Electronic Companion: A Unified Theory of Robust and Distributionally Robust Optimization via the Primal-Worst-Equals-Dual-Best Principle

Jianzhe Zhen[1], Daniel Kuhn[2], and Wolfram Wiesemann[3]

[1]*School of Economics and Management, University of Chinese Academy of Sciences, China*
[2]*College of Management of Technology, École Polytechnique Fédérale de Lausanne, Switzerland*
[3]*Imperial College Business School, Imperial College London, United Kingdom*

July 18, 2023

A Basic Concepts of Convex Analysis

Throughout the paper we use the following key concepts of convex analysis. The *domain* of a function $f : \mathbb{R}^{d_{\boldsymbol{x}}} \to \overline{\mathbb{R}}$ is defined as $\mathrm{dom}(f) = \{\boldsymbol{x} \in \mathbb{R}^{d_{\boldsymbol{x}}} \mid f(\boldsymbol{x}) < +\infty\}$. The *epigraph* of f is defined as $\mathrm{epi}(f) = \{(\boldsymbol{x}, \tau) \in \mathbb{R}^{d_{\boldsymbol{x}}} \times \mathbb{R} \mid f(\boldsymbol{x}) \leq \tau\}$. The function f is *proper* if $f(\boldsymbol{x}) > -\infty$ for all $\boldsymbol{x} \in \mathbb{R}^{d_{\boldsymbol{x}}}$ and $f(\boldsymbol{x}) < +\infty$ for at least one $\boldsymbol{x} \in \mathbb{R}^{d_{\boldsymbol{x}}}$, implying that $\mathrm{dom}(f) \neq \emptyset$. In addition, f is *closed* if f is lower semicontinuous and either $f(\boldsymbol{x}) > -\infty$ for all $\boldsymbol{x} \in \mathbb{R}^{d_{\boldsymbol{x}}}$ or $f(\boldsymbol{x}) = -\infty$ for all $\boldsymbol{x} \in \mathbb{R}^{d_{\boldsymbol{x}}}$.

We now define the notions of conjugate functions and perspective functions.

Definition A.1 (Conjugate Function)**.** *The conjugate of a function $f : \mathbb{R}^{d_{\boldsymbol{x}}} \to \overline{\mathbb{R}}$ is the function $f^* : \mathbb{R}^{d_{\boldsymbol{x}}} \to \overline{\mathbb{R}}$ defined through $f^*(\boldsymbol{w}) = \sup_{\boldsymbol{x}} \left\{\boldsymbol{x}^\top \boldsymbol{w} - f(\boldsymbol{x})\right\}$. The conjugate $(f^*)^*$ of f^* is called the biconjugate of f and is abbreviated as f^{**}.*

The *indicator function* $\delta_{\mathcal{X}} : \mathbb{R}^{d_{\boldsymbol{x}}} \to \overline{\mathbb{R}}$ of a set $\mathcal{X} \subseteq \mathbb{R}^{d_{\boldsymbol{x}}}$ is defined through $\delta_{\mathcal{X}}(\boldsymbol{x}) = 0$ if $\boldsymbol{x} \in \mathcal{X}$ and $\delta_{\mathcal{X}}(\boldsymbol{x}) = +\infty$ if $\boldsymbol{x} \notin \mathcal{X}$. The *support function* $\delta^*_{\mathcal{X}} : \mathbb{R}^{d_{\boldsymbol{x}}} \to \overline{\mathbb{R}}$ of a set $\mathcal{X} \subseteq \mathbb{R}^{d_{\boldsymbol{x}}}$ is defined through $\delta^*_{\mathcal{X}}(\boldsymbol{w}) = \sup_{\boldsymbol{x} \in \mathcal{X}} \{\boldsymbol{x}^\top \boldsymbol{w}\}$. Note that the support function of $\mathcal{X}$ coincides with the conjugate of the indicator function of $\mathcal{X}$, which justifies our notation.

Definition A.2 (Perspective Functions)**.** *The convex perspective of a proper, closed and convex function* $f : \mathbb{R}^{d_{\boldsymbol{x}}} \to \overline{\mathbb{R}}$ *is the function* $\underline{f} : \mathbb{R}^{d_{\boldsymbol{x}}} \times \mathbb{R}_+ \to \overline{\mathbb{R}}$ *defined through* $\underline{f}(\boldsymbol{x}, t) = tf(\boldsymbol{x}/t)$ *if* $t > 0$ *and* $\underline{f}(\boldsymbol{x}, 0) = \delta^*_{\mathrm{dom}(f^*)}(\boldsymbol{x})$*. Similarly, the concave perspective of a function* f *for which* $-f$ *is proper, closed and convex is the function* $\overline{f} : \mathbb{R}^{d_{\boldsymbol{x}}} \times \mathbb{R}_+ \to \overline{\mathbb{R}}$ *defined through* $\overline{f}(\boldsymbol{x}, t) = tf(\boldsymbol{x}/t)$ *if* $t > 0$ *and* $\overline{f}(\boldsymbol{x}, 0) = -\delta^*_{\mathrm{dom}((-f)^*)}(\boldsymbol{x})$*.*

One can show that for $t > 0$, the epigraph of $\underline{f}(\cdot, t)$ coincides with the epigraph of f multiplied by t. Moreover, the epigraph of $\underline{f}$ coincides with the closure of the cone generated by $\mathrm{epi}(f) \times \{1\} \subseteq \mathbb{R}^{d_{\boldsymbol{x}}} \times \mathbb{R}$. Finally, our definitions of the convex and concave perspectives satisfy

$$\underline{f}(\boldsymbol{x}, 0) = \liminf_{(\boldsymbol{x}', t') \to (\boldsymbol{x}, 0)} t' f\left(\boldsymbol{x}'/t'\right) \qquad \text{and} \qquad \overline{f}(\boldsymbol{x}, 0) = \limsup_{(\boldsymbol{x}', t') \to (\boldsymbol{x}, 0)} t' f\left(\boldsymbol{x}'/t'\right) \tag{A.1}$$

for convex and concave f, respectively, and $\boldsymbol{x} \in \mathbb{R}^{d_{\boldsymbol{x}}}$ (Rockafellar, 1970, p. 67 and Theorem 13.3). For ease of notation, we henceforth use $tf(\boldsymbol{x}/t)$ to denote both $\underline{f}(\boldsymbol{x}, t)$ and $\overline{f}(\boldsymbol{x}, t)$. The correct interpretation of $0f(\boldsymbol{x}/0)$ will be clear from the context. Specifically, $0f(\boldsymbol{x}/0)$ should be interpreted as $\underline{f}(\boldsymbol{x}, 0)$ if f is convex and as $\overline{f}(\boldsymbol{x}, 0)$ if f is concave. This convention is justified in view of (A.1).

By construction, the convex perspective of a proper, closed and convex function is guaranteed to be proper, closed and convex; see Proposition C.2. The next example shows that alternative constructions of the convex perspective that are sometimes adopted in the literature fail to be closed.

