Online Supplementary Material
Robust Data-Driven Quasiconcave Optimization

Jian Wu, William B. Haskell, Wenjie Huang, and Huifu Xu

8. Alternative Representation

This section gives an alternative representation of our worst-case objective function which highlights the role of a set of convex performance functions and targets for all levels of the worst-case objective. This result shows that $\psi_{\mathcal{U}}$ can be represented by a set of J convex functions, and a continuum of targets for all levels. We explicitly determine this set of convex functions (which are closely related to the upper level sets of $\psi_{\mathcal{U}}$) as well as the targets. This result is based on the choice model in Brown et al. (2012) and its representation in terms of a family of convex risk measures and targets.

We first derive the representation of $\psi_{\mathcal{U}}$. We can represent $\psi_{\mathcal{U}}$ in terms of a set of convex functions (one for every input θ_j for $j \in [J]$) and targets (one for every level $\upsilon \leq \upsilon_{\max}$). First we define constants $c_j \triangleq -\inf \left\{ m \in \mathbb{R} \mid m\vec{1}_N \geq \sum_{\theta\in\mathcal{D}_j} \tilde{\theta}\cdot p_\theta,\ \sum_{\theta\in\mathcal{D}_j} p_\theta = 1,\ p \in \mathbb{R}^j_{\geq 0} \right\}$, $j \in [J]$, and upper level sets $\mathcal{A}_j \triangleq \left\{ x \in \mathcal{X} \mid x \geq \sum_{\theta\in\mathcal{D}_j} \tilde{\theta}\cdot p_\theta + c_j,\ \sum_{\theta\in\mathcal{D}_j} p_\theta = 1,\ p \in \mathbb{R}^j_{\geq 0} \right\}$, $j \in [J]$, corresponding to each $\mathcal{D}_j$ (note these correspond to the characterization of the upper level sets in Theorem 2). Then, we define convex functions $\mu_j : \mathbb{R}^N \to \mathbb{R}$ via: $\mu_j(x) \triangleq \inf_{m\in\mathbb{R}} \left\{ m \mid x + m\vec{1}_N \in \mathcal{A}_j \right\}$, $\forall x \in \mathcal{X}$, for all $j \in [J]$. We can interpret $\mu_j(x)$ as the minimum amount which must be added to every component of x to put it in the upper level set $\mathcal{A}_j$, so smaller values of $\mu_j(x)$ mean x is closer to/deeper into $\mathcal{A}_j$. The value $\mu_j(x)$ is the cost of putting x into $\mathcal{A}_j$.

Next we establish the properties of $\{\mu_j\}_{j\in[J]}$. We say μ_j is translation invariant (with respect to the direction $\vec{1}_N$) if $\mu_j(x + \beta\vec{1}_N) = \mu_j(x) - \beta$ for all $\beta \in \mathbb{R}$. We say μ_j is normalized if $\mu_j(0) = 0$ for all $j \in [J]$. We require the following technical condition so that μ_j are normalized.

ASSUMPTION 2. *We have* $0 \in \mathcal{X}$.

Assumption 2 can be met without loss of generality by translation of the data points in Θ.

PROPOSITION 7. *Suppose Assumption 2 holds.*
(i) $\{\mu_j\}_{j\in[J]}$ *are monotone, convex, translation invariant, and normalized.*
(ii) $\mu_j \geq \mu_{j+1}$ *for all* $j \in [J-1]$.

(i) First, note that all of the sets $\mathcal{A}_j$ are monotone by construction. If $x \in \mathcal{A}_j$ and $y \geq x$, then $y \in \mathcal{A}_j$. Next, choose $x \leq y$, then $y + m\vec{1}_N \geq x + m\vec{1}_N$ for all $m \in \mathbb{R}$. It follows that $\mu_j(x) \geq \mu_j(y)$.

All $\mathcal{A}_j$ are convex because they are polyhedra. Pick $x_1, x_2 \in \mathcal{X}$, $\epsilon > 0$, and $\lambda \in [0,1]$. For $i = 1,2$, let m_i satisfy $x_i + m_i\vec{1}_N$ and $m_i \leq \mu_j(x_i) + \epsilon$. Then $x_i + m_i\vec{1}_N \in \mathcal{A}_j$ for $i = 1,2$, and so $\lambda(x_1 + m_1\vec{1}_N) + (1-\lambda)(x_2 + m_2\vec{1}_N) = \lambda x_1 + (1-\lambda)x_2 + (\lambda m_1 + (1-\lambda)m_2)\vec{1}_N \in \mathcal{A}_j$, by convexity of $\mathcal{A}_j$. It follows that $\mu_j(\lambda x_1 + (1-\lambda)x_2) \leq \lambda m_1 + (1-\lambda)m_2 \leq \lambda\mu_j(x_1) + (1-\lambda)\mu_j(x_2) + \epsilon$. Since ϵ was arbitrary, we conclude that μ_j is convex.

To check translation invariance, for any $\beta \in \mathbb{R}$ we have:

$$\begin{aligned}\mu_j(x+\beta\vec{1}_N) &= \inf_{m\in\mathbb{R}}\left\{m \mid x+(m+\beta)\vec{1}_N \in \mathcal{A}_j\right\}\\ &= \inf_{y\in\mathbb{R}}\left\{y-\beta \mid x+y\,\vec{1}_N \in \mathcal{A}_j\right\}\\ &= \inf_{m\in\mathbb{R}}\left\{m \mid x+m\,\vec{1}_N \in \mathcal{A}_j\right\}-\beta,\end{aligned}$$

which shows that μ_j is translation invariant.

To check normalization, let $x=0$ by Assumption 2 so $x+m\,\vec{1}_N = m\,\vec{1}_N$. By the definition of c_j, for any $m>0$, we have: $x+m\,\vec{1}_N - c_j = m\,\vec{1}_N - c_j \geq \sum_{\theta\in\mathcal{D}_j}\tilde{\theta}\cdot p_\theta$ for some p. It then follows that $x+m\,\vec{1}_N \in \mathcal{A}_j$ by definition of μ_j. Now, for any $m<0$, there is no p such that $x+m\,\vec{1}_N - c_j \geq \sum_{\theta\in\mathcal{D}_j}\tilde{\theta}\cdot p_\theta$, which means that $x+m\,\vec{1}_N \notin \mathcal{A}_j$. Hence, we conclude $\mu_j(0)=0$.

