. Then there exists $(\eta_*, \gamma_*, \lambda_*) \in \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}^n$ satisfying (non triviality):*

$$(\eta_*, \gamma_*, \lambda_*) \neq 0,$$

as well as (positivity and complementary slackness):

$$\eta_* = 0 \text{ or } 1, \ \gamma_* \geq 0, \ \langle \gamma_*, g(x_*) \rangle = 0,$$

and (maximization condition), $\forall x \in S$:

$$\eta_* f(x) - \langle \gamma_*, g(x) \rangle - \langle \lambda_*, h(x) \rangle \leq \eta_* f(x_*) - \langle \gamma_*, g(x_*) \rangle - \langle \lambda_*, h(x_*) \rangle = \eta_* f(x_*).$$

∎

The proof of this proposition entails the analysis of the convex cone:

$$C = \{(f(x) - \delta, g(x) + \Delta, h(x)) : \delta \geq 0, \Delta \geq 0, x \in S\},$$

around $z_* = (f(x_*), 0, 0)$, which must lie on its boundary. One then uses the fact that there must exist a vector at z_* that points towards the exterior of C and such that the scalar product of this vector with any vector $z - z_*$ (where z is in C) is negative.

First order condition

In the above theorem, the usual stationarity condition (first order condition) is replaced by a maximization condition. It is clear that if x_* is in the interior of S and that f and g are differentiable at x_*, the maximization condition implies:

$$\eta_* \nabla f(x) - \langle \gamma_*, \nabla g(x) \rangle - \langle \lambda_*, \nabla h(x) \rangle = 0.$$

Sufficiency

When $\eta_* = 1$ (the normal case), the maximization condition reads:

$$f(x) - \langle \gamma_*, g(x) \rangle - \langle \lambda_*, h(x) \rangle \leq f(x_*), \forall x \in S.$$

Since $\gamma_* \geq 0$, we have $\langle \gamma_*, g(x) \rangle \leq 0$ whenever x satisfies the inequality constraint $g(x) \leq 0$. In addition, under the set of constraints, $h(x) = 0$. We can thus conclude that:
$$f(x) \leq f(x_*), \forall x \in S, \ g(x) \leq 0, \ h(x) = 0,$$

and we can conclude that x_* is optimum, i.e. the conditions in the theorem are also sufficient.

Slater condition

Theorem (Slater conditions): *Assume that there exists a strictly admissible point x_0 for P, i.e.*:

$$x_0 \in \text{int} S, \ g(x_0) < 0, \ h(x_0) = 0,$$

and the affine functions of the equality constraints are independent. Then the multiplier of the Kuhn-Tucker Theorem satisfies $\eta_* = 1$.

∎

Saddle point property

Assume x_* and $(1, \gamma_*, \lambda_*)$ are the multiplier of problem P. The maximization condition in the Kuhn-Tucker theorem reads:

$$f(x) - \langle \gamma_*, g(x) \rangle - \langle \lambda_*, h(x) \rangle \leq f(x_*) - \langle \gamma_*, g(x_*) \rangle - \langle \lambda_*, h(x_*) \rangle = f(x_*), \forall x \in S.$$

Now, since $\gamma_* \geq 0$ and $g(x_*) \leq 0$, we have $\langle \gamma, g(x_*) \rangle \geq 0$ for all $\gamma \geq 0$ with equality for γ_* (complementary slackness). We also have $\langle \lambda, h(x_*) \rangle = 0$. As a result we have:

$$f(x) - \langle \gamma_*, g(x) \rangle - \langle \lambda_*, h(x) \rangle \leq f(x_*) \leq f(x_*) - \langle \gamma, g(x_*) \rangle - \langle \lambda, h(x_*) \rangle,$$

$$\forall x \in S, \gamma \geq 0, \lambda.$$

This is the famous saddle point property for the Lagrangian of the constrained optimization problem.

Multipliers as derivatives of the value function

One can consider the following maximization problem, noted $P(\alpha, \beta)$:

$$\max_x f(x)$$

$$\text{s.t. :}$$

$$x \in S,$$

$$h(x) = \beta,$$

$$g(x) \leq \alpha.$$

Theorem: *Assuming there is a solution x_* to problem P where the Slater condition is satisfied. Define $V(\alpha, \beta)$ as the value function of $P(\alpha, \beta)$. Then, $V(\cdot, \cdot)$ is concave with values in $[-\infty, +\infty)$. The vector $(1, \gamma, \lambda)$ is a multiplier associated to x_* if and only if (γ, λ) is such that:*

$$V(\alpha, \beta) \leq V(0, 0) + \langle \gamma, \alpha \rangle + \langle \lambda, \beta \rangle.$$

∎

The vector (γ, λ) is called the superdifferential of the concave value function at $(0, 0)$. If $V(\cdot, \cdot)$ is differentiable, γ is the partial derivative with respect to α and λ is the partial derivative with respect to β.