Example A.1 (Perspective Functions)**.** *Define* $f : \mathbb{R} \to \overline{\mathbb{R}}$ *through* $f(x) = \delta_{\{x_0\}}(x)$ *for some* $x_0 \neq 0$*, and note that* f *is proper, closed and convex. An elementary calculation shows that the convex perspective of* f *is given by* $\underline{f}(x, t) = \delta_{\{tx_0\}}(x)$ *for every* $x \in \mathbb{R}$ *and* $t \geq 0$*, which is also proper, closed and convex. Note, however, that* $\underline{f}(x, 0) = \delta_{\{0\}}(x) \neq +\infty = \lim_{t \downarrow 0} tf(x/t)$*, where the last equality holds because* $x_0 \neq 0$*. This example shows that if one were to define* $\underline{f}(x, 0) = \lim_{t \downarrow 0} tf(x/t)$*, as is sometimes done in the literature, then the resulting perspective would fail to be closed. As a second example, define* $f : \mathbb{R} \to \overline{\mathbb{R}}$ *through* $f(x) = |x|$*, and note that* f *is again proper, closed and convex. The convex perspective of* f *is given by* $\underline{f}(x, t) = |x|$ *for every* $x \in \mathbb{R}$ *and* $t \geq 0$*, which is also proper, closed and convex. This example shows that if one were to define* $\underline{f}(x, 0) = \delta_{\{0\}}(x)$*, as is sometimes done in the literature, then the resulting perspective would fail to be closed.*

Throughout the paper we use the following terminology for optimization problems, which is in line with Rockafellar (1970). Any assignment of real values to the decision variables of an optimization problem is a solution. A solution is feasible in an optimization problem if it satisfies all the constraints and attains an objective value other than $+\infty$ ($-\infty$) in a minimization (maximization) problem; otherwise, it is infeasible. An optimization problem is feasible if it has at least one feasible solution. We refer to the feasible region of an optimization problem as the set containing all of its feasible solutions. A feasible optimization problem is solvable if its optimal value is attained by a feasible solution, whereas an infeasible optimization problem is solved by any (necessarily infeasible) solution. Whenever the domain of a variable in an optimization problem is omitted, it is understood to be the entire space (whose definition will be clear from the context).

In Section 3 we study functions $f(\boldsymbol{x}, \boldsymbol{z})$ with two arguments, where the first argument $\boldsymbol{x}$ represents a decision variable, while the second argument $\boldsymbol{z}$ represents an exogenous uncertain parameter. Below we show how the notions of conjugates and perspectives are extended to such functions.

Definition A.3 (Partial Conjugates). *The partial conjugate of a function $f : \mathbb{R}^{d_{\boldsymbol{x}}} \times \mathbb{R}^{d_{\boldsymbol{z}}} \to \overline{\mathbb{R}}$ with respect to its first argument is the function $f^{*1} : \mathbb{R}^{d_{\boldsymbol{x}}} \times \mathbb{R}^{d_{\boldsymbol{z}}} \to \overline{\mathbb{R}}$ defined through $f^{*1}(\boldsymbol{w}, \boldsymbol{z}) = \sup_{\boldsymbol{x} \in \mathbb{R}^{d_{\boldsymbol{x}}}} \left\{ \boldsymbol{w}^\top \boldsymbol{x} - f(\boldsymbol{x}, \boldsymbol{z}) \right\}$. Likewise, the partial conjugate of f with respect to its second argument is the function $f^{*2} : \mathbb{R}^{d_{\boldsymbol{x}}} \times \mathbb{R}^{d_{\boldsymbol{z}}} \to \overline{\mathbb{R}}$ defined through $f^{*2}(\boldsymbol{x}, \boldsymbol{y}) = \sup_{\boldsymbol{z} \in \mathbb{R}^{d_{\boldsymbol{z}}}} \left\{ \boldsymbol{y}^\top \boldsymbol{z} - f(\boldsymbol{x}, \boldsymbol{z}) \right\}$.*

Definition A.4 (Partial Perspectives). *If $f : \mathbb{R}^{d_{\boldsymbol{x}}} \times \mathbb{R}^{d_{\boldsymbol{z}}} \to \overline{\mathbb{R}}$ is proper, closed and convex in its first argument, then we define its convex partial perspective $\underline{f} : \mathbb{R}^{d_{\boldsymbol{x}}} \times \mathbb{R}_+ \times \mathbb{R}^{d_{\boldsymbol{z}}} \to \overline{\mathbb{R}}$ through $\underline{f}(\boldsymbol{x}, t, \boldsymbol{z}) = t f(\boldsymbol{x}/t, \boldsymbol{z})$ if $t > 0$ and $\underline{f}(\boldsymbol{x}, 0, \boldsymbol{z}) = \delta^*_{\mathrm{dom}(f^{*1}(\cdot, \boldsymbol{z}))}(\boldsymbol{x})$. If $-f$ is proper, closed and convex in its second argument, then we define its concave partial perspective $\overline{f} : \mathbb{R}^{d_{\boldsymbol{x}}} \times \mathbb{R}^{d_{\boldsymbol{z}}} \times \mathbb{R}_+ \to \overline{\mathbb{R}}$ through $\overline{f}(\boldsymbol{x}, \boldsymbol{z}, t) = t f(\boldsymbol{x}, \boldsymbol{z}/t)$ if $t > 0$ and $\overline{f}(\boldsymbol{x}, \boldsymbol{z}, 0) = -\delta^*_{\mathrm{dom}((-f)^{*2}(\boldsymbol{x}, \cdot))}(\boldsymbol{z})$.*

For ease of notation, throughout the paper we use $t f(\boldsymbol{x}/t, \boldsymbol{z})$ and $t f(\boldsymbol{x}, \boldsymbol{z}/t)$ to denote $\underline{f}(\boldsymbol{x}, t, \boldsymbol{z})$ and $\overline{f}(\boldsymbol{x}, \boldsymbol{z}, t)$, respectively. The correct interpretation will always be clear from the context.

Section 5 makes extensive use of proper convex cones, which we define next.

Definition A.5 (Proper Convex Cone). *A convex cone $\mathcal{C} \subseteq \mathbb{R}^{d_{\mathcal{C}}}$ is called proper if it is closed, solid (i.e., it has nonempty interior) and pointed (i.e., it contains no line).*

Section 5 also uses a generalized Slater condition for sets represented by conic inequalities.

Definition A.6 (Slater Condition for Sets). *The vector $\boldsymbol{x}^{\mathrm{S}}$ is a Slater point of the set $\mathcal{X}$ represented by $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^{d_{\boldsymbol{x}}} \mid \boldsymbol{f}_i(\boldsymbol{x}) \preceq_{\mathcal{C}_i} \boldsymbol{0} \;\; \forall i \in \mathcal{I},\; h_j(\boldsymbol{x}) = 0 \;\; \forall j \in \mathcal{J}\}$, where $\mathcal{C}_i$ is a proper convex cone for all $i \in \mathcal{I}$, if* (i) *$\boldsymbol{x}^{\mathrm{S}} \in \mathrm{ri}(\mathrm{dom}(\boldsymbol{f}_i))$ and $\boldsymbol{x}^{\mathrm{S}} \in \mathrm{ri}(\mathrm{dom}(h_j))$ for all i and j;* (ii) *$\boldsymbol{x}^{\mathrm{S}} \in \mathcal{X}$; and*

(iii) $\boldsymbol{f}_i(\boldsymbol{x}^{\mathrm{S}}) \prec_{\mathcal{C}_i} \mathbf{0}$ *for all* $i \in \mathcal{I}$ *with the exception of those for which* $\boldsymbol{f}_i$ *is affine and* $\mathcal{C}_i$ *is the non-negative orthant. The Slater point* $\boldsymbol{x}^{\mathrm{S}}$ *is strict if* $\boldsymbol{f}_i(\boldsymbol{x}^{\mathrm{S}}) \prec_{\mathcal{C}_i} \mathbf{0}$ *for all* $i \in \mathcal{I}$.

We close with an example that describes vector- and matrix-valued functions that are convex with respect to proper convex cones but have components that fail to be convex in the usual sense.