(ii) By construction, $\mathcal{A}_j \subset \mathcal{A}_{j+1}$ since $\mathcal{D}_j \subset \mathcal{D}_{j+1}$ for all $j\in[J-1]$. If $x+m\,\vec{1}_N \in \mathcal{A}_j$, then $x+m\,\vec{1}_N \in \mathcal{A}_{j+1}$. Consequently, $\mu_{j+1}(x) \leq \mu_j(x)$.

Proposition 7 has close connections with the theory of risk measures. In the language of risk measures, $\mathcal{A}_j$ are "acceptance sets" which describe a set of satisfactory positions. Notice that they are convex sets, and also monotone so if $x\in\mathcal{A}_j$ is acceptable then all $y\geq x$ are as well. We have $\mathcal{A}_j \subset \mathcal{A}_{j+1}$ so the acceptance sets become more stringent as j decreases. Corresponding to each acceptance set $\mathcal{A}_j$, μ_j is called a convex risk measure which returns the minimum amount of cash m needed to make $x+m$ acceptable. Then, we have the relationship $\mathcal{A}_j = \{x : \mu_j(x)\leq 0\}$.

Now we define the target function $\tau(\upsilon) \triangleq \upsilon/L - c_{\kappa(\upsilon)}$ for all levels $\upsilon \leq \upsilon_{\max}$. We specify that $\tau(\upsilon)$ is the target for all N components of x (so $\tau(\upsilon)\vec{1}_N$ appears in the following expression to compare to x).

THEOREM 4. *Suppose Assumption 2 holds.*

(i) The target function $\tau(\upsilon)=\upsilon/L - c_{\kappa(\upsilon)}$ is non-decreasing in υ.

(ii) We have

$$\psi_{\mathcal{U}}(x) = \sup_{\upsilon\leq\upsilon_{\max}}\left\{\upsilon \mid \mu_{\kappa(\upsilon)}(x-\tau(\upsilon)\vec{1}_N)\leq 0\right\},\ \forall x\in\mathcal{X}. \tag{8.1}$$

F irst, we have $\psi_{\mathcal{U}}(x) = \sup\{\upsilon\leq\upsilon_{\max} : x\in\mathcal{A}(\psi_{\mathcal{U}},\upsilon)\}$ for all $x\in\mathcal{X}$ by Proposition 3. By Theorem 2, we can write $\mathcal{A}(\psi_{\mathcal{U}},\upsilon)$ explicitly as:

$$\mathcal{A}(\psi_{\mathcal{U}},\upsilon) = \left\{x \mid x \geq \sum_{j=1}^{\kappa(\upsilon)}\tilde{\theta}\cdot p_\theta + \upsilon/L,\ \sum_{j=1}^{\kappa(\upsilon)} p_\theta = 1,\ p\in\mathbb{R}^j_{\geq 0}\right\}.$$

It then follows that:

$$\begin{aligned}\psi_{\mathcal{U}}(x) &= \sup\left\{\upsilon\leq\upsilon_{\max} \mid x \geq \sum_{j=1}^{\kappa(\upsilon)}\tilde{\theta}\cdot p_\theta + \upsilon/L,\ \sum_{j=1}^{\kappa(\upsilon)} p_\theta = 1,\ p\geq 0\right\}\\ &= \sup\left\{\upsilon\leq\upsilon_{\max} \mid x-\upsilon/L + c_{\kappa(\upsilon)} \in \mathcal{A}_{\kappa(\upsilon)}\right\}\\ &= \sup\left\{\upsilon\leq\upsilon_{\max} \mid \mu_{\kappa(\upsilon)}(x-\upsilon/L+c_{\kappa(\upsilon)})\leq 0\right\}\\ &= \sup\left\{\upsilon\leq\upsilon_{\max} \mid \mu_{\kappa(\upsilon)}(x-\tau(\upsilon))\leq 0\right\}.\end{aligned}$$

In the above display: the second equality follows from the definition of μ_j where $j = \kappa(\upsilon)$; the third equality follows from the fact that $\mu_j(x) \le 0$ for $x \in \mathcal{X}$ if and only if $x \in \mathcal{A}_j$; and the last equality follows from the definition of the target function $\tau(\upsilon)$.

Eq. (8.1) can be viewed as a decomposition of $\psi_{\mathcal{U}}$ into a set of convex cost functions $\{\mu_j\}_{j\in[J]}$ and a target function $\tau(\upsilon)$. Note that there is a convex cost function for every input $j \in [J]$ in the original data sample.

We can alternatively choose upper level sets and targets to construct an evaluation function $\psi \in \mathcal{F}_{\mathrm{QCo}}$. We can generalize the more specific Theorem 4 with the following ingredients:

- An index function $\kappa : \mathbb{R} \to [L]$ that is decreasing for $L \ge 1$.
- A collection $\{\mu_l\}$ of monotone, convex, translation invariant, and normalized functions. Furthermore, $\{\mu_l\}$ satisfy $\mu_l \ge \mu_{l+1}$ for all $l \in [L-1]$.
- A target function $\vec{\tau} : \mathbb{R} \to \mathbb{R}^N$ that is component-wise non-decreasing in υ, where $\vec{\tau}(\upsilon) = (\tau_1(\upsilon), \ldots, \tau_N(\upsilon))$ and $\tau_n(\upsilon)$ is the target for component n at level υ.

This setup allows more flexibility in designing a valuation function $\psi : \mathcal{X} \to \mathbb{R}$ defined by:

$$\psi(x) = \sup\{\upsilon : \mu_{\kappa(\upsilon)}(x - \vec{\tau}(\upsilon)) \le 0\},\ x \in \mathcal{X}. \tag{8.2}$$

Eq. (8.2) can be interpreted as the largest valuation υ such that x is acceptable relative to the target $\vec{\tau}(\upsilon)$.

THEOREM 5. *Let ψ be defined by Eq. (8.2). Then $\psi \in \mathcal{F}_{QCo}$.*

F irst we verify that ψ is monotone. Choose $x \in \mathcal{X}$, $\epsilon > 0$, and υ such that $\mu_{\kappa(\upsilon)}(x - \vec{\tau}(\upsilon)) \le 0$ and $\upsilon \ge \psi(x) - \epsilon$. For any $y \ge x$, since all $\{\mu_l\}$ are monotone we have $\mu_{\kappa(\upsilon)}(y - \vec{\tau}(\upsilon)) \le 0$. It follows that $\psi(y) \ge \upsilon \ge \psi(x) - \epsilon$. Since ϵ was arbitrary, we conclude $\psi(y) \ge \psi(x)$.

