Discrete choice prox-functions on the simplex

Vladimir Shikhman

Chemnitz University of Technology, Germany

in collaboration with David Müller and Yurii Nesterov

Guiding principles I have learnt from Yurii at CORE

- Everything in the world is **convex**, but we need to find a right perspective to see this.
- A good **economic** model should be simple enough to be mathematically tractable, but involved enough to remain interesting (not only to mathematicians).

$$\Downarrow$$

ALGORITHMIC APPROACH TO MICROECONOMICS:

How to explain the behavior of economic agents by algorithms from convex optimization?

PROX-FUNCTIONS

d is a prox-function on a closed convex set $Q \subset \mathbb{R}^n$ if

(1) d is **continuous** with the domain containing Q.

(2) d is β-**strongly convex** on Q with respect to a norm $\|\cdot\|$, i. e. there exists a constant $\beta > 0$ such that for all $x, y \in Q$ and $\alpha \in [0,1]$ it holds:

$$d(\alpha x + (1-\alpha)y) \leq \alpha d(x) + (1-\alpha)d(y) - \frac{\beta}{2}\alpha(1-\alpha)\|x-y\|^2.$$

(3) The computation of the **convex conjugate**

$$d^*(s) = \max_{x \in Q} \langle s, x \rangle - d(x)$$

is simple, i. e. the unique maximizer $x(s)$ can be easily obtained for any $s \in \mathbb{R}^n$.

DUAL AVERAGING
Nesterov 2013

$$\min_{x \in Q} \; f(x)$$

$$x_{k+1} = \underset{x \in Q}{\operatorname{argmin}} \left\{ \left\langle \frac{1}{k+1} \sum_{\ell=0}^{k} \nabla f(x_\ell), x \right\rangle + \frac{d(x)}{\sqrt{k+1}} \right\},$$

where

$$d(x) \geq d(x_0) = 0.$$

Convergence result:

$$f\left(\frac{1}{k+1} \sum_{\ell=0}^{k} x_\ell \right) - f(x^*) \leq \frac{d(x^*)}{\sqrt{k+1}} + \frac{1}{2\beta(k+1)} \sum_{\ell=0}^{k} \frac{\|\nabla f(x_\ell)\|_*^2}{\sqrt{\ell+1}}$$

Upper and lower bounds, and convexity parameter for $d(x)$ on Q are crucial

ADVANTAGES of PROX-FUNCTIONS

(1) The complexity bounds for optimization methods heavily depend on the size of the feasible set Q. This value has been traditionally defined with respect to Euclidean norm. However, the size of Q, measured with respect to another norm, can be smaller. By introducing prox-functions, which are strongly convex with respect to an appropriate norm $\|\cdot\|$, it is possible to take into account a particular geometry of the feasible set Q.

(2) Prox-functions often allow natural interpretations of the iteration steps within the convex optimization framework. This feature is important in order to explain agents' behavioral dynamics as being driven by unintentional optimization.

HERE: $Q = \Delta$ **simplex,** $\|\cdot\|_1$ **norm, probabilistic interpretation**

ADDITIVE RANDOM UTILITY MODELS
McFadden 1978

ARUM aims to model the discrete choice from a finite number of alternatives $\{1, \ldots, n\}$ by a rational decision-maker prone to some random errors. Accordingly, the *i*-th alternative is endowed with the utility

$$u^{(i)} + \varepsilon^{(i)},$$

where

- $u^{(i)}$ is its **deterministic part**,
- $\epsilon^{(i)}$ is the corresponding **random error**.