Example A.2 ($\mathcal{C}$-Convex Functions)**.** *The set* $\mathbb{S}^n_+$ *of all positive semidefinite matrices represents a proper convex cone in the space* $\mathbb{S}^n$ *of symmetric* $n \times n$*-matrices, and its interior is given by the set* $\mathbb{S}^n_{++}$ *of positive definite matrices. The matrix inversion* $\boldsymbol{F} : \mathbb{S}^n \to \mathbb{S}^n \cup \{+\infty_{\mathbb{S}^n_+}\}$ *defined through*

$$\boldsymbol{F}(\boldsymbol{X}) = \begin{cases} \boldsymbol{X}^{-1} & \text{if } \boldsymbol{X} \in \mathbb{S}^n_{++} \\ +\infty_{\mathbb{S}^n_+} & \text{otherwise} \end{cases}$$

is an example of an $\mathbb{S}^n_+$*-convex function. To see this, note that* $\mathrm{dom}(\boldsymbol{F}) = \mathbb{S}^n_{++}$ *is convex and that the* $\mathbb{S}^n_+$*-epigraph of* $\boldsymbol{F}$ *can be represented as*

$$\mathrm{epi}_{\mathbb{S}^n_+}(\boldsymbol{F}) = \left\{ (\boldsymbol{X}, \boldsymbol{Y}) \in \mathbb{S}^n_{++} \times \mathbb{S}^n \,\middle|\, \boldsymbol{X}^{-1} \preceq_{\mathbb{S}^n_+} \boldsymbol{Y} \right\} = \left\{ (\boldsymbol{X}, \boldsymbol{Y}) \in \mathbb{S}^n_{++} \times \mathbb{S}^n \,\middle|\, \begin{pmatrix} \boldsymbol{X} & \boldsymbol{I}_n \\ \boldsymbol{I}_n & \boldsymbol{Y} \end{pmatrix} \succeq_{\mathbb{S}^{2n}_+} \mathbf{0} \right\},$$

where $\boldsymbol{I}_n$ *stands for the identity matrix in* $\mathbb{S}^n$*. The second equality in the above expression follows from a standard Schur complement argument. Thus,* $\mathrm{epi}_{\mathbb{S}^n_+}(\boldsymbol{F})$ *is manifestly convex. Other* $\mathcal{C}$*-convex functions can be constructed as follows. If* $\mathcal{C} \subseteq \mathbb{R}^{d_{\mathcal{C}}}$ *is a proper convex cone,* $g(\boldsymbol{x}, \boldsymbol{y})$ *is a Borel-measurable function that is convex in* $\boldsymbol{x}$ *for every fixed* $\boldsymbol{y} \in \mathcal{C}$ *and* μ *is a Borel measure on* $\mathcal{C}$*, then* $\boldsymbol{f}(\boldsymbol{x}) = \int_{\mathcal{C}} \boldsymbol{y} \cdot g(\boldsymbol{x}, \boldsymbol{y})\, \mu(\mathrm{d}\boldsymbol{y})$ *is* $\mathcal{C}$*-convex (provided the integral exists) because* $\boldsymbol{\lambda}^\top \boldsymbol{f}$ *is convex for every* $\boldsymbol{\lambda} \in \mathcal{C}^* \backslash \{\mathbf{0}\}$*. To see this, recall that* $\boldsymbol{\lambda}^\top \boldsymbol{y} \geq 0$ *for all* $\boldsymbol{y} \in \mathcal{C}$ *and that convexity is preserved by integration against a non-negative weighting function (Boyd and Vandenberghe, 2004, § 3.2.1).*

B Proofs

Proof of Theorem 1. If (P) or (D) is infeasible, the statement trivially holds. In the remainder of the proof, we thus assume that both (P) and (D) are feasible. Set $\mathcal{C} = \cap_{i \in \mathcal{I}_0} \mathrm{dom}(f_i)$, which is nonempty and convex by the feasibility of (P) and assumption **(F)**, respectively. However, $\mathcal{C}$ is not necessarily closed because not every proper, closed and convex function has a closed domain.

Next, define the Lagrangian $\mathscr{L}:\mathbb{R}^{d_{\boldsymbol{x}}}\times\mathbb{R}^I\to\overline{\mathbb{R}}$ associated with problem (P) through

$$\mathscr{L}(\boldsymbol{x},\boldsymbol{\lambda})=\begin{cases} f_0(\boldsymbol{x})+\sum_{i\in\mathcal{I}}\lambda_i f_i(\boldsymbol{x}) & \text{if } \boldsymbol{x}\in\mathcal{C},\ \boldsymbol{\lambda}\geq\mathbf{0},\\ -\infty & \text{if } \boldsymbol{x}\in\mathcal{C},\ \boldsymbol{\lambda}\not\geq\mathbf{0},\\ +\infty & \text{otherwise.}\end{cases}$$

As the objective and constraint functions of problem (P) are proper and convex by assumption **(F)**, the Lagrangian $\mathscr{L}(\boldsymbol{x},\boldsymbol{\lambda})$ is proper and convex in $\boldsymbol{x}$ for every fixed $\boldsymbol{\lambda}\geq\mathbf{0}$. As $\mathcal{C}$ may fail to be closed, however, $\mathscr{L}(\boldsymbol{x},\boldsymbol{\lambda})$ is not necessarily closed in $\boldsymbol{x}$ even if $\boldsymbol{\lambda}\geq\mathbf{0}$. One also easily verifies that $-\mathscr{L}(\boldsymbol{x},\boldsymbol{\lambda})$ is proper, closed and convex in $\boldsymbol{\lambda}$ for every fixed $\boldsymbol{x}\in\mathcal{C}$.

Using the Lagrangian, the primal problem (P) can be expressed as the min-max problem

$$\inf_{\boldsymbol{x}\in\mathcal{C}} f(\boldsymbol{x}),\quad \text{where}\quad f(\boldsymbol{x})=\sup_{\boldsymbol{\lambda}\geq\mathbf{0}}\mathscr{L}(\boldsymbol{x},\boldsymbol{\lambda})=\begin{cases} f_0(\boldsymbol{x}) & \text{if } f_i(\boldsymbol{x})\leq 0\ \ \forall i\in\mathcal{I},\\ +\infty & \text{otherwise.}\end{cases}\tag{B.1}$$

Below we will show that the dual problem (D) can be bounded above by the max-min problem

$$\sup_{\boldsymbol{\lambda}\geq\mathbf{0}} g(\boldsymbol{\lambda}),\quad \text{where}\quad g(\boldsymbol{\lambda})=\inf_{\boldsymbol{x}\in\mathcal{C}}\mathscr{L}(\boldsymbol{x},\boldsymbol{\lambda}).\tag{B.2}$$

The statement of the theorem then follows because

$$\inf\,(\mathrm{P}) \;=\; \inf_{\boldsymbol{x}\in\mathcal{C}} f(\boldsymbol{x}) \;=\; \inf_{\boldsymbol{x}\in\mathcal{C}}\sup_{\boldsymbol{\lambda}\geq\mathbf{0}}\mathscr{L}(\boldsymbol{x},\boldsymbol{\lambda}) \;\geq\; \sup_{\boldsymbol{\lambda}\geq\mathbf{0}}\inf_{\boldsymbol{x}\in\mathcal{C}}\mathscr{L}(\boldsymbol{x},\boldsymbol{\lambda}) \;=\; \sup_{\boldsymbol{\lambda}\geq\mathbf{0}} g(\boldsymbol{\lambda}) \;\geq\; \sup\,(\mathrm{D}),\tag{B.3}$$

where the first inequality is a direct consequence of the classical min-max inequality. To see that the second inequality holds, fix any $\boldsymbol{\lambda}\geq\mathbf{0}$ and note that