Next we verify that ψ is quasiconcave. Choose $x_1, x_2 \in \mathcal{X}$. Without loss of generality, suppose $\psi(x_1) \le \psi(x_2)$. Then choose $\epsilon > 0$, and υ_1, υ_2 such that $\upsilon_1 \le \upsilon_2$ and $\upsilon_i \ge \psi(x_i) - \epsilon$ for $i = 1, 2$. We have $\mu_{\kappa(\upsilon_i)}(x_i - \vec{\tau}(\upsilon_i)) \le 0$ for $i = 1, 2$ by definition of ψ. In addition, we have: $\mu_{\kappa(\upsilon_1)}(x_2 - \vec{\tau}(\upsilon_1)) \le \mu_{\kappa(\upsilon_1)}(x_2 - \vec{\tau}(\upsilon_2)) \le \mu_{\kappa(\upsilon_2)}(x_2 - \vec{\tau}(\upsilon_2)) \le 0$, where the first inequality follows by monotonicity of $\mu_{\kappa(\upsilon_1)}$ and the second inequality follows by monotonicity of μ_l in l. By convexity of $\mu_{\kappa(\upsilon_1)}$, we have: $\mu_{\kappa(\upsilon_1)}(\lambda x_1 + (1-\lambda)x_2 - \vec{\tau}(\upsilon)) \le 0$, $\forall \lambda \in [0, 1]$. Then, we have $\psi(\lambda x_1 + (1-\lambda)x_2) \ge \upsilon_1 \ge \min\{\psi(x_1), \psi(x_2)\} - \epsilon$. Since ϵ was arbitrary, we have $\psi(\lambda x_1 + (1-\lambda)x_2) \ge \min\{\psi(x_1), \psi(x_2)\}$.

The target function for our worst-case evaluation function $\psi_{\mathcal{U}}$ in Eq. (8.1) is $\vec{\tau}(\upsilon) = \tau(\upsilon)\vec{1}_N$ where all the components are identical, and the convex loss functions $\{\mu_j\}_{j\in[J]}$ are based on the upper level sets $\{\mathcal{A}_j\}_{j\in[J]}$ corresponding to our data sample. Eq. (8.2) generalizes this construction by giving more flexibility in the choice of convex performance functions and targets.

9. Permutation Invariance

In this section, we incorporate the property of permutation invariance into our robust choice model. This property naturally emerges in many applications. For instance, we may be evaluating the utility of a group of identical customers, so interchanging their order should not change the overall evaluation. Permutation invariance also corresponds to the property of law invariance of choice functions in the decision theory literature.

To define a set of permutation invariant functions in $\mathfrak{F}$, we first decompose the components of $x \in \mathbb{R}^N$ into groups of equal size. Fix $M \ge 1$ to be the number of groups, and suppose $N = MK$ where $K \ge 1$ is the size of each group (so

M evenly divides N). Then let the set $\Omega = \{\omega_1, \ldots, \omega_M\}$ index groups, so we can write $x = (x(\omega_1), \ldots, x(\omega_M))$ where each $x(\omega_m) = (x_{K(m-1)+1}, \ldots, x_{Km}) \in \mathbb{R}^K$ for $m \in [M]$. For example, in multi-objective stochastic optimization, M is the number of problem scenarios (for optimization under uncertainty) and K is the number of problem attributes. This setup gives us flexibility in the implementation of permutation invariance, depending on the problem at hand.

Now let Σ denote the set of all permutations $[M] \to [M]$, and $\sigma \in \Sigma$ denote a specific permutation. Then we define $\sigma(x) = (x(\omega_{\sigma(1)}), \ldots, x(\omega_{\sigma(M)}))$ to be a permutation of x according to σ. It changes the order of the groups defined above (but does not change the order of elements within each group).

DEFINITION 7. (a) A function $f : \mathcal{X} \to \mathbb{R}$ is permutation invariant if $f(x) = f(\sigma(x))$ for all $\sigma \in \Sigma$.

(b) We let $\mathcal{F}^{\dagger}_{\text{QCo}}$ denote the set of all permutation invariant functions in $\mathcal{F}_{\text{QCo}}$.

We let $\mathcal{U}^{\dagger} = \mathcal{U}^{\dagger}(D, L) = \{f \in \mathcal{F}^{\dagger}_{\text{QCo}} \cap \mathcal{F}_{\text{Lip}}(L) : f(\theta) \geq \hat{v}(\theta),\ \forall \theta \in \Theta\}$, denote our ambiguity set for the permutation invariant case. In particular, the permutation invariant ambiguity set $\mathcal{U}^{\dagger} \subset \mathcal{U}$ is a strict subset of the original ambiguity set.

The corresponding worst-case evaluation function is $\psi_{\mathcal{U}^{\dagger}}$, and our RO problem becomes:

$$\mathscr{P}(\mathcal{U}^{\dagger}) : \max_{z \in \mathcal{Z}} \psi_{\mathcal{U}^{\dagger}}(G(z)).$$

We will solve $\mathscr{P}(\mathcal{U}^{\dagger})$ using the same approach we did for $\mathscr{P}(\mathcal{U})$, computing the upper level sets and using binary search. Our first step is to obtain the upper level sets of $\psi_{\mathcal{U}^{\dagger}}$. We introduce the following additional notation:

- $\vec{x}_k = (x_k(w_m))_{m=1}^M$ is the vector of components corresponding specifically to component $k \in [K]$ of x across all M groups. Then, $\sigma(\vec{x}_k) = (x_k(\omega_{\sigma(m)}))_{m=1}^M$ is a permutation of $\vec{x}_k$ which just permutes the order of the groups for fixed component k.
- θ_k is the vector of components corresponding to component $k \in [K]$ of input $\theta \in \mathbb{R}^N$ across all groups;
- $s_k \in \mathbb{R}^M$ is the subgradient of f at θ corresponding to component $k \in [K]$ across all groups;
- $y(\theta) \in \mathbb{R}^M$ and $w(\theta) \in \mathbb{R}^M$ are auxiliary decision variables for every $\theta \in \mathcal{D}_j$. Let $y = \{y(\theta')\}_{\theta' \in \mathcal{D}_j}$ and $w = \{w(\theta')\}_{\theta' \in \mathcal{D}_j}$.