A rational decision-maker chooses alternatives with the maximal utility, so that the corresponding surplus is given by the expectation

$$E(u) = \mathbb{E}_{\epsilon} \left(\max_{1 \leq i \leq n} u^{(i)} + \epsilon^{(i)} \right).$$

CHOICE PROBABILITIES

Assumption

The random vector ϵ follows a joint distribution with finite mean that is absolutely continuous with respect to the Lebesgue measure and fully supported on $\mathbb{R}^n$.

$$\Downarrow$$

Surplus function E is convex and differentiable, in particular, its partial derivatives can be expressed as choice probabilities:

$$\frac{\partial E(u)}{\partial u^{(i)}} = \mathbb{P}\left(u^{(i)} + \epsilon^{(i)} = \max_{1 \leq i \leq n} u^{(i)} + \epsilon^{(i)}\right), \quad i = 1, \ldots, n.$$

CONVEX CONJUGATE of SURPLUS FUNCTION

$$E^*(p) = \sup_{u \in \mathbb{R}^n} \langle p, u \rangle - E(u)$$

Theorem (Continuity of E^*)

The convex conjugate E^* is continuous on its domain $\operatorname{dom} E^*$ which coincides with the simplex Δ.

Corollary (Upper bound for E^*)

The convex conjugate E^* is bounded from above on its domain Δ, namely it holds:

$$E^*(p) \leq - \min_{1 \leq i \leq n} \mathbb{E}_\epsilon \left(\epsilon^{(i)} \right) \quad \text{for all } p \in \Delta.$$

STRONG CONVEXITY: GENERAL CASE

Theorem (Strong convexity of E^*)

Let the differences $\epsilon^{(j)} - \epsilon^{(i)}$ of random utility shocks have modes $\bar{z}_{i,j} \in \mathbb{R}$, $i \neq j$, i.e. there density functions are bounded:

$$g_{ij}(z) \leq g_{ij}\left(\bar{z}_{ij}\right).$$

Then, the corresponding convex conjugate E^* is β-strongly convex with respect to the $\|\cdot\|_1$ norm, where the convexity parameter is given by

$$\beta = \frac{1}{2\sum\limits_{i=1}^{n}\sum\limits_{j\neq i} g_{i,j}\left(\bar{z}_{i,j}\right)}.$$

STRONG CONVEXITY: IID CASE

Corollary (Strong convexity of E^* for IID utility shocks)

Let the random utility shocks $\epsilon^{(i)}$, $i = 1, \ldots, n$, be independent and identically distributed with the common probability density function f having a mode $\bar{z} \in \mathbb{R}$. Then, the corresponding convex conjugate E^* is β-strongly convex with respect to the $\|\cdot\|_1$ norm, where the convexity parameter is given by

$$\beta = \frac{1}{2n(n-1)f(\bar{z})}.$$

Dependent on dimension n

MULTINOMIAL LOGIT

IID random utility shocks $\epsilon^{(i)}$, $i = 1, \ldots, n$, each of them following the Gumbel distribution with zero mode and variance $\mu\pi/\sqrt{6}$.

$$E(u) = \mu \ln \sum_{i=1}^{n} e^{u^{(i)}/\mu} + \mu\gamma.$$

$$\mathbb{P}\left(u^{(i)} + \epsilon^{(i)} = \max_{1 \le i \le n} u^{(i)} + \epsilon^{(i)}\right) = \frac{e^{u^{(i)}/\mu}}{\sum_{i=1}^{n} e^{u^{(i)}/\mu}}, \quad i = 1, \ldots, n.$$

$$E^*(p) = \mu \sum_{i=1}^{n} p^{(i)} \ln p^{(i)} - \mu\gamma = \mu H(p) - \mu\gamma,$$

where H is the (negative) entropy. Due to Pinsker inequality, H is 1-strongly convex with respect to $\|\cdot\|_1$. Hence, E^* is μ-strongly convex with respect to $\|\cdot\|_1$.

GENERALIZED EXTREME VALUE MODELS

Random utility shocks of GEV follow the joint distribution given by the probability density function

$$f_\epsilon\left(y^{(1)}, \ldots, y^{(n)}\right) = \frac{\partial^n \exp\left(-G\left(e^{-y^{(1)}}, \ldots, e^{-y^{(n)}}\right)\right)}{\partial y^{(1)} \cdots \partial y^{(n)}},$$

where the generating function $G : \mathbb{R}_+^n \to \mathbb{R}_+$ fulfills:

(G1) G is homogeneous of degree $1/\mu > 0$.