$$\begin{aligned}
g(\boldsymbol{\lambda}) &= \inf_{\boldsymbol{x}}\left\{ f_0(\boldsymbol{x})+\sum_{\substack{i\in\mathcal{I}:\\ \lambda_i>0}}\lambda_i f_i(\boldsymbol{x})+\sum_{\substack{i\in\mathcal{I}:\\ \lambda_i=0}}\delta_{\mathrm{dom}(f_i)}(\boldsymbol{x})\right\}\\
&= -\sup_{\boldsymbol{x}}\left\{ \mathbf{0}^\top\boldsymbol{x}-f_0(\boldsymbol{x})-\sum_{\substack{i\in\mathcal{I}:\\ \lambda_i>0}}\lambda_i f_i(\boldsymbol{x})-\sum_{\substack{i\in\mathcal{I}:\\ \lambda_i=0}}\delta_{\mathrm{dom}(f_i)}(\boldsymbol{x})\right\}\\
&\geq -\inf_{\{\boldsymbol{w}_i\}_{i\in\mathcal{I}_0}}\left\{ f_0^*(\boldsymbol{w}_0)+\sum_{\substack{i\in\mathcal{I}:\\ \lambda_i>0}}(\lambda_i f_i)^*(\boldsymbol{w}_i)+\sum_{\substack{i\in\mathcal{I}:\\ \lambda_i=0}}\delta^*_{\mathrm{dom}(f_i)}(\boldsymbol{w}_i)\;\middle|\;\sum_{i\in\mathcal{I}_0}\boldsymbol{w}_i=\mathbf{0}\right\}\\
&= -\inf_{\{\boldsymbol{w}_i\}_{i\in\mathcal{I}_0}}\left\{ f_0^*(\boldsymbol{w}_0)+\sum_{\substack{i\in\mathcal{I}:\\ \lambda_i>0}}\lambda_i f_i^*(\boldsymbol{w}_i/\lambda_i)+\sum_{\substack{i\in\mathcal{I}:\\ \lambda_i=0}}\delta^*_{\mathrm{dom}(f_i)}(\boldsymbol{w}_i)\;\middle|\;\sum_{i\in\mathcal{I}_0}\boldsymbol{w}_i=\mathbf{0}\right\}\\
&= \sup_{\{\boldsymbol{w}_i\}_{i\in\mathcal{I}_0}}\left\{ -f_0^*(\boldsymbol{w}_0)-\sum_{i\in\mathcal{I}}\lambda_i f_i^*(\boldsymbol{w}_i/\lambda_i)\;\middle|\;\sum_{i\in\mathcal{I}_0}\boldsymbol{w}_i=\mathbf{0}\right\},
\end{aligned}\tag{B.4}$$

where the first equality expresses $g(\boldsymbol{\lambda})$ as the optimal value of an unconstrained minimization problem, in which any solution $\boldsymbol{x} \notin \mathcal{C}$ adopts an infinite objective value. The objective function of this minimization problem is proper because $\mathcal{C} \neq \emptyset$, and it is convex because the functions f_i, $i \in \mathcal{I}_0$, are all convex. The inequality in the above expression follows from Proposition C.4, which asserts that the conjugate of a sum of proper convex functions provides a lower bound on the infimal convolution of the conjugates of these functions. The third equality follows from Theorem 16.1 by Rockafellar (1970), which asserts that the conjugate of a positive multiple of a proper convex function equals the perspective of the conjugate of this function. The fourth equality holds due to our definition of the convex perspective and because $\delta^*_{\mathrm{dom}(f_i)} = \delta^*_{\mathrm{dom}(f_i^{**})}$ by virtue of Theorem 12.2 by Rockafellar (1970). Substituting the lower bound (B.4) on $g(\boldsymbol{\lambda})$ into (B.2) then shows that the dual problem (D) is indeed bounded above by (B.2). This observation completes the proof. $\square$

The proof of Theorem 2 relies on the following auxiliary result.

Lemma B.1. *The Lagrangian dual of problem* (D) *is equivalent to problem* (P).

Proof of Lemma B.1. We express (D) as the max-min problem

$$\sup_{\boldsymbol{\lambda} \geq \mathbf{0}} \sup_{\{\boldsymbol{w}_i\}_{i \in \mathcal{I}_0}} \inf_{\boldsymbol{x}} -f_0^*(\boldsymbol{w}_0) - \sum_{i \in \mathcal{I}} \lambda_i f_i^*(\boldsymbol{w}_i/\lambda_i) + \sum_{i \in \mathcal{I}_0} \boldsymbol{w}_i^\top \boldsymbol{x}, \tag{B.5}$$

where the equality constraint involving $\{\boldsymbol{w}_i\}_{i \in \mathcal{I}_0}$ is enforced implicitly through the embedded minimization over the Lagrange multiplier $\boldsymbol{x}$. Interchanging the order of the maximization and minimization operators in (B.5), we obtain the standard Lagrangian dual of the dual problem (D).

$$\inf_{\boldsymbol{x}} \sup_{\boldsymbol{\lambda} \geq \mathbf{0}} \sup_{\{\boldsymbol{w}_i\}_{i \in \mathcal{I}_0}} -f_0^*(\boldsymbol{w}_0) - \sum_{i \in \mathcal{I}} \lambda_i f_i^*(\boldsymbol{w}_i/\lambda_i) + \sum_{i \in \mathcal{I}_0} \boldsymbol{w}_i^\top \boldsymbol{x} \tag{B.6}$$

For any fixed $\boldsymbol{x}$, the embedded maximization problems in (B.6) evaluate to $f(\boldsymbol{x})$ as defined in (B.1).

Indeed, a direct calculation reveals that

$$
\begin{aligned}
&\sup_{\boldsymbol{\lambda}\geq \mathbf{0}} \sup_{\{\boldsymbol{w}_i\}_{i\in\mathcal{I}_0}} \left\{ -f_0^*\left(\boldsymbol{w}_0\right) - \sum_{i\in\mathcal{I}} \lambda_i f_i^*\left(\boldsymbol{w}_i/\lambda_i\right) + \sum_{i\in\mathcal{I}_0} \boldsymbol{w}_i^\top \boldsymbol{x} \right\} \\
&= \sup_{\boldsymbol{\lambda}\geq \mathbf{0}} \left\{ \sup_{\boldsymbol{w}_0} \left\{ \boldsymbol{w}_0^\top \boldsymbol{x} - f_0^*(\boldsymbol{w}_0) \right\} + \sum_{\substack{i\in\mathcal{I}:\\ \lambda_i>0}} \sup_{\boldsymbol{w}_i} \left\{ \boldsymbol{w}_i^\top \boldsymbol{x} - \lambda_i f_i^*\left(\boldsymbol{w}_i/\lambda_i\right) \right\} \right. \\
&\qquad\qquad\qquad\qquad\qquad\qquad \left. + \sum_{\substack{i\in\mathcal{I}:\\ \lambda_i=0}} \sup_{\boldsymbol{w}_i} \left\{ \boldsymbol{w}_i^\top \boldsymbol{x} - \delta^*_{\mathrm{dom}(f_i)}(\{\boldsymbol{w}_i\}_i) \right\} \right\} \\
&= \sup_{\boldsymbol{\lambda}\geq \mathbf{0}} \left\{ f_0(\boldsymbol{x}) + \sum_{\substack{i\in\mathcal{I}:\\ \lambda_i>0}} \lambda_i f_i(\boldsymbol{x}) + \sum_{\substack{i\in\mathcal{I}:\\ \lambda_i=0}} \delta_{\mathrm{cl}(\mathrm{dom}(f_i))}(\boldsymbol{x}) \right\} = f(\boldsymbol{x}).
\end{aligned}
$$

Here, the first equality follows from a regrouping of terms, the definition of the convex perspective and Theorem 12.2 by Rockafellar (1970), which applies because f_i, $i \in \mathcal{I}_0$, is proper, closed and convex by assumption **(F)**. The second equality exploits the fact that each inner maximization evaluates a conjugate. The explicit expressions for these conjugates follow from Theorems 12.2 and 13.2 and the remarks before Theorem 16.1 in the monograph by Rockafellar (1970), where we again use the fact that f_i, $i \in \mathcal{I}_0$, is proper, closed and convex. The third equality, finally, follows from a case distinction. Thus, the Lagrangian dual of (D) is equivalent to (P). □