We define the following LP, which is used to check the upper level sets of $\psi_{\mathcal{U}^{\dagger}}$:

$$\begin{aligned}
\mathscr{P}^{\dagger}_{\mathsf{LP}}(\theta; \mathcal{D}_j) : \min_{\upsilon, \xi, y, w} \quad & \upsilon \\
\text{s.t.} \quad & \langle \vec{1}_M, y(\theta') \rangle + \langle \vec{1}_M, w(\theta') \rangle - \langle \xi, \theta \rangle + \upsilon - \hat{v}(\theta') \geq 0, \quad \forall \theta' \in \mathcal{D}_j, \\
& \sum_{k=1}^{K} \theta'_k s_k^{\top} - y(\theta') \vec{1}_M^{\top} - \vec{1}_M w(\theta')^{\top} \geq 0, \quad \forall \theta' \in \mathcal{D}_j, \\
& \xi \geq 0, \|\xi\|_1 \leq L.
\end{aligned}$$

In particular, $\mathscr{P}^{\dagger}_{\mathsf{LP}}(\theta; \mathcal{D}_j)$ has a polynomial number of constraints even though there is an exponential number of permutations. This feature is essential for our binary search algorithm to be tractable.

PROPOSITION 8. *Fix $x \in \mathbb{R}^N$ and let $\upsilon = \psi_{\mathcal{U}^{\dagger}}(x)$. Then $\psi_{\mathcal{U}^{\dagger}}(x) = \min\{\hat{v}(\theta_j), \mathsf{val}(\mathscr{P}^{\dagger}_{LP}(x; \mathcal{D}_j))\}$ for $j = \kappa(\upsilon)$.*

W e first introduce the disjunctive programming problem

$$\begin{aligned}
\mathscr{P}^{\dagger}(x;\mathcal{D}_j):\ \min_{\upsilon\in\mathbb{R},\,\xi\in\mathbb{R}^N}\quad & \upsilon && (9.2a)\\
\text{s.t.}\quad & \upsilon+\max\{\langle\xi,\sigma(\theta)-x\rangle,0\}\geq\hat{v}(\theta),\ \forall\theta\in\mathcal{D}_j,\ \sigma\in\Sigma, && (9.2b)\\
& \xi\geq 0,\ \|\xi\|_1\leq L. && (9.2c)
\end{aligned}$$

By the same reasoning as Proposition 4, we have that $\psi_{\mathcal{U}^{\dagger}}(x)=\mathsf{val}(\mathscr{P}^{\dagger}(x;\mathcal{D}_j))$. Next let $\Sigma(\mathcal{D}_j)\triangleq\{\sigma(\theta):\theta\in\mathcal{D}_j,\sigma\in\Sigma\}$ and introduce the LP:

$$\begin{aligned}
\mathscr{P}_{LP}(x;\Sigma(\mathcal{D}_j)):\ \min_{\upsilon\in\mathbb{R},\,\xi\in\mathbb{R}^N}\quad & \upsilon && (9.3a)\\
\text{s.t.}\quad & \upsilon+\langle\xi,\sigma(\theta)-x\rangle\geq\hat{v}(\theta),\ \forall\theta\in\mathcal{D}_j,\ \sigma\in\Sigma, && (9.3b)\\
& \xi\geq 0,\ \|\xi\|_1\leq L. && (9.3c)
\end{aligned}$$

By the same reasoning as Proposition 5, we have $\psi_{\mathcal{U}^{\dagger}}(x)=\min\{\hat{v}(\theta_j),\mathsf{val}(\mathscr{P}_{LP}(x;\Sigma(\mathcal{D}_j)))\}$.

We note that Eq. (9.3b) has an exponential number of constraints due to the index $\sigma\in\Sigma$. We obtain a reduction of the cut generation problem as follows. Constraint (9.3b) is equivalent to:

$$\min_{\sigma\in\Sigma}\langle\xi,\sigma(\theta')\rangle-\langle\xi,\theta\rangle+\upsilon-\hat{v}(\theta')\geq 0,\quad\forall\theta'\in\mathcal{D}_t. \tag{9.4}$$

We will reduce the optimization problem $\min_{\sigma\in\Sigma}\langle\xi,\sigma(\theta')\rangle$ in Eq. (9.4). Recall that ξ_k is the subgradient of f at θ corresponding to attribute $k\in[K]$. The optimal value of $\min_{\sigma\in\Sigma}\langle\xi,\sigma(\theta')\rangle$ in Eq. (9.4) is equal to the optimal value of:

$$\begin{aligned}
\min_{Q\in\mathbb{R}^{M\times M}}\quad & \sum_{k=1}^{K}\xi_k^{\top}Q\,\theta'_k && (9.5a)\\
\text{s.t.}\quad & Q^{\top}\vec{1}_M=\vec{1}_M, && (9.5b)\\
& Q\vec{1}_M=\vec{1}_M, && (9.5c)\\
& Q_{m,l}\in\{0,1\},\quad\forall l,\,m\in[M], && (9.5d)
\end{aligned}$$

which is a linear assignment problem. Here Q is the permutation matrix corresponding to σ so that $Q\,\theta'_k=\sigma(\theta'_k)$ for all $k\in[K]$ (the permutation must be the same for all attributes, hence we only have a single permutation matrix Q). Problem (9.5) can be solved exactly by relaxing the binary constraints $Q_{m,l}\in\{0,1\}$ to $0\leq Q_{m,l}\leq 1$ for all $l,\,m\in[M]$. Strong duality holds for the relaxed problem, and the optimal value of the relaxed problem is equal to:

$$\begin{aligned}
\max_{w,y\in\mathbb{R}^M}\quad & \langle\vec{1}_M,w\rangle+\langle\vec{1}_M,y\rangle\\
\text{s.t.}\quad & \sum_{k=1}^{K}\theta'_k\xi_k^{\top}-w\vec{1}_M^{\top}-\vec{1}_M y^{\top}\geq 0.
\end{aligned}$$

It follows that constraint (9.4) is satisfied if and only if there exists w and y such that:

$$\begin{aligned}
&\langle\vec{1}_M,w\rangle+\langle\vec{1}_M,y\rangle-\langle\xi,\theta\rangle+\upsilon-\hat{v}(\theta')\geq 0,\\
&\sum_{k=1}^{K}\theta'_k\xi_k^{\top}-w\vec{1}_M^{\top}-\vec{1}_M y^{\top}\geq 0,
\end{aligned}$$

which gives the desired form of $\mathscr{P}^{\dagger}_{LP}(x;\,\mathcal{D}_j)$.

The dual of $\mathscr{P}^{\dagger}_{\mathsf{LP}}(x;\mathcal{D}_j)$ leads to an explicit characterization of the upper level sets of $\psi_{\mathcal{U}^\dagger}$ in the following theorem.

