‘MATH+ECON+CODE’ MASTERCLASS ON MATCHING MODELS, OPTIMAL TRANSPORT AND APPLICATIONS

Alfred Galichon (New York University)

Spring 2019
Wednesday: “Optimal transport II”
Block 8. A short tutorial on convex analysis

- A short tutorial on convex analysis

- [OTME], Ch. 6
- Rockafellar (1970). *Convex analysis*. Princeton.

Section 1

Theory

- Assume that P and Q have a convex support with nonempty interior. Recall that if a dual minimizer (u, v) exists, u and v are related by

$$v(y) = \max_{x \in \mathbb{R}^d} \left\{ x^\intercal y - u(x) \right\} \tag{1}$$

$$u(x) = \max_{y \in \mathbb{R}^d} \left\{ x^\intercal y - v(y) \right\} \tag{2}$$

  (we can always assign the value $+\infty$ to u outside of the support of P and same for v).

- This expression is a fundamental tool in convex analysis: it is called the *Legendre-Fenchel transform*, which is defined in general by:

DEFINITION

The Legendre-Fenchel transform of u is defined by

$$u^*(y) = \sup_{x \in \mathbb{R}^d} \left\{ x^\intercal y - u(x) \right\}. \tag{3}$$

PROPOSITION

The following holds:
(i) u^* *is convex.*
(ii) $u_1 \leq u_2$ *implies* $u_1^* \geq u_2^*$*.*
(iii) (Fenchel's inequality): $u(x) + u^*(y) \geq x^\intercal y$*.*
(iv) $u^{**} \leq u$ *with equality iff* u *is convex.*

As an immediate corollary of (iv), we get the fundamental result:

PROPOSITION

If u *is convex, then* $u = (u^*)^*$*. The converse holds true.*

EXAMPLE

One has:
(i) For $u(x) = |x|^2/2$, one gets $u^*(y) = |y|^2/2$.
(ii) For $u(x) = \sum_i \lambda_i x_i^2/2$, $\lambda_i > 0$, one gets $u^*(y) = \sum_i \lambda_i^{-1} y_i^2/2$.
(iii) The entropy function

$$u(x) = \begin{cases} \sum_{i=1}^d x_i \ln x_i \text{ for } x \geq 0,\ \sum_{i=1}^d x_i = 1 \\ +\infty \text{ otherwise} \end{cases}$$

has a Legendre transform which is the log-partition function, a.k.a. logit function

$$u^*(y) = \ln\left(\sum_{i=1}^d e^{y_i}\right).$$

We now restate the demand sets of workers and firms in terms of subdifferentials of convex functions. For this, let us recall the basic economic interpretation of relations (1)-(2), which we had previously spelled out: Expression (1) captures the problem of a firm of type y, which hires a worker x who offers the best trade-off between production if hired by y (that is $\Phi(x, y) = x^{\intercal} y$) and wage $u(x)$. Thus, firm y will be willing to match with any worker whithin the set of maximizers of (1), while worker x will be willing to match with any firm whithin the set of maximizers of (2). The set of maximizers of (1) and of (2) are called *subdifferentials* of v and u,

DEFINITION

Let $u : \mathbb{R}^d \to \mathbb{R}$. The subdifferential of u at x, denoted $\partial u(x)$, is the set of $y \in \mathbb{R}^d$ such that $\forall \tilde{x} \in \mathbb{R}^d$, $u(\tilde{x}) \geq u(x) + y^\intercal (\tilde{x} - x)$.

- The definition does *not* require u to be convex; however, if u is convex, Definition 5 immediately implies that

$$\partial u(x) = \arg\max_y \left\{ x^\intercal y - u^*(y) \right\}, \tag{4}$$

  hence the subdifferential of a convex function is always nonempty (while the subdifferential of a non-convex function can be empty in general).
- When u is differentiable and convex, then

$$\partial u(x) = \left\{ \nabla u(x) \right\}.$$

EXAMPLE

When $u(x) = |x|$, one has $\partial u(x) = \{-1\}$ if $x < 0$, $\{+1\}$ if $x > 0$, and $[-1, +1]$ if $x = 0$.

It also follows that if u is a convex function, the following statements are equivalent:

$$\text{(i) } u(x) + u^{*}(y) = x^{\intercal} y \tag{5}$$

$$\text{(ii) } y \in \partial u(x) \tag{6}$$

$$\text{(iii) } x \in \partial u^{*}(y). \tag{7}$$

Going back to our worker-firm example, this has a straightforward economic interpretation. If worker x chooses firm y, then y maximizes $x^{\intercal}\tilde{y} - u^{*}(\tilde{y})$ over $\tilde{y}$, thus $y \in \partial u(x)$. This means that while worker x's equilibrium wage $u(x)$ is in general greater or equal than the value $x^{\intercal}y - u^{*}(y)$ she can extract from firm y, those two values necessarily coincide if x and y are willing to match, in which case $u(x) + u^{*}(y) = x^{\intercal}y$.

These considerations allow us to relate the solutions to the primal and dual problems. Recall that in the finite-dimensional case, where the primal and the dual problems are related by a complementary slackness condition. In the present case, let $(X, Y) \sim \pi$ be a solution to the primal problem, and (u, u^*) be a solution to the dual problem. Then almost surely X and Y are willing to match, which, by the previous discussion, implies that

$$u(X) + u^*(Y) = X^\intercal Y, \tag{8}$$

or equivalently $Y \in \partial u(X)$ or in turn $X \in \partial u^*(Y)$. In other words, the support of π is included in the set $\{(x, y) : u(x) + u^*(y) = x^\intercal y\}$. This condition appears as the correct generalization of the complementary slackness condition in the finite-dimensional case. Without surprise, taking the expectation with respect to π of equality (8) yields the equality between the value of the dual problem on the left-hand side, and the value of the primal problem on the right-hand side.