Proof of Theorem 2. In view of assertion *(i)*, assume first that (P) admits a Slater point. In this case (P) is feasible, and its infimum is strictly smaller than $+\infty$. If the infimum of (P) evaluates to $-\infty$, then the supremum of (D) also amounts to $-\infty$ by weak duality (see Theorem 1), and (D) is solvable because any solution of an infeasible problem is optimal according to our convention. In the remainder we may thus assume that the infimum of (P) is finite. In this case, we will first show that both inequalities in (B.3) collapse to equalities, which implies that the duality gap between (P) and (D) vanishes. Indeed, the first inequality in (B.3) becomes tight due to Proposition 5.3.6 by Bertsekas (2009), which applies because (P) has a finite infimum and admits a Slater point

$$
\boldsymbol{x}^{\mathrm{S}} \in \bigcap_{i\in\mathcal{I}_0} \mathrm{ri}(\mathrm{dom}(f_i)) = \mathrm{ri}\Big(\bigcap_{i\in\mathcal{I}_0} \mathrm{dom}(f_i) \Big),
$$

where the above equality holds due to Proposition 2.42 by Rockafellar and Wets (2009). The second inequality in (B.3) becomes tight due to Proposition C.4 and because the existence of a Slater point guarantees that $\cap_{i\in\mathcal{I}_0}\mathrm{ri}(\mathrm{dom}(f_i))\neq\emptyset$. To establish the solvability of (D), note that (B.2) is solved by some $\boldsymbol{\lambda}^\star$ due to Proposition 5.3.6 by Bertsekas (2009) and that the parametric problem (B.4) for $\boldsymbol{\lambda}=\boldsymbol{\lambda}^*$ is solved by some $\{\boldsymbol{w}_i^\star\}_{i\in\mathcal{I}_0}$ due to Proposition C.4 and because $\cap_{i\in\mathcal{I}_0}\mathrm{ri}(\mathrm{dom}(f_i))\neq\emptyset$. By construction, $(\boldsymbol{\lambda}^\star,\{\boldsymbol{w}_i^\star\}_{i\in\mathcal{I}_0})$ thus constitutes an optimal solution for (D).

Assume now that (D) admits a Slater point. Similar arguments as in the previous paragraph show that strong duality and solvability of (P) trivially hold if the supremum of (D) evaluates to $+\infty$, and we may thus assume that the supremum of (D) is finite. Strong duality between (P) and (D) as well as the solvability of (P) then follow from Lemma B.1 and Proposition 5.3.6 by Bertsekas (2009), which applies because (D) has a finite supremum and admits a Slater point that satisfies all explicit (linear) constraints and resides in the relative interior of the objective function.

As for assertion *(ii)*, assume first that the feasible region of (P) is nonempty and bounded. This ensures via assumption **(F)** that the function f in (B.1) is proper and has compact sublevel sets. Strong duality between (P) and (D) as well as solvability of (P) then follow from Lemma B.1 as well as Proposition 5.5.4 by Bertsekas (2009).

Finally, assume that the feasible region of (D) is nonempty and bounded, which ensures via our definition of the convex perspective and assumption **(F)** that the (negative) optimal value function of the inner minimization problem of (B.5) in the proof of Lemma B.1, which is given by

$$-\inf_{\boldsymbol{x}} -f_0^*(\boldsymbol{w}_0)-\sum_{i\in\mathcal{I}}\lambda_i f_i^*(\boldsymbol{w}_i/\lambda_i)+\sum_{i\in\mathcal{I}_0}\boldsymbol{w}_i^\top\boldsymbol{x}=\begin{cases} f_0^*(\boldsymbol{w}_0)+\sum_{i\in\mathcal{I}}\lambda_i f_i^*(\boldsymbol{w}_i/\lambda_i) & \text{if } \sum_{i\in\mathcal{I}_0}\boldsymbol{w}_i=\boldsymbol{0},\\ +\infty & \text{otherwise,}\end{cases}$$

is proper and has compact sublevel sets. Strong duality and solvability of (D) thus follow from Lemma B.1 and Proposition 5.5.4 by Bertsekas (2009). □

Proof of Proposition C.6 We construct a strict Slater point for (D) from *(i)* a (possibly infeasible) solution $(\{\boldsymbol{w}_i^+\}_i,\boldsymbol{\lambda}^+)$ to (D) that resides in the relative interior of the domain of the objective function of (D) and *(ii)* a point $(\{\boldsymbol{w}_i^-\}_i,\boldsymbol{\lambda}^-)$ that resides in (but not necessarily in the relative interior of) the domain of the objective function of (D) and that offsets any infeasibility of $(\{\boldsymbol{w}_i^+\}_i,\boldsymbol{\lambda}^+)$.

By assumption **(F)**, the function f_i is proper, and by Theorem 12.2 of Rockafellar (1970), its conjugate f_i^* inherits properness from f_i for each $i\in\mathcal{I}_0$. Thus, there exists $\boldsymbol{w}_i^+\in\mathrm{ri}(\mathrm{dom}(f_i^*))$ for every $i\in\mathcal{I}_0$. Setting $\boldsymbol{\lambda}^+=\boldsymbol{1}>\boldsymbol{0}$, it is then easy to verify that $(\{\boldsymbol{w}_i^+\}_i,\boldsymbol{\lambda}^+)$ resides within the relative interior of the domain of the objective function of (D). However, because $\boldsymbol{w} = \sum_{i\in\mathcal{I}} \boldsymbol{w}_i^+$ may differ from $\mathbf{0}$, the solution $(\{\boldsymbol{w}_i^+\}_i, \boldsymbol{\lambda}^+)$ may nevertheless be infeasible in (D).

To construct the point $(\{\boldsymbol{w}_i^-\}_i, \boldsymbol{\lambda}^-)$, we consider the following variant of (P), where we add the linear term $\boldsymbol{w}^\top \boldsymbol{x}$ to the objective function with the fixed gradient $\boldsymbol{w} \in \mathbb{R}^{d_{\boldsymbol{x}}}$.

$$\begin{array}{ll} \inf & f_0(\boldsymbol{x}) + \boldsymbol{w}^\top \boldsymbol{x} \\ \text{s.t.} & f_i(\boldsymbol{x}) \leq 0 \qquad \forall i \in \mathcal{I} \\ & \boldsymbol{x} \text{ free} \end{array} \tag{P$_{\boldsymbol{w}}$}$$

As the conjugate of the new objective function $f_0(\boldsymbol{x})+\boldsymbol{w}^\top \boldsymbol{x}$ evaluated at $\boldsymbol{w}_0$ amounts to $f_0^*(\boldsymbol{w}_0-\boldsymbol{w})$, the variable substitution $\boldsymbol{w}_0 \leftarrow \boldsymbol{w}_0 - \boldsymbol{w}$ allows us to express the problem dual to (P$_{\boldsymbol{w}}$) as

$$\begin{array}{ll} \sup & -f_0^*\left(\boldsymbol{w}_0\right) - \displaystyle\sum_{i\in\mathcal{I}} \lambda_i f_i^*\left(\boldsymbol{w}_i/\lambda_i\right) \\ \text{s.t.} & \displaystyle\sum_{i\in\mathcal{I}_0} \boldsymbol{w}_i = -\boldsymbol{w} \\ & \boldsymbol{w}_i \text{ free} \qquad\qquad \forall i \in \mathcal{I}_0 \\ & \boldsymbol{\lambda} \geq \mathbf{0}. \end{array} \tag{D$_{\boldsymbol{w}}$}$$

By construction, (P$_{\boldsymbol{w}}$) and (P) share the same feasible region, which is nonempty and bounded by assumption, whereas (D$_{\boldsymbol{w}}$) and (D) share the same objective function. Similar arguments as in the proof of Theorem 2 *(ii)* thus imply that (P$_{\boldsymbol{w}}$) and (D$_{\boldsymbol{w}}$) share the same (finite) optimal value, which in turn ensures that problem (D$_{\boldsymbol{w}}$) admits a feasible solution $(\{\boldsymbol{w}_i^-\}_i, \boldsymbol{\lambda}^-)$. By construction, this solution resides within the domain of the common objective function of (D$_{\boldsymbol{w}}$) and (D) but not necessarily within its relative interior.