(G1) $G\left(x^{(1)}, \ldots, x^{(i)}, \ldots, x^{(n)}\right) \to \infty$ as $x^{(i)} \to \infty$, $i = 1, \ldots, n$.

(G3) For the partial derivatives of G with respect to k distinct variables it holds:

$$\frac{\partial^k G\left(x^{(1)}, \ldots, x^{(n)}\right)}{\partial x^{(i_1)} \cdots \partial x^{(i_k)}} \quad \begin{array}{ll} \geq 0 & \text{if } k \text{ is odd} \\ \leq 0 & \text{if } k \text{ is even.} \end{array}$$

STRONG CONVEXITY: GEV

Theorem (Strong convexity of E^* for GEV)

Let a generating function G for GEV satisfy the following inequality for all $x = \left(x^{(1)}, \ldots, x^{(n)}\right)^T \in \mathbb{R}^n_+$:

$$\sum_{i=1}^{n} \frac{\partial^2 G(x)}{\partial x^{(i)2}} \cdot x^{(i)2} \leq M \cdot G(x)$$

with some constant $M \in \mathbb{R}$. Then, the corresponding convex conjugate E^* is β-strongly convex with respect to the $\|\cdot\|_1$ norm, where the convexity parameter is given by

$$\beta = \frac{1}{2(\mu M + 1) - {}^{1}\!/_{\mu}}.$$

GENERALIZED NESTED LOGIT MODELS

GNL is a particular class of GEV models with

$$G(x) = \sum_{\ell \in L} \left(\sum_{i=1}^{n} \left(\sigma_{i\ell} \cdot x^{(i)} \right)^{1/\mu_\ell} \right)^{\mu_\ell/\mu},$$

where

- L is a generic set of nests.
- $\sigma_{i\ell}$ is the share of the i-th alternative within the ℓ-th nest.
- μ_ℓ describes the variance of the random errors while choosing alternatives within the ℓ-th nest.
- μ describes the variance of the random errors while choosing among the nests.

For the function G to fulfill (G1)-(G3) we require:

$$\mu_\ell \leq \mu \quad \text{for all } \ell \in L.$$

TWO-STAGES of CHOICE PROCESS

(1) the probability of choosing the ℓ-th nest is

$$q_\ell = \frac{e^{v_\ell/\mu}}{\sum_{\ell \in L} e^{v_\ell/\mu}}, \text{ where } v_\ell = \mu_\ell \ln \left(\sum_{i=1}^{n} \left(\sigma_{i\ell} \cdot e^{u^{(i)}} \right)^{1/\mu_\ell} \right)$$

stands for the utility attached to the ℓ-th nest;

(2) the probability of choosing the i-th alternative within the ℓ-th nest is

$$p_{i\ell} = \frac{\left(\sigma_{i\ell} \cdot e^{u^{(i)}} \right)^{1/\mu_\ell}}{\sum_{i=1}^{n} \left(\sigma_{i\ell} \cdot e^{u^{(i)}} \right)^{1/\mu_\ell}}.$$

Overall, the choice probability of the i-th alternative is

$$\mathbb{P} \left(u^{(i)} + \epsilon^{(i)} = \max_{1 \le i \le n} u^{(i)} + \epsilon^{(i)} \right) = \sum_{\ell \in L} q_\ell \cdot p_{i\ell}.$$

STRONG CONVEXITY: GNL

Corollary (Strong convexity of E^* for GNL)

For GNL the corresponding convex conjugate E^* is β-strongly convex with respect to the $\|\cdot\|_1$ norm, where the convexity parameter is given by

$$\beta = \frac{1}{\frac{2}{\min\limits_{\ell \in L} \mu_\ell} - {}^1\!/\!_\mu}.$$