THEOREM 6. *The upper level set of $\psi_{\mathcal{U}^\dagger}$ at level $\upsilon \in (-\infty, \upsilon_{\max}]$ satisfies:*

$$\mathcal{A}(\psi_{\mathcal{U}^\dagger}, \upsilon) = \left\{ x \in \mathcal{X} \;\middle|\; \begin{array}{l} \sum_{\theta \in \mathcal{D}_{\kappa(\upsilon)}} \hat{\nu}(\theta) \cdot p_\theta - L\, q \geq \upsilon, \\ \sum_{\theta \in \mathcal{D}_{\kappa(\upsilon)}} \rho_\theta^\top \cdot \theta_k - \vec{x}_k \leq q, \quad \forall k \in [K] \\ \sum_{\theta \in \mathcal{D}_{\kappa(\upsilon)}} p_\theta = 1,\ p \in \mathbb{R}^j_{\geq 0},\ q \geq 0, \\ \vec{1}_M^\top \rho_\theta = p_\theta \vec{1}_M^\top,\ \rho_\theta \vec{1}_M = p_\theta \vec{1}_M,\ \rho_\theta \in \mathbb{R}^M_{\geq 0} \times \mathbb{R}^M_{\geq 0},\ \forall \theta \in \mathcal{D}_{\kappa(\upsilon)} \end{array} \right\}.$$

F irst fix $j = \kappa(\upsilon)$. The dual of $\mathscr{P}^{\dagger}_{\mathsf{LP}}(x;\,\mathcal{D}_j)$ is as follows. First define variables $p \in \mathbb{R}^j_{\geq 0}$, $q \in \mathbb{R}$, and $\{\rho_\theta\}_{\theta \in \mathcal{D}_j}$ where $\rho_\theta \in \mathbb{R}^M_{\geq 0} \times \mathbb{R}^M_{\geq 0}$. The dual to $\mathscr{P}^{\dagger}_{\mathsf{LP}}(x;\,\mathcal{D}_j)$ is then:

$$D^\dagger(x;\mathcal{D}_j):\ \max_{p,q,\{\rho_\theta\}_{\theta\in\mathcal{D}_j}} \quad \sum_{\theta \in \mathcal{D}_j} \hat{\nu}(\theta) \cdot p_\theta - L\, q \tag{9.6a}$$

$$\text{s.t.} \quad \sum_{\theta \in \mathcal{D}_j} \rho_\theta^\top \theta_k - \vec{x}_k \leq q, \quad \forall k \in [K], \tag{9.6b}$$

$$\sum_{\theta \in \mathcal{D}_j} p_\theta = 1,\ p \in \mathbb{R}^j_{\geq 0},\ q \geq 0, \tag{9.6c}$$

$$\vec{1}_M^\top \rho_\theta = p_\theta \vec{1}_M^\top,\ \rho_\theta \vec{1}_M = p_\theta \vec{1}_M,\ \rho_\theta \in \mathbb{R}^M_{\geq 0} \times \mathbb{R}^M_{\geq 0}, \quad \forall \theta \in \mathcal{D}_j. \tag{9.6d}$$

We interpret constraint (9.6b) in the component-wise sense.

Problem $\mathscr{P}^{\dagger}_{\mathsf{LP}}(x;\,\mathcal{D}_j)$ is always feasible and its optimal value is lower bounded, so we have $\mathsf{val}(\mathscr{P}^{\dagger}_{\mathsf{LP}}(x;\,\mathcal{D}_j)) = \mathsf{val}(D^\dagger(x;\mathcal{D}_j))$ for all $x \in \mathcal{X}$ by strong duality. The conclusion then follows from Proposition 8.

Theorem 6 shows that the permutation invariant upper level sets also have a polyhedral structure (with a polynomial number of variables and constraints). With the upper level sets in hand, our strategy for solving $\mathscr{P}(\mathcal{U}^\dagger)$ is analogous to the base case. First note $\mathscr{P}(\mathcal{U}^\dagger)$ is equivalent to

$$\max_{z \in \mathcal{Z},\, \upsilon \in \mathbb{R}} \{\upsilon : G(z) \in \mathcal{A}(\psi_{\mathcal{U}^\dagger}, \upsilon)\}, \tag{9.7}$$

where we want to find the largest value υ such that there exists $z \in \mathcal{Z}$ with $G(z) \in \mathcal{A}(\psi_{\mathcal{U}^\dagger}, \upsilon)$.

Fix $j \in [J]$, and introduce decision variables $\upsilon \in \mathbb{R}$, $z \in \mathcal{Z}$, $p \in \mathbb{R}^j$, $q \in \mathbb{R}$, and $\{\rho_\theta\}_{\theta \in \mathcal{D}_j}$ where $\rho_\theta \in \mathbb{R}^M \times \mathbb{R}^M$. We define $\vec{g}_k(z) = (g_k(z, w_m))_{m=1}^M$ to be the vector of components of $G(z)$ corresponding specifically to component $k \in [K]$ of $x(\omega_m)$ over all groups $m \in [M]$. Consider the following problem similar to $\mathscr{G}(\mathcal{D}_j)$:

$$\mathscr{G}^\dagger(\mathcal{D}_j):\ \max_{z,p,q,\rho,\upsilon} \quad \upsilon \tag{9.8a}$$

$$\text{s.t.} \quad \sum_{\theta \in \mathcal{D}_j} \hat{\nu}(\theta) \cdot p_\theta - L\, q \geq \upsilon, \tag{9.8b}$$

$$\sum_{\theta \in \mathcal{D}_j} \rho_\theta^\top \cdot \theta_k - \vec{g}_k(z) \leq q, \quad \forall k \in [K], \tag{9.8c}$$

$$\sum_{\theta \in \mathcal{D}_j} p_\theta = 1,\ p \in \mathbb{R}^j_{\geq 0},\ q \geq 0, \tag{9.8d}$$

$$\vec{1}_M^\top \rho_\theta = p_\theta \vec{1}_M^\top,\ \rho_\theta \vec{1}_M = p_\theta \vec{1}_M,\ \rho_\theta \in \mathbb{R}^M_{\geq 0} \times \mathbb{R}^M_{\geq 0}, \quad \forall \theta \in \mathcal{D}_j, \tag{9.8e}$$

which is a convex optimization problem.

We show in the following proposition that we can bound the optimal value of Problem (9.7) using $\mathscr{G}^\dagger(\mathcal{D}_j)$ for $j = \kappa(\upsilon)$. This result follows from Theorem 6 and the same argument as Proposition 6.