Next, define $(\boldsymbol{w}_i^{\mathrm{S}}, \lambda_i^{\mathrm{S}}) = \frac{1}{2}(\boldsymbol{w}_i^+, \lambda_i^+) + \frac{1}{2}(\boldsymbol{w}_i^-, \lambda_i^-)$ for every $i \in \mathcal{I}_0$. By the line segment principle of Bertsekas (2009, Proposition 1.3.1), the constructed solution $(\{\boldsymbol{w}_i^{\mathrm{S}}\}_i, \boldsymbol{\lambda}^{\mathrm{S}})$ belongs to the relative interior of the domain of the objective function of (D). In addition, we have

$$\sum_{i\in\mathcal{I}} \boldsymbol{w}_i^{\mathrm{S}} \;=\; \frac{1}{2}\sum_{i\in\mathcal{I}} \boldsymbol{w}_i^+ + \frac{1}{2}\sum_{i\in\mathcal{I}} \boldsymbol{w}_i^- \;=\; \frac{1}{2}\boldsymbol{w} - \frac{1}{2}\boldsymbol{w} \;=\; \mathbf{0}$$

and $\boldsymbol{\lambda}^{\mathrm{S}} > \mathbf{0}$. Therefore, the solution $(\{\boldsymbol{w}_i^{\mathrm{S}}\}_i, \boldsymbol{\lambda}^{\mathrm{S}})$ constitutes a strict Slater point for (D). □

Proof of Theorem 3. For any fixed $\boldsymbol{z}_i \in \mathcal{Z}$, $i \in \mathcal{I}_0$, the problems (P-W) and (D-B) collapse to instances of (P-S) and (D-S), respectively, and the following inequalities are due to Theorem 1.

$$
\begin{aligned}
& \inf_{\boldsymbol{x}\in\mathcal{X}(\boldsymbol{z}_1,\ldots,\boldsymbol{z}_I)} f_0(\boldsymbol{x},\boldsymbol{z}_0) \geq \sup_{\substack{\sum_{i\in\mathcal{I}_0}\boldsymbol{w}_i=\boldsymbol{0}\\ \boldsymbol{\lambda}\geq\boldsymbol{0}}} -f_0^{*1}\left(\boldsymbol{w}_0,\boldsymbol{z}_0\right) - \sum_{i\in\mathcal{I}} \lambda_i f_i^{*1}(\boldsymbol{w}_i/\lambda_i,\boldsymbol{z}_i) \\
\Longrightarrow \quad & \sup_{\{\boldsymbol{z}_i\}_{i\in\mathcal{I}_0}\subseteq\mathcal{Z}} \inf_{\boldsymbol{x}\in\mathcal{X}(\boldsymbol{z}_1,\ldots,\boldsymbol{z}_I)} f_0(\boldsymbol{x},\boldsymbol{z}_0) \geq \sup_{\{\boldsymbol{z}_i\}_{i\in\mathcal{I}_0}\subseteq\mathcal{Z}} \sup_{\substack{\sum_{i\in\mathcal{I}_0}\boldsymbol{w}_i=\boldsymbol{0}\\ \boldsymbol{\lambda}\geq\boldsymbol{0}}} -f_0^{*1}\left(\boldsymbol{w}_0,\boldsymbol{z}_0\right) - \sum_{i\in\mathcal{I}} \lambda_i f_i^{*1}(\boldsymbol{w}_i/\lambda_i,\boldsymbol{z}_i),
\end{aligned}
$$

where $\mathcal{X}(\boldsymbol{z}_1,\ldots,\boldsymbol{z}_I) = \left\{\boldsymbol{x}\in\mathbb{R}^{d_{\boldsymbol{x}}} \mid f_i(\boldsymbol{x},\boldsymbol{z}_i)\leq 0 \;\; \forall i\in\mathcal{I}\right\}$. Note that the right-hand side of the second inequality is equivalent to (D-B), while left-hand side is upper bounded by (P-W) because

$$
\inf_{\boldsymbol{x}\in\mathcal{X}(\boldsymbol{z}_1,\ldots,\boldsymbol{z}_I)} \sup_{\{\boldsymbol{z}_i\}_{i\in\mathcal{I}_0}\subseteq\mathcal{Z}} f_0(\boldsymbol{x},\boldsymbol{z}_0) \quad \geq \quad \sup_{\{\boldsymbol{z}_i\}_{i\in\mathcal{I}_0}\subseteq\mathcal{Z}} \inf_{\boldsymbol{x}\in\mathcal{X}(\boldsymbol{z}_1,\ldots,\boldsymbol{z}_I)} f_0(\boldsymbol{x},\boldsymbol{z}_0)
$$

due to the min-max inequality. $\square$

Proof of Proposition 4. To show that (P-W′) upper bounds (P-W), we dualize the embedded maximization problems in (P-W) that evaluate the worst-case uncertainty realizations in the objective and the constraint functions. Specifically, for any fixed $i\in\mathcal{I}_0$ and $\boldsymbol{x}\in\mathbb{R}^{d_{\boldsymbol{x}}}$, we have

$$
\sup_{\boldsymbol{z}_i\in\mathcal{Z}} f_i(\boldsymbol{x},\boldsymbol{z}_i) = -\begin{cases} \inf & -f_i(\boldsymbol{x},\boldsymbol{z}_i) \\ \text{s.t.} & c_\ell(\boldsymbol{z}_i)\leq 0 \qquad \forall \ell\in\mathcal{L} \\ & \boldsymbol{z}_i \text{ free} \end{cases} \tag{B.7a}
$$

$$
\leq -\begin{cases} \sup & -(-f_i)^{*2}\left(\boldsymbol{x},\boldsymbol{y}_{i0}\right) - \sum_{\ell\in\mathcal{L}} \nu_{i\ell} c_\ell^*\left(\boldsymbol{y}_{i\ell}/\nu_{i\ell}\right) \\ \text{s.t.} & \sum_{\ell\in\mathcal{L}_0} \boldsymbol{y}_{i\ell} = \boldsymbol{0} \\ & \boldsymbol{y}_{i\ell} \text{ free} \qquad \forall \ell\in\mathcal{L}_0 \\ & \nu_{i\ell}\geq 0 \qquad \forall \ell\in\mathcal{L}, \end{cases} \tag{B.7b}
$$

where the inequality follows from Theorem 1, which applies because the assumptions **(RF)** and **(C)** imply that (B.7a) satisfies assumption **(F)** from Section 2. Interchanging the minus sign and the supremum operator in (B.7b) results in a minimization problem. Substituting the resulting minimization problem into (P-W) for every $i\in\mathcal{I}_0$ and then merging the infimum operators in the objective and removing the infimum operators in the constraints yields (P-W′), and thus the infimum of (P-W) is indeed smaller or equal to that of (P-W′). Note that if the optimal solution of (B.7b) is not attained for some $i \in \mathcal{I}$, then removing the infimum operators in the constraints may lead to a further restriction of the problem and therefore result in a higher optimal value.

As for assertion *(ii)*, assume that $\mathcal{Z}$ admits a Slater point $\boldsymbol{z}^{\mathrm{S}}$. As $\mathrm{dom}(-f_i(\boldsymbol{x}, \cdot)) = \mathbb{R}^{d_{\boldsymbol{x}}}$ due to assumption **(RF)**, $\boldsymbol{z}^{\mathrm{S}}$ is also a Slater point for the minimization problem in (B.7a). By Theorem 2 *(i)*, the duality gap between (B.7a) and (B.7b) thus vanishes, and (B.7b) is solvable. This implies that the infima of (P-W) and (P-W′) coincide and that any optimizer of (P-W) can be combined with optimizers of the dual subproblems (B.7b) for $i \in \mathcal{I}_0$ to construct an optimizer for (P-W′).