Independent of dimension n

NESTED LOGIT

Let in GNL for every alternative i there be a unique nest $\ell_i \in L$ with $\sigma_{i\ell_i} = 1$, and $\mu = 1$. Then, the nests $N_\ell = \{i \mid \ell_i = \ell\}$ are mutually exclusive, and the generating function is

$$G(x) = \sum_{\ell \in L} \left(\sum_{i \in N_\ell} x^{(i)1/\mu_\ell} \right)^{\mu_\ell}$$

Fosgerau, Melo, de Palma, Shum, 2017:

$$E^*(p) = \sum_{\ell \in L} \mu_\ell \sum_{i \in N_\ell} p^{(i)} \ln p^{(i)} + \sum_{\ell \in L} (1 - \mu_\ell) \left(\sum_{i \in N_\ell} p^{(i)} \right) \ln \left(\sum_{i \in N_\ell} p^{(i)} \right).$$

The same order of magnitude:

$$\beta = \frac{1}{\frac{2}{\min\limits_{\ell \in L} \mu_\ell} - 1/\mu} = \frac{1}{\frac{2}{\min\limits_{\ell \in L} \mu_\ell} - 1} > \frac{1}{2} \min_{\ell \in L} \mu_\ell.$$

ORDERED GEV
Small 1987

is a GNL model with

$$L = \{1, \ldots, n+m\}, \quad \mu = 1,$$

$$\sigma_{i\ell} > 0 \text{ for all } \ell \in \{i, \ldots, i+m\},$$

$$\sigma_{i\ell} = 0 \text{ for all } \ell \in L \backslash \{i, \ldots, i+m\}.$$

There are $n+m$ overlapping nests $N_\ell = \{i \mid \ell - m \leq i \leq \ell\}$, and every alternative lies exactly in $m+1$ of them, namely $i \in N_\ell$ for $\ell = i, \ldots, i+m$. Then, the generating function is

$$G(x) = \sum_{\ell=1}^{n+m} \left(\sum_{i \in N_\ell} \left(\sigma_{i\ell} x^{(i)} \right)^{1/\mu_\ell} \right)^{\mu_\ell}.$$

PAIRED COMBINATORIAL LOGIT
Koppelman, Wen 2000

is a GNL model with

$$L = \{(i,j) \in \{1,\dots,n\} \mid i \neq j\}, \quad \mu = 1,$$

$$\sigma_{i\ell} = \begin{cases} \dfrac{1}{2(n-1)} & \text{if } \ell = (i,j),(j,i) \text{ with } j \neq i, \\ 0 & \text{else.} \end{cases}$$

There are $n^2 - n$ nests corresponding to the pairs of alternatives, and every alternative lies in $2(n-1)$ of them.

$$G(x) = \sum_{\ell=(i,j), i \neq j} \left(\left(\sigma_{i\ell} x^{(i)} \right)^{1/\mu_\ell} + \left(\sigma_{j\ell} x^{(j)} \right)^{1/\mu_\ell} \right)^{\mu_\ell}.$$

PRINCIPLES OF DEFFERENTIATION GEV
Bresnahan, Stern, Trajtenberg 1997

is a GNL model with

$$L = \dot{\bigcup_{d \in D}} L_d, \quad \mu = 1, \quad \mu_\ell = \mu_d \text{ for all } \ell \in L_d,$$

$$\sigma_{i\ell} = \begin{cases} \sigma_d & \text{if } i \in N_{\ell d} \text{ and } \ell \in L_d, \\ 0 & \text{else,} \end{cases}$$

where

$$\{1, \ldots, n\} = \dot{\bigcup_{\ell \in L_d}} N_{\ell d}.$$

The set D represents the dimensions of alternatives. For d-th dimension alternatives are clustered into disjoint nests $N_{\ell d}$, $\ell \in L_d$.

$$G(x) = \sum_{d \in D} \sigma_d \sum_{\ell \in L_d} \left(\sum_{i \in N_{\ell d}} \left(x^{(i)} \right)^{1/\mu_d} \right)^{\mu_d}.$$