PROPOSITION 9. *Choose level* $\upsilon \in (-\infty, \upsilon_{\max}]$ *and* $j = \kappa(\upsilon)$*, then* $\max_{z\in\mathcal{Z}} \psi_{\mathcal{U}^\dagger}(G(z)) \geq \upsilon$ *if and only if* $\mathsf{val}(\mathscr{G}^\dagger(\mathcal{D}_j)) \geq \upsilon$.

We now present the details of our binary search algorithm for Problem (9.7). It mirrors Algorithm 1 except now we replace each instance of $\mathscr{G}(\mathcal{D}_j)$ with the modified problem $\mathscr{G}^\dagger(\mathcal{D}_j)$.

Algorithm 3: Binary search for $\mathscr{P}(\mathcal{U}^\dagger)$

Initialization: data sample $D = \{(\theta, \hat{v}(\theta))\}_{\theta\in\Theta}$, $j_1 = J$, $j_2 = 1$;

while $j_1 \neq j_2$ **do**

  Set $j := \lfloor \frac{j_1+j_2}{2} \rfloor$, and compute $\upsilon_j = \mathsf{val}(\mathscr{G}^\dagger(\mathcal{D}_j))$ with optimal solution z^*;

  if $\upsilon_j \leq \hat{v}(\theta_{j+1})$ **then** set $j_2 := j + 1$;

  else set $j_1 := j$;

end

Set $j := \lfloor \frac{j_1+j_2}{2} \rfloor$, and compute $\upsilon_j = \mathsf{val}(\mathscr{G}^\dagger(\mathcal{D}_j))$ with optimal solution z^*;

return z^* and $\psi_{\mathcal{U}^\dagger}(G(z^*)) = \min\{\upsilon_j, \hat{v}(\theta_j)\}$.

The complexity of Algorithm 3 for solving Problem (9.7) follows the same reasoning as Theorem 3.

THEOREM 7. *Algorithm 3 returns an optimal solution* z^* *of* $\mathscr{P}(\mathcal{U}^\dagger)$*, after solving at most* $\log J$ *instances of Problem* (9.8).

We note the same order of complexity in Theorem 7 that we saw in Theorem 3. Like the base case, the binary search algorithm for Problem (9.7) is highly scalable in both the size of the problem instance and the level J (which grows with the size of the dataset D). The difference between these two algorithms is reflected in the form of the feasibility problems that are solved for each level.

10. Proofs for Section 2

10.1. Proof of Proposition 1

For $z_1, z_2 \in \mathcal{Z}$ and $\lambda \in [0,1]$, we have:

$$\begin{aligned}
\rho_f(\lambda z_1 + (1-\lambda)z_2) &= f(G(\lambda z_1 + (1-\lambda)z_2)) \\
&\geq f(\lambda G(z_1) + (1-\lambda)G(z_2)) \\
&\geq \min\{\rho_f(z_1), \rho_f(z_2)\},
\end{aligned}$$

where the first inequality uses monotonicity of f and concavity of G, and the second inequality uses quasiconcavity of f.

11. Proofs for Section 4

11.1. Proof of Proposition 3

(i) This part is immediate by the definition of upper level set:

$$\sup\{v \in \mathbb{R} : x \in \mathcal{A}(f, \upsilon)\} = \sup\{v \in \mathbb{R} : f(x) \geq v\} = f(x).$$

(ii) We have the following equalities:

$$\begin{aligned}\mathcal{A}(\psi_{\mathcal{F}}, \upsilon) =& \{x \in \mathcal{X} : \psi_{\mathcal{F}}(x) \geq \upsilon\}\\ =& \{x \in \mathcal{X} : \inf_{f \in \mathcal{F}} f(x) \geq \upsilon\}\\ =& \{x \in \mathcal{X} : f(x) \geq \upsilon, \forall f \in \mathcal{F}\}\\ =& \cap_{f \in \mathcal{F}} \mathcal{A}(f, \upsilon),\end{aligned}$$

at level $\upsilon \in \mathbb{R}$.

11.2. Proof of Proposition 4

We have that $\psi_{\mathcal{U}}(x)$ is equal to the optimal value of:

$$\mathscr{P}(x; \mathcal{D}_J) : \min_{\upsilon, \xi} \quad \upsilon \tag{11.1a}$$

$$\text{s.t.} \quad \upsilon + \max\{\langle \xi, \theta - x \rangle, 0\} \geq \hat{v}(\theta), \forall \theta \in \mathcal{D}_J, \tag{11.1b}$$

$$\xi \geq 0, \|\xi\|_1 \leq L. \tag{11.1c}$$

Let $\upsilon^* = \mathsf{val}(\mathscr{P}(x; \mathcal{D}_J))$. Constraints (11.1b) are not binding for all θ with $\hat{v}(\theta) < \upsilon^*$. By definition of $\kappa(\upsilon^*)$, $\mathscr{P}(x; \mathcal{D}_{\kappa(\upsilon^*)})$ is then equivalent to $\mathscr{P}(x; \mathcal{D}_J)$ (it has the same optimal value and set of optimal solutions).

11.3. Proof of Proposition 5

First we have $\upsilon = \mathsf{val}(\mathscr{P}(x;\, \mathcal{D}_j))$ by Proposition 4. There are two cases to check: (i) $\hat{v}(\theta_{j+1}) < \upsilon < \hat{v}(\theta_j)$; and (ii) $\upsilon = \hat{v}(\theta_j)$.

For the first case, since $\upsilon < \hat{v}(\theta)$ for all $\theta \in \mathcal{D}_j$, we must have $\langle \xi, \theta - x \rangle > 0$ for all $\theta \in \mathcal{D}_j$. It follows that $\mathscr{P}(x;\, \mathcal{D}_j)$ and $\mathscr{P}_{LP}(x;\, \mathcal{D}_j)$ are equivalent, and $\upsilon = \mathsf{val}(\mathscr{P}_{LP}(x;\, \mathcal{D}_j))$. For the second case, we have $\mathsf{val}(\mathscr{P}_{LP}(x;\, \mathcal{D}_j)) \geq \mathsf{val}(\mathscr{P}(x;\, \mathcal{D}_j))$, and so $\mathsf{val}(\mathscr{P}_{LP}(x;\, \mathcal{D}_j)) \geq \hat{v}(\theta_j)$.