As for assertion *(iii)*, assume finally that $\mathcal{Z}$ is compact and that problem (P-W) admits a strict Slater point $\boldsymbol{x}^{\mathrm{S}}$. In this case, the functions $F_i(\boldsymbol{x}) = \sup_{\boldsymbol{z}_i \in \mathcal{Z}} f_i(\boldsymbol{x}, \boldsymbol{z}_i)$, $i \in \mathcal{I}_0$, are convex and continuous in $\boldsymbol{x}$ by virtue of assumption **(RF)**. Indeed, $F_i(\boldsymbol{x})$ is convex and closed because $f_i(\boldsymbol{x}, \boldsymbol{z}_i)$ is convex and closed in $\boldsymbol{x}$ for every fixed $\boldsymbol{z}_i$. Moreover, $F_i(\boldsymbol{x})$ is finite for every fixed $\boldsymbol{x}$ due to Weierstrass' extreme value theorem, which applies because $\mathcal{Z}$ is compact and $-f_i(\boldsymbol{x}, \boldsymbol{z}_i)$ is closed (and thus lower semicontinuous) in $\boldsymbol{z}_i$. As any convex function is continuous on the relative interior of its domain, we may thus conclude that each F_i is continuous on $\mathbb{R}^{d_{\boldsymbol{x}}}$. By forming convex combinations with the strict Slater point $\boldsymbol{x}^{\mathrm{S}}$, one can now use the continuity and convexity of the functions F_i, $i \in \mathcal{I}_0$, to prove that any $\boldsymbol{x}$ feasible in (P-W) can be represented as a limit of strict Slater points for (P-W). Therefore, (P-W) is equivalent to

$$
\begin{array}{ll}
\inf & \sup\limits_{\boldsymbol{z}_0 \in \mathcal{Z}} f_0(\boldsymbol{x}, \boldsymbol{z}_0) \\
\text{s.t.} & \sup\limits_{\boldsymbol{z}_i \in \mathcal{Z}} f_i(\boldsymbol{x}, \boldsymbol{z}_i) < 0 \qquad \forall i \in \mathcal{I} \\
& \boldsymbol{x} \text{ free.}
\end{array} \tag{B.8}
$$

We can now dualize the embedded maximization problems in (B.8) as in the proof of assertion *(i)*. The compactness of $\mathcal{Z}$ implies via Theorem 2 *(ii)* that the duality gap between (B.7a) and (B.7b) vanishes. The minimization problems resulting from interchanging the minus sign and the supremum operator in (B.7b) can then be substituted back into (B.8), the infimum operators in the objective can be merged, and the infimum operators in the constraints can be removed to obtain a variant of (P-W′) with strict inequalities. Note that because the constraints in (B.8) are strict, the infimum operators in the constraints may indeed be removed without restricting the problem even if the corresponding subproblems are not solvable. Next, we argue that the strict inequalities in the resulting problem can again be relaxed to weak inequalities without changing the problem's optimal value. By Remark C.1, this is the case if problem (P-W$'$) admits a strict Slater point. Such a strict Slater point can be constructed by combining the strict Slater point $\boldsymbol{x}^{\mathrm{S}}$ of (P-W) with strict Slater points $(\{\boldsymbol{y}_{i\ell}^{\mathrm{S}}, \nu_{i\ell}^{\mathrm{S}}\}_{\ell})$ for the dual subproblems, $i \in I_0$, which exist thanks to Proposition C.6. Thus, the infima of (P-W) and (P-W$'$) are indeed equal.

Finally, to see that the solvability of (P-W$'$) implies the solvability of (P-W), assume that $(\boldsymbol{x}^{\star}, \{\boldsymbol{y}_{i\ell}^{\star}, \nu_{i\ell}^{\star}\}_{i,\ell})$ solves (P-W$'$). The above reasoning then implies that the optimal value of (P-W$'$) amounts to $F_0(\boldsymbol{x}^{\star})$, which in turn shows that $\boldsymbol{x}^{\star}$ solves (P-W). $\square$

Proof of Proposition 5. As for *(i)*, we prove that any feasible solution to (D-B) corresponds to a feasible solution to (D-B$'$) with the same objective value. To this end, select any $(\{\boldsymbol{w}_i, \boldsymbol{z}_i\}_i, \boldsymbol{\lambda})$ feasible in (D-B) and define $\boldsymbol{v}_i = \lambda_i \boldsymbol{z}_i$ for $i \in \mathcal{I}$. We show that $(\{\boldsymbol{w}_i\}_i, \boldsymbol{z}_0, \boldsymbol{\lambda}, \{\boldsymbol{v}_i\}_i)$ is feasible in (D-B$'$) and attains the same objective value. Indeed, it is clear that $\lambda_i c_\ell(\boldsymbol{v}_i/\lambda_i) \leq 0$ for all $\ell \in \mathcal{L}$ and $i \in \mathcal{I}$ with $\lambda_i > 0$. If $\lambda_i = 0$ for some $i \in \mathcal{I}$, on the other hand, we have $\boldsymbol{v}_i = \boldsymbol{0}$ and

$$0 c_\ell(\boldsymbol{0}/0) \;=\; \delta^*_{\mathrm{dom}(c_\ell^*)}(\boldsymbol{0}) \;=\; 0 \qquad \forall \ell \in \mathcal{L},$$

where the first equality follows from the definition of the convex perspective, while the second equality holds because c_ℓ^* inherits properness from c_ℓ (Rockafellar, 1970, Theorem 12.2) and because the support function of $\mathrm{dom}(c_\ell^*) \neq \emptyset$ vanishes at the origin. All other constraints of (D-B$'$) are trivially satisfied. Next, we show that the objective value of $(\{\boldsymbol{w}_i, \boldsymbol{z}_i\}_i, \boldsymbol{\lambda})$ in (D-B) equals that of $(\{\boldsymbol{w}_i\}_i, \boldsymbol{z}_0, \boldsymbol{\lambda}, \{\boldsymbol{v}_i\}_i)$ in (D-B$'$). Indeed, it is clear that $\lambda_i f_i^{*1}(\boldsymbol{w}_i/\lambda_i, \boldsymbol{v}_i/\lambda_i) = \lambda_i f_i^{*1}(\boldsymbol{w}_i/\lambda_i, \boldsymbol{z}_i)$ for all $i \in \mathcal{I}$ with $\lambda_i > 0$. If $\lambda_i = 0$ for some $i \in \mathcal{I}$, on the other hand, then $\boldsymbol{v}_i = \boldsymbol{0}$, and the convex perspective function $0 f_i^{*1}(\boldsymbol{w}_i/0, \boldsymbol{0}/0)$ is defined as the support function of the domain of $(f_i^{*1})^*$. As $(f_i^{*1})^* = (-f_i)^{*2}$ due to Proposition C.7, we may thus conclude that

$$\begin{aligned}
0 f_i^{*1}(\boldsymbol{w}_i/0, \boldsymbol{0}/0) \;&=\; \delta^*_{\mathrm{dom}((-f_i)^{*2})}(\boldsymbol{w}_i, \boldsymbol{0}) \\
&=\; \sup_{\boldsymbol{x}} \left\{ \boldsymbol{w}_i^\top \boldsymbol{x} \mid \exists \boldsymbol{y} \in \mathbb{R}^{d_{\boldsymbol{z}}} : (-f_i)^{*2}(\boldsymbol{x}, \boldsymbol{y}) < +\infty \right\} \\
&=\; \delta_{\{\boldsymbol{0}\}}(\boldsymbol{w}_i) \\
&=\; \delta^*_{\mathrm{dom}(f_i(\cdot, \boldsymbol{z}_i))}(\boldsymbol{w}_i) \;=\; 0 f_i^{*1}(\boldsymbol{w}_i/0, \boldsymbol{z}_i),
\end{aligned}$$

where the third equality follows from Rockafellar (1970, Theorem 12.2), which ensures that for any $\boldsymbol{x} \in \mathbb{R}^{d_{\boldsymbol{x}}}$ the partial conjugate $(-f_i)^{*2}(\boldsymbol{x}, \cdot)$ of the proper function $-f_i(\boldsymbol{x}, \cdot)$ is also proper.