CROSS MOMENT MODEL
Mishra, Natarajan, Tao, Teo 2012

In CMM the surplus function is maximized w.r.t. random error with zero mean and a given covariance Σ:

$$Z(u) = \max_{\epsilon \sim (0,\Sigma)} \mathbb{E}_\epsilon \left(\max_{1 \le i \le n} u^{(i)} + \epsilon^{(i)} \right).$$

With random error $\epsilon(u) \sim (0, \Sigma)$ that maximizes the surplus function, the corresponding choice probabilities are

$$p^{(i)} = \mathbb{P} \left(u^{(i)} + \epsilon^{(i)}(u) = \max_{1 \le i \le n} u^{(i)} + \epsilon^{(i)}(u) \right), \quad i = 1, \dots, n.$$

DUAL REPRESENTATION in CMM
Ahipasaoglu, Li, Natarajan 2019

$$Z(u) = \max_{p \in \Delta} \langle p, u \rangle + \operatorname{tr}\left(\left(\Sigma^{1/2}\left(\operatorname{diag}(p) - pp^T\right)\Sigma^{1/2}\right)^{1/2}\right).$$

The solution $p \in \Delta$ of the latter optimization problem provides the choice probabilities. Hence, the convex conjugate of Z is

$$Z^*(p) = -\operatorname{tr}\left(\left(\Sigma^{1/2}\left(\operatorname{diag}(p) - pp^T\right)\Sigma^{1/2}\right)^{1/2}\right).$$

Z^* is shown to be strongly convex on the simplex w.r.t. Euclidean norm. Z^* can be therefore used as a discrete choice prox-function on the simplex as well.

SIMPLICITY

Theorem (Simplicity)

The unique maximizer of the optimization problem

$$E(u) = \sup_{p \in \Delta} \langle p, u \rangle - E^*(p)$$

is given by the choice probabilities

$$p^{(i)} = \mathbb{P}\left(u^{(i)} + \epsilon^{(i)} = \max_{1 \leq i \leq n} u^{(i)} + \epsilon^{(i)} \right), \quad i = 1, \ldots, n.$$

LOWER BOUND

Corollary (Lower bound for E^*)

The unique minimizer p_0 of the convex conjugate E^* consists of the choice probabilities with respect to the zero-utility, i. e.

$$p_0^{(i)} = \mathbb{P}\left(\epsilon^{(i)} = \max_{1 \leq i \leq n} \epsilon^{(i)}\right), \quad i = 1, \dots, n.$$

Moreover, it holds:

$$E^*(p) \geq E^*(p_0) = -\mathbb{E}_\epsilon\left(\max_{1 \leq i \leq n} \epsilon^{(i)}\right) \quad \textit{for all } p \in \Delta.$$

UTILITY MAXIMIZATION

$$\max_{\substack{x \geq 0 \\ \langle \pi, x \rangle = 1}} \quad \underbrace{\min_{1 \leq i \leq n} \frac{(Qx)^{(i)}}{\sigma^{(i)}}}_{\textit{Leontieff utility } U(x)}$$

- consumption goods $j = 1, \ldots, m$
- prices of goods $\pi \in \mathbb{R}^m_+$
- demand for goods $x \in \mathbb{R}^m_+$
- qualities of goods $i = 1, \ldots, n$
- quality standards $\sigma \in \mathbb{R}^n_+$
- quality matrix $Q = (q_{ij}) \in \mathbb{R}^{n \times m}$ with q_{ij} denoting the amount of quality i while consuming one unit of good j

spend 1 € for demand x in order to maximize
the worst ratio of consumption/standard qualities