11.4. Proof of Theorem 2

Let $\mathscr{D}_{LP}(x; \mathcal{D}_j)$ denote the LP dual of $\mathscr{P}_{LP}(x;\, \mathcal{D}_j)$. Problem $\mathscr{P}_{LP}(x;\, \mathcal{D}_j)$ is automatically feasible, e.g., take $\upsilon = \max\{\hat{v}(\theta) : \theta \in \mathcal{D}_j\}$ and $\xi = 0$. Furthermore, under [Lip] the optimal value of $\mathscr{P}_{LP}(x;\, \mathcal{D}_j)$ is lower bounded by $\min\{\hat{v}(\theta) - L\|\theta - x\|_1 : \theta \in \mathcal{D}_j\} > -\infty$. Thus, LP strong duality holds between $\mathscr{P}_{LP}(x;\, \mathcal{D}_j)$ and $\mathscr{D}_{LP}(x;\, \mathcal{D}_j)$, and we have $\mathsf{val}(\mathscr{P}_{LP}(x;\, \mathcal{D}_j)) = \mathsf{val}(\mathscr{D}_{LP}(x;\, \mathcal{D}_j))$ for all $x \in \mathcal{X}$.

For $\hat{v}(\theta_{j+1}) < \upsilon \leq \hat{v}(\theta_j)$ and $j \in [J]$, by Proposition 5 we know that $x \in \mathcal{A}(\psi_{\mathcal{U}}, \upsilon)$ if and only if $\mathsf{val}(\mathscr{P}_{LP}(x;\, \mathcal{D}_j)) \geq \upsilon$. By strong duality, this is equivalent to the inequality $\mathsf{val}(\mathscr{D}_{LP}(x; \mathcal{D}_j)) \geq \upsilon$. However, since $\mathscr{D}_{LP}(x; \mathcal{D}_j)$ is a maximization problem, this latter inequality reduces to the feasibility problem:

$$\left\{(p, q) : \sum_{\theta \in \mathcal{D}_j} \hat{v}(\theta) p_\theta - L\, q \geq \upsilon,\ \sum_{\theta \in \mathcal{D}_j} \theta \cdot p_\theta - x \leq q \cdot \vec{1}_N,\ \sum_{\theta \in \mathcal{D}_j} p_\theta = 1,\ p \in \mathbb{R}^j_{\geq 0},\ q \geq 0\right\}.$$

Since $\left(\sum_{\theta\in\mathcal{D}_j}\hat{v}(\theta)\cdot p_\theta - \upsilon\right)/L \geq q$ and $\sum_{\theta\in\mathcal{D}_j}\theta\cdot p_\theta - x \leq q\cdot\vec{1}_N$, we can eliminate q from the above feasibility problem to obtain:

$$\left\{p : \sum_{\theta\in\mathcal{D}_j}\theta\cdot p_\theta - x \leq \left(\sum_{\theta\in\mathcal{D}_j}\hat{v}(\theta)\cdot p_\theta - \upsilon\right)/L\cdot\vec{1}_N,\ \sum_{\theta\in\mathcal{D}_j} p_\theta = 1,\ p\in\mathbb{R}^j_{\geq 0}\right\}.$$

We then have

$$\sum_{\theta\in\mathcal{D}_j}\theta\cdot p_\theta - \left(\sum_{\theta\in\mathcal{D}_j}\hat{v}(\theta)\cdot p_\theta - \upsilon\right)/L\cdot\vec{1}_N = \sum_{\theta\in\mathcal{D}_j}(\theta - \hat{v}(\theta)/L)\,p_\theta + \upsilon/L\cdot\vec{1}_N,$$

and the desired result follows from the definition of $\tilde{\theta} = \theta - \hat{v}(\theta)/L$.

12. Additional Results for Non-Monotone Functions

We can extend our results from Sections 4 and 5 to cover non-monotone quasiconcave functions. Let $\mathcal{F}_{\text{NQCo}} \subset \mathfrak{F}$ be the set of all quasiconcave functions $f : \mathcal{X}\to\mathbb{R}$ (which are not necessarily monotone). We modify the definition of our ambiguity set to: $\mathcal{U} = \mathcal{U}(D, L) = \{f\in\mathcal{F}_{\text{NQCo}}\cap\mathcal{F}_{\text{Lip}}(L) : f(\theta)\geq\hat{v}(\theta),\ \forall\theta\in\Theta\}$. For this setting, the non-monotone counterpart of $\mathscr{P}_{LP}(x;\,\mathcal{D}_j)$ is obtained by simply dropping the constraint $\xi\geq 0$ to get:

$$\begin{aligned}
\mathscr{P}_{LP}(x;\,\mathcal{D}_j) : \min_{\upsilon\in\mathbb{R},\,\xi\in\mathbb{R}^N} \quad & \upsilon && (12.1a)\\
\text{s.t.} \quad & \upsilon + \langle\xi, \theta - x\rangle \geq \hat{v}(\theta),\ \forall\theta\in\mathcal{D}_j, && (12.1b)\\
& \|\xi\|_1 \leq L. && (12.1c)
\end{aligned}$$

We now use the above form of $\mathscr{P}_{LP}(x;\,\mathcal{D}_j)$ to compute $\psi_{\mathcal{U}}(x)$ for fixed $x\in\mathcal{X}$. The dual to Problem $\mathscr{P}_{LP}(x;\,\mathcal{D}_j)$ is:

$$\begin{aligned}
\mathscr{D}_{LP}(x;\,\mathcal{D}_j) : \quad \max_{\{p_\theta\}_{\theta\in\mathcal{D}_j},\,t\in\mathbb{R}_{\geq 0}} \quad & \sum_{\theta\in\mathcal{D}_j} p_\theta\,\hat{v}(\theta) - Lt\\
\text{s.t.} \quad & -t\vec{1}_N \leq \sum_{\theta\in\mathcal{D}_j} p_\theta\,(\theta - x) \leq t\vec{1}_N,\\
& \sum_{\theta\in\mathcal{D}_j} p_\theta = 1,\\
& p_\theta \geq 0,\quad \forall\theta\in\mathcal{D}_j.
\end{aligned}$$