Remark

The most confusing aspect in the multiplier rule is perhaps the choice of the sign of the multiplier in the case of inequality constraints. In other words, how should we choose to formulate the Lagrangian? The logic is as follows. When the inequality constraint is of the form $g(x) \leq 0$ and the problem is a maximization one, the multiplier rule should be that at the optimum, increasing f should require increasing g. Gradients therefore have to be colinear (up to the adjustment for the equality constraints) with a positive constant of proportionality. In the expression below:

$$\nabla f(x) = \langle \gamma_*, \nabla g(x) \rangle + \langle \lambda_*, \nabla h(x) \rangle,$$

γ should be positive. This pins down the form of the Lagrangian. Should the constraint be $g(x) \geq 0$, the sign of γ would have to be reversed.

Envelope Theorem

The penultimate section gives the derivatives of the value function with respect to the parameters of the constraint. One can understand this rule adapting an argument usually coined the envelope theorem. The envelope theorem concerns the maximization problem:

$$\max_x f(x,t)$$

$$\text{s.t.} :$$

$$x \in S,$$

where t is for instance a real parameter, S is a convex subset of $\mathbb{R}$ (say) and f is assumed continuously differentiable in both variables. Whenever this problem has an interior solution (i.e. in the interior of S) $x^*(t)$ for t in an interval $I =]t_0, t_1[$, then, the value function attached to this problem satisfies (assuming this value function is differentiable)[1]:

$$V'(t) = \frac{\partial f}{\partial t}(x^*(t), t),$$

in this interval. In other words, the derivative of the value function can be computed as if the command was fixed at $x^*(t)$ in the marginal experiment. Indeed, we have:

$$\frac{\partial f}{\partial x}(x^*(t), t) = 0,$$

so that this term drops out in the variation of f:

$$dV = \frac{\partial f}{\partial x}(x^*(t), t)dx + \frac{\partial f}{\partial t}(x^*(t), t)dt.$$

---

[1] For a rigorous discussion of envelope theorems, you might want to check Milgrom-Segal, 2002, 'Envelope Theorems for Arbitrary Choice Sets', *Econometrica* 70 (2):583-601.

The envelope theorem derives its name from the geometric situation. For each t, the function $f(\cdot, t)$ corresponds to a curve $\mathcal{C}_t$. The curve corresponding to $V(\cdot)$ is the upper envelope of the $\mathcal{C}_t$'s. Now the tangent to the upper envelope at point t is equal to the tangent to the supporting curve $\mathcal{C}_t$ at this same point.

Applied to the constrained optimization context, this means that we can evaluate the change in the value function when constraints change by choosing a second best candidate to the choice of x. This second best should make full use of the relaxation of the constraints however. The reasoning is sketched below.

Let $x^*(\alpha, \beta)$ be the optimal solution as a function of parameters. Consider a small change from $(0, 0)$ to (α, β). We can decompose the change in the value function as follows. We move the choice of x from $x^*(0,0)$ to $\hat{x}(\alpha, \beta)$ where $\hat{x}(\alpha, \beta)$ is built to keep saturated constraints satisfied for the new parameters (α, β) (I write the change in g as if all inequality constraints were saturated; one should in fact keep track of the indices of saturated constraints but the argument goes through unchanged, using the slackness condition in the last step below):

$$\langle \nabla_{x^*(0,0)} g, \hat{x}(\alpha, \beta) - x^*(0,0) \rangle = \alpha,$$

$$\langle \nabla_{x^*(0,0)} h, \hat{x}(\alpha, \beta) - x^*(0,0) \rangle = \beta.$$

We then move from $\hat{x}(\alpha, \beta)$ to $x^*(\alpha, \beta)$. Both points $\hat{x}(\alpha, \beta)$ and $x^*(\alpha, \beta)$ satisfy the constraints $g(x) \leq \alpha$ and $h(x) = \beta$, and given that $x^*(\alpha, \beta)$ is optimal under the constraint, the change in the value of f along this move is negligible. The change in the value function can thus be computed as:

$$\langle \nabla_{x^*(0,0)} f, \hat{x}(\alpha, \beta) - x^*(0,0) \rangle.$$

Using the multiplier rule to replace $\nabla_{x^*(0,0)} f$ by $\langle \gamma_*, \nabla g(x) \rangle + \langle \lambda_*, \nabla h(x) \rangle$, this leads to (after some easy calculations):

$$\langle \nabla_{x^*(0,0)} f, \hat{x}(\alpha, \beta) - x^*(0,0) \rangle = \langle \gamma, \alpha \rangle + \langle \lambda, \beta \rangle.$$

Links

- Link to pdf