More can be said when u is differentiable at x. In that case, it is not hard to show that $\partial u(x) = \{\nabla u(x)\}$, i.e. contains only one point, which is $\nabla u(x) = (\partial u(x) / \partial x_i)_i$, the vector of partial derivatives of u, or gradient of u. Similarly, if u^* is differentiable at y, then $\partial u^*(y) = \{\nabla u^*(y)\}$. Hence, if u and v are differentiable, then the equivalence between (6) and (7) implies that $y = \nabla u(x)$ if and only if $x = \nabla u^*(x)$, that is

$$(\nabla u)^{-1} = \nabla u^*. \tag{9}$$

Alternatively, relation (9) can be seen as a duality between first-order conditions and the envelope theorem. First order conditions in the firm's problem (1) implies that if worker x is chosen by firm y, then $\nabla u(x) = y$, but the envelope theorem implies that the gradient in y of the firm's indirect profit $u^*(y)$ is given by $\nabla u^*(y) = x$, where x is chosen by y. Thus the first-order conditions and the envelope theorem are "conjugate" in the sense of convex analysis.

EXAMPLE

When $u(x) = \sum_i \lambda_i x_i^2 / 2$, $\lambda_i > 0$, recall that $u^*(y) = \sum_i \lambda_i^{-1} y_i^2 / 2$. Define $\Lambda = diag(\lambda)$. One has $\nabla u(x) = \Lambda x$ and $\nabla u^*(y) = \Lambda^{-1} y$.

Assume both u and u^* are stricly convex and differentiable. Then it can be show that their Hessians are invertible at all points, and that if $y = \nabla u\left(x\right)$, then

$$D^2 u^*\left(y\right) = \left(D^2 u\left(x\right)\right)^{-1}.$$

This can be obtained by differentiating the relationship $\nabla u^*\left(y\right) = \left(\nabla u\right)^{-1}\left(y\right)$.

Section 2

EXERCISES

EXERCISE

Compute the Legendre-Fenchel transforms of the following functions:
(i) $u(x) = x^{\intercal}\Sigma x/2$, *where* Σ *is a positive definite matrix, one has* $u^*(y) = y^{\intercal}\Sigma^{-1}y/2$.
(ii) Let $p > 1$ *and* $u(x) = \frac{1}{p}\|x\|^p$, *where* $\|.\|$ *is the Euclidean norm. Then* $u^*(y) = \frac{1}{q}\|y\|^q$, *where* $q > 1$ *such that* $1/p + 1/q = 1$.
(iii) $u(x) = 1\{x \in [0,1]\}$.

EXERCISE

Give the subdifferentials of the following functions from $\mathbb{R}$ *to* $\mathbb{R}$*:*
(a) $u(x) = \max(x, 0)$.
(b) $u(x) = \max(f(x), g(x))$, *where both* f *and* g *are convex and differentiable.*
(c) $u(x) = \max_{1\leq i\leq n}\{a_i x + b_i\}$, *where* $a_1 < a_2 < \ldots < a_n$.
(d) $u(x) = -x^2$.

Consider the entropy function

$$u(x) = \begin{cases} \sum_{i=1}^{d} x_i \ln x_i \text{ for } x \geq 0,\ \sum_{i=1}^{d} x_i = 1 \\ +\infty \text{ otherwise} \end{cases}.$$

As it is defined on the simplex, it is not a differentiable function from $\mathbb{R}^d$ to $\mathbb{R}$. Instead, let us take $x_d = 1 - \sum_{i=1}^{d-1} x_i$, and let us view u as a function $\tilde{u}$ from $\mathbb{R}^{d-1}$ to $\mathbb{R}$. We define

$$\tilde{u}(x) = \sum_{i=1}^{d-1} x_i \ln x_i + \left(1 - \sum_{i=1}^{d-1} x_i\right) \ln \left(1 - \sum_{i=1}^{d-1} x_i\right)$$

if $x \geq 0$, $\sum_{i=1}^{d-1} x_i \leq 1$, $\tilde{u}(x) = +\infty$ otherwise.

EXERCISE

Show that:
(a) The Legendre transform of $\tilde{u}$ is a function of $\mathbb{R}^{d-1}$ to $\mathbb{R}$ given by

$$\tilde{u}^* (y) = \ln \left(\sum_{i=1}^{d-1} e^{y_i} + 1 \right).$$

(b) The gradient of $\tilde{u}$ is a vector in $\mathbb{R}^{d-1}$ given by

$$\nabla \tilde{u} (x) = \left(\ln \left(\frac{x_i}{1 - \sum_{i=1}^{d-1} x_i} \right) \right)_{1 \leq i \leq d-1}$$

(c) The gradient of $\tilde{u}^*$ is a vector in $\mathbb{R}^{d-1}$ given by

$$\nabla \tilde{u}^* (y) = \left(\frac{e^{y_i}}{\sum_{i=1}^{d-1} e^{y_i} + 1} \right)_{1 \leq i \leq d-1}$$

(d) Compute $D^2 \tilde{u}$ and $D^2 \tilde{u}^*$.