DUAL PROBLEM

$$\max_{\substack{x \geq 0 \\ \langle \pi, x \rangle = 1}} \underbrace{\min_{1 \leq i \leq n} \frac{(Qx)^{(i)}}{\sigma^{(i)}}}_{\text{Leontieff utility } U(x)} = \max_{\substack{x \geq 0 \\ \langle \pi, x \rangle = 1}} \underbrace{\min_{\lambda \in \Delta}}_{\text{simplex}} \left\langle \frac{Qx}{\sigma}, \lambda \right\rangle$$

$$\max_{\substack{x \geq 0 \\ \langle \pi, x \rangle = 1}} \min_{\substack{\lambda \geq 0 \\ \langle \sigma, \lambda \rangle = 1}} \langle Qx, \lambda \rangle \overset{\text{duality}}{=} \min_{\substack{\lambda \geq 0 \\ \langle \sigma, \lambda \rangle = 1}} \max_{\substack{x \geq 0 \\ \langle \pi, x \rangle = 1}} \langle Qx, \lambda \rangle$$

$$\min_{\substack{\lambda \geq 0 \\ \langle \sigma, \lambda \rangle = 1}} \max_{\substack{x \geq 0 \\ \langle \pi, x \rangle = 1}} \left\langle x, Q^T \lambda \right\rangle = \min_{\substack{\lambda \geq 0 \\ \langle \sigma, \lambda \rangle = 1}} \underbrace{\max_{1 \leq j \leq m} \frac{\left(Q^T \lambda\right)^{(j)}}{\pi^{(j)}}}_{\text{Overvaluation } \Phi(\lambda)}$$

adjust internal prices of qualities λ in order to minimize the best quality/price ratio of goods

BUYING & CONSUMING

$$\max_{\substack{x \geq 0 \\ \langle \pi, x \rangle = 1}} \quad \min_{1 \leq i \leq n} \frac{(Qx)^{(i)}}{\sigma^{(i)}} \quad = \quad \min_{\substack{\lambda \geq 0 \\ \langle \sigma, \lambda \rangle = 1}} \quad \max_{1 \leq j \leq m} \frac{(Q^T \lambda)^{(j)}}{\pi^{(j)}}$$

$$\begin{tikzcd}
\boxed{\begin{array}{c}\text{Internal prices} \\ \text{of qualities } \lambda\end{array}} \arrow[r] & \boxed{\begin{array}{c}\text{Buy goods with} \\ \text{best quality/price} \\ \text{ratio } \frac{(Q^T\lambda)^{(j)}}{\pi^{(j)}}\end{array}} \arrow[d, "x"] \\
\boxed{\begin{array}{c}\text{Choose qualities with} \\ \text{worst consumption/standard} \\ \text{ratio } \frac{(Qx)^{(i)}}{\sigma^{(i)}}\end{array}} \arrow[u, "\lambda^+", "\textit{choice probabilities}"'] & \boxed{\begin{array}{c}\text{Experience qualities } Qx \\ \text{by consuming goods}\end{array}} \arrow[l]
\end{tikzcd}$$

SCALING of INTERNAL PRICES

$$\min_{\substack{\lambda \geq 0 \\ \langle \sigma, \lambda \rangle = 1}} \underbrace{\max_{1 \leq j \leq m} \frac{\left(Q^T \lambda\right)^{(j)}}{\pi^{(j)}}}_{\Phi(\lambda)}$$

Using new variables

$$p^{(i)} = \sigma^{(i)} \lambda^{(i)}, \quad i = 1, \ldots, n,$$

the dual problem reads:

$$\min_{p \in \Delta} \Psi(p),$$

where the objective function is

$$\Psi(p) = \Phi\left(\frac{p}{\sigma}\right).$$

DUAL AVERAGING
Nesterov 2013

1. Compute $\nabla\Psi\left(p_k\right)$.

2. Set $s_{k+1} = \dfrac{1}{k+1}\displaystyle\sum_{\ell=0}^{k} \nabla\Psi\left(p_\ell\right)$.

3. Update $p_{k+1} = \arg\min\limits_{p\in\Delta}\left\{\langle s_{k+1}, p\rangle + \dfrac{d(p)}{\sqrt{k+1}}\right\}$.

$$d(p) := \underbrace{E^*(p) - E^*\left(p_0\right)}_{\textit{discrete-choice prox-function}}$$