Then, the constraint $\mathsf{val}(\mathscr{D}_{LP}(x;\mathcal{D}_j))\geq\upsilon$ can be reformulated as the system of inequalities:

$$\begin{aligned}
\sum_{\theta\in\mathcal{D}_j} p_\theta\,\hat{v}(\theta) - Lt \geq \upsilon,\\
-t\vec{1}_N \leq \sum_{\theta\in\mathcal{D}_j} p_\theta\,(\theta - x) \leq t\vec{1}_N,\\
\sum_{\theta\in\mathcal{D}_j} p_\theta = 1,\\
p_\theta \geq 0,\quad \forall\theta\in\mathcal{D}_j,
\end{aligned}$$

which determine a convex (polyhedral) set. Eliminating the dual variable t, we obtain the non-monotone counterpart of Problem $\mathcal{G}(\mathcal{D}_j)$ which is

$$\begin{aligned}
\mathcal{G}(\mathcal{D}_j): \max_{z\in\mathcal{Z},\, \upsilon\in\mathbb{R},\, p\in\mathbb{R}^j_{\geq 0}} \quad & \upsilon \\
\text{s.t.} \quad & G(z) \geq \sum_{\theta\in\mathcal{D}_j} \left(\theta - (\hat{v}(\theta)/L)\vec{1}_N\right)\cdot p_\theta + (\upsilon/L)\vec{1}_N, \\
& G(z) \leq \sum_{\theta\in\mathcal{D}_j} \left(\theta + (\hat{v}(\theta)/L)\vec{1}_N\right)\cdot p_\theta - (\upsilon/L)\vec{1}_N, \\
& \sum_{\theta\in\mathcal{D}_j} p_\theta = 1.
\end{aligned}$$

Binary search (analogous to Algorithm 1) can then be applied to compute the robust optimal solution using the modified $\mathcal{G}(\mathcal{D}_j)$.

13. Proofs for Section 5

13.1. Proof of Proposition 6

For any $\upsilon \in (-\infty, \upsilon_{\max}]$ and $j \in [J]$, we define the feasibility problem:

$$\begin{aligned}
(\mathcal{F}_\upsilon(\mathcal{D}_j)) \quad \text{Find} \quad & z\in\mathcal{Z},\, p\in\mathbb{R}^j_{\geq 0} \\
\text{s.t.} \quad & G(z) \geq \sum_{\theta\in\mathcal{D}_j} \tilde{\theta}\cdot p_\theta + (\upsilon/L)\vec{1}_N, \\
& \sum_{\theta\in\mathcal{D}_j} p_\theta = 1,
\end{aligned}$$

in the variables (z, p). By Theorem 2, feasibility of the inequality $\psi_{\mathcal{U}}(G(z)) \geq \nu$ for $z \in \mathcal{Z}$ is equivalent to feasibility of $\mathcal{F}_\upsilon(\mathcal{D}_j)$ for $j = \kappa(\upsilon)$. We then have the following chain of equivalences:

$$\begin{aligned}
\{\exists z\in\mathcal{Z},\, \psi_{\mathcal{U}}(G(z)) \geq \upsilon\} &\iff \left\{\mathcal{F}_\upsilon(\mathcal{D}_j) \text{ feasible}\right\}, \\
\left\{\max_{z\in\mathcal{Z}} \psi_{\mathcal{U}}(G(z)) \geq \upsilon\right\} &\iff \left\{\max\left\{\upsilon' \mid \mathcal{F}_{\upsilon'}(\mathcal{D}_j) \text{ feasible}\right\} \geq \upsilon\right\}, \\
\left\{\max_{z\in\mathcal{Z}} \psi_{\mathcal{U}}(G(z)) \geq \upsilon\right\} &\iff \{\mathsf{val}(\mathcal{G}(\mathcal{D}_j)) \geq \upsilon\},
\end{aligned}$$

using the fact that $\mathcal{G}(\mathcal{D}_j)$ is equivalent to $\max\left\{\upsilon' \mid \mathcal{F}_{\upsilon'}(\mathcal{D}_j) \text{ feasible}\right\}$.

13.2. Proof of Theorem 3

Let ν denote the optimal value of Problem (8). By Proposition 6, $\nu \geq \hat{v}(\theta_j)$ if and only if $\mathsf{val}(\mathcal{G}(\mathcal{D}_j)) \geq \hat{v}(\theta_j)$. Hence the binary search procedure in Algorithm 1 finds $\zeta \in [J]$ such that $\hat{v}(\theta_{\zeta+1}) < \nu \leq \hat{v}(\theta_\zeta)$ within $O(\log J)$ iterations.

To see this, note that for any $j \leq \zeta - 1$, we have $\nu \leq \hat{v}(\theta_\zeta) \leq \hat{v}(\theta_{j+1})$. That is, by Proposition 6 we have $\mathsf{val}(\mathcal{G}(\mathcal{D}_{j+1})) \leq \hat{v}(\theta_{j+1})$. Therefore, $\mathsf{val}(\mathcal{G}(\mathcal{D}_j)) \leq \hat{v}(\theta_{j+1})$ by the monotonicity of $\mathsf{val}(\mathcal{G}(\mathcal{D}_j))$ in j by Observation 1. On the other hand, for any $j \geq \zeta + 1$, we have $\nu > \hat{v}(\theta_{\zeta+1}) \geq \hat{v}(\theta_j)$. Therefore, by Proposition 6 we have $\mathsf{val}(\mathcal{G}(\mathcal{D}_j)) > \hat{v}(\theta_j) \geq \hat{v}(\theta_{j+1})$. This proves that binary search returns the correct index ζ. Let $(\tilde{z}^*, \tilde{p}^*, \tilde{v}^*)$ be the optimal solution of $\mathcal{G}(\mathcal{D}_\zeta)$ for this ζ. We want to show that $\tilde{z}^*$ is an optimal solution of Problem (8).

First, let $\upsilon^* = \psi_{\mathcal{U}}(G(z^*))$ by Proposition 6 so $\nu \geq \upsilon^*$. Suppose $\nu > \upsilon^*$, then there is some $\tilde{z} \in \mathcal{Z}$ with $\psi_{\mathcal{U}}(G(\tilde{z})) = \nu > \upsilon^*$. It follows that $(\tilde{z}, \tilde{v}, \tilde{p})$ are feasible for $\mathcal{G}(\mathcal{D}_{\tilde{\zeta}})$ for $\tilde{\zeta} = \kappa(\nu) \leq \zeta$. However, this contradicts optimality of $(\tilde{z}^*, \tilde{p}^*, \tilde{v}^*)$ being the optimal solution of $\mathcal{G}(\mathcal{D}_\zeta)$, since any feasible solution for $\mathcal{G}(\mathcal{D}_{\zeta'})$ for $\zeta' \leq \zeta$ can be extended to a feasible solution for $\mathcal{G}(\mathcal{D}_\zeta)$ with the same optimal value.