STEP 1

$$\nabla \Psi (p_k) = \frac{\nabla \Phi (\lambda_k)}{\sigma} = \frac{Q y_k / \pi}{\sigma},$$

where the sharing vector $y_k \in \Delta$ fulfills

$$y_k^{(j)} = 0 \quad \text{for } j \notin J(\lambda_k)$$

and the active index set $J(\lambda_k)$ contains goods with the best quality/price ratio estimated by means of internal prices λ_k.
We set the demand at the k-th iteration as

$$x_k = \frac{y_k}{\pi}.$$

Overall, we obtain:

$$\nabla \Psi (p_k) = \frac{Q x_k}{\sigma}.$$

STEP 2

We set

$$s_{k+1} = \frac{1}{k+1} \sum_{\ell=0}^{k} \nabla \Psi \left(p_\ell\right) = \frac{1}{k+1} \sum_{\ell=0}^{k} \frac{Qx_\ell}{\sigma} = \frac{Q\bar{x}_k}{\sigma}$$

with the average demand

$$\bar{x}_k = \frac{1}{k+1} \sum_{\ell=0}^{k} x_\ell.$$

Thus, s_{k+1} relates the average consumption $Q\bar{x}_k$ to standards σ.

STEP 3

$$p_{k+1} = \arg\min_{p \in \Delta} \left\{ \langle s_{k+1}, p \rangle + \frac{E^*(p) - E^*(p_0)}{\sqrt{k+1}} \right\}.$$

Due to simplicity, we equivalently obtain for $i = 1, \ldots, n$:

$$p_{k+1}^{(i)} = \mathbb{P}\left(s_{k+1}^{(i)} - \frac{\epsilon^{(i)}}{\sqrt{k+1}} = \min_{1 \le i \le n} s_{k+1}^{(i)} - \frac{\epsilon^{(i)}}{\sqrt{k+1}} \right).$$

For the internal prices we have:

$$\lambda_{k+1}^{(i)} = \frac{1}{\sigma^{(i)}} \mathbb{P}\left(s_{k+1}^{(i)} - \frac{\epsilon^{(i)}}{\sqrt{k+1}} = \min_{1 \le i \le n} s_{k+1}^{(i)} - \frac{\epsilon^{(i)}}{\sqrt{k+1}} \right).$$

Thus, the internal price $\lambda_{k+1}^{(i)}$ of the i-th quality is proportional to the probability of detecting its average consumption $\left(Q\bar{x}_k\right)^{(i)}$ as the lowest one as compared to the standard $\sigma^{(i)}$.

CONVERGENCE RESULT

Theorem

The duality gap between (P) and (D) evaluated at the average demand and the average internal prices is closing at the optimal rate $O\left(\frac{1}{\sqrt{k+1}}\right)$. Namely, it holds for $k \geq 0$:

$$0 \leq \Phi\left(\bar{\lambda}_k\right) - U\left(\bar{x}_k\right) \leq \left(D + \frac{M^2}{\beta}\right) \frac{1}{\sqrt{k+1}},$$

where

$$M = \max_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}} \frac{|q_{i,j}|}{\sigma^{(i)} \cdot \pi^{(j)}}, \quad D = \mathbb{E}_\epsilon \left(\max_{1 \leq i \leq n} \epsilon^{(i)} \right) - \min_{1 \leq i \leq n} \mathbb{E}_\epsilon \left(\epsilon^{(i)} \right).$$

Literature

- Müller, David ; Nesterov, Yurii ; Shikhman, Vladimir. Discrete choice prox-functions on the simplex. In: Mathematics of Operations Research, to appear.

- Müller, David ; Nesterov, Yurii ; Shikhman, Vladimir. Dynamic pricing under nested logit demand. In: Pure and Applied Functional Analysis, to appear.