Thus, for any $\boldsymbol{x} \in \mathbb{R}^{d_{\boldsymbol{x}}}$, there exists $\boldsymbol{y} \in \mathbb{R}^{d_{\boldsymbol{z}}}$ with $(-f_i)^{*2}(\boldsymbol{x}, \boldsymbol{y}) < +\infty$, which implies that $\boldsymbol{x}$ is actually free, and the supremum evaluates to $+\infty$ unless $\boldsymbol{w}_i = \boldsymbol{0}$. The fourth equality holds because $\mathrm{dom}(f_i(\cdot, \boldsymbol{z}_i)) = \mathbb{R}^{d_{\boldsymbol{x}}}$ for every $\boldsymbol{z}_i \in \mathbb{R}^{d_{\boldsymbol{z}}}$, and the last equality follows from the definition of the partial convex perspective and from Theorem 12.2 of Rockafellar (1970), which implies that $(f_i^{*1})^{*1}(\cdot, \boldsymbol{z}_i) = f_i(\cdot, \boldsymbol{z}_i)$. In summary, we have shown that the optimal value of (D-B) does not exceed that of (D-B$'$), and thus assertion *(i)* follows.

Assume now that $(\{\boldsymbol{w}_i^{\mathrm{S}}, \boldsymbol{z}_i^{\mathrm{S}}\}_i, \boldsymbol{\lambda}^{\mathrm{S}})$ is a strict Slater point for problem (D-B), which implies that $\boldsymbol{\lambda}^{\mathrm{S}} > \boldsymbol{0}$. In that case, $(\{\boldsymbol{w}_i^{\mathrm{S}}\}_i, \boldsymbol{z}_0^{\mathrm{S}}, \boldsymbol{\lambda}^{\mathrm{S}}, \{\boldsymbol{\upsilon}_i^{\mathrm{S}}\}_i)$ with $\boldsymbol{\upsilon}_i^{\mathrm{S}} = \lambda_i^{\mathrm{S}} \cdot \boldsymbol{z}_i^{\mathrm{S}}$, $i \in \mathcal{I}$, is a strict Slater point for problem (D-B$'$). To prove assertion *(ii)*, we show that any feasible solution to (D-B$'$) corresponds to a sequence of feasible solutions to (D-B) that asymptotically attain a non-inferior objective value. This implies that the optimal value of (D-B$'$) is smaller or equal to that of (D-B), and together with assertion *(i)* we can then conclude that the optimal values of (D-B) and (D-B$'$) coincide. To this end, select any solution $(\{\boldsymbol{w}_i\}_i, \boldsymbol{z}_0, \boldsymbol{\lambda}, \{\boldsymbol{\upsilon}_i\}_i)$ feasible in (D-B$'$) and any $\epsilon > 0$. As the feasible region of (D-B$'$) is convex and the objective function of (D-B$'$) is concave, there exists $\theta \in (0, 1)$ such that the solution $(\{\boldsymbol{w}_i^\epsilon\}_i, \boldsymbol{z}_0^\epsilon, \boldsymbol{\lambda}^\epsilon, \{\boldsymbol{\upsilon}_i^\epsilon\}_i)$ defined through

$$(\{\boldsymbol{w}_i^\epsilon\}_i, \boldsymbol{z}_0^\epsilon, \boldsymbol{\lambda}^\epsilon, \{\boldsymbol{\upsilon}_i^\epsilon\}_i) = \theta \cdot (\{\boldsymbol{w}_i\}_i, \boldsymbol{z}_0, \boldsymbol{\lambda}, \{\boldsymbol{\upsilon}_i\}_i) \; + \; (1-\theta) \cdot (\{\boldsymbol{w}_i^{\mathrm{S}}\}_i, \boldsymbol{z}_0^{\mathrm{S}}, \boldsymbol{\lambda}^{\mathrm{S}}, \{\boldsymbol{\upsilon}_i^{\mathrm{S}}\}_i) \tag{B.9}$$

is feasible in (D-B$'$) and attains an objective function value that is at least as large as that of $(\{\boldsymbol{w}_i\}_i, \boldsymbol{z}_0, \boldsymbol{\lambda}, \{\boldsymbol{\upsilon}_i\}_i)$ minus ϵ. Setting $\boldsymbol{z}_i^\epsilon = \boldsymbol{\upsilon}_i^\epsilon / \lambda_i^\epsilon$ for all $i \in \mathcal{I}$, which is possible because $\boldsymbol{\lambda}^\epsilon > \boldsymbol{0}$, it is clear that $(\{\boldsymbol{w}_i^\epsilon, \boldsymbol{z}_i^\epsilon\}_i, \boldsymbol{\lambda}^\epsilon)$ is feasible in (D-B) and attains the same objective value as $(\{\boldsymbol{w}_i^\epsilon\}_i, \boldsymbol{z}_0^\epsilon, \boldsymbol{\lambda}^\epsilon, \{\boldsymbol{\upsilon}_i^\epsilon\}_i)$ in (D-B$'$). As $(\{\boldsymbol{w}_i\}_i, \boldsymbol{z}_0, \boldsymbol{\lambda}, \{\boldsymbol{\upsilon}_i\}_i)$ and $\epsilon > 0$ were chosen arbitrarily, the supremum of (D-B) is thus at least as large as that of (D-B$'$). Together with assertion *(i)*, we thus conclude that the suprema of (D-B) and (D-B$'$) coincide. Moreover, since our proof of assertion *(i)* has shown that any feasible solution to (D-B) corresponds to a feasible solution to (D-B$'$) with the same objective value, (D-B$'$) is solvable whenever (D-B) is solvable.

Assume now that $\mathcal{Z}$ is bounded. To prove assertion *(iii)*, we show that any feasible solution to (D-B$'$) corresponds to a feasible solution to (D-B) with the same objective value. Together with assertion *(i)*, this implies that the optimal values of (D-B) and (D-B$'$) coincide. To this end, select any solution $(\{\boldsymbol{w}_i\}_i, \boldsymbol{z}_0, \boldsymbol{\lambda}, \{\boldsymbol{\upsilon}_i\}_i)$ feasible in (D-B$'$), and define $\boldsymbol{z}_i = \boldsymbol{\upsilon}_i / \lambda_i$ if $\lambda_i > 0$ and $\boldsymbol{z}_i = \boldsymbol{z}_0$ if $\lambda_i = 0$, $i \in \mathcal{I}$. Lemma C.8 *(i)* implies that if $\lambda_i = 0$, then $\boldsymbol{\upsilon}_i$ must be a recession direction for the uncertainty set $\mathcal{Z}$. As $\mathcal{Z}$ is nonempty and bounded, this in turn implies that $\boldsymbol{\upsilon}_i = \boldsymbol{0}$. Using the same reasoning as in the proof of assertion *(i)*, one can thus show that $\lambda_i c_\ell(\boldsymbol{v}_i/\lambda_i) = \lambda_i c_\ell(\boldsymbol{z}_i)$ for all $\ell \in \mathcal{L}$ and $i \in \mathcal{I}$ and $\lambda_i f_i^{*1}(\boldsymbol{w}_i/\lambda_i, \boldsymbol{v}_i/\lambda_i) = \lambda_i f_i^{*1}(\boldsymbol{w}_i/\lambda_i, \boldsymbol{z}_i)$ for all $i \in \mathcal{I}$. This implies that $(\{\boldsymbol{w}_i, \boldsymbol{z}_i\}_i, \boldsymbol{\lambda})$ is feasible in (D-B) and attains the same objective value as $(\{\boldsymbol{w}_i\}_i, \boldsymbol{z}_0, \boldsymbol{\lambda}, \{\boldsymbol{v}_i\}_i)$ in (D-B$'$). Note that our proof of assertion *(i)* has shown that any feasible solution to (D-B) corresponds to a feasible solution to (D-B$'$) with the same objective value, and our proof of assertion *(iii)* has shown that any feasible solution to (D-B$'$) corresponds to a feasible solution to (D-B) with the same objective value. Since the optimal values of both problems coincide, we can conclude that (D-B) is solvable if and only if (D-B$'$) is solvable. $\square$

Proof of Theorem 6. We show that (P-W$'$) and (D-B$'$) can be viewed as instances of (P) and (D), respectively. Assertions *(i)*, *(ii)* and *(iii)* can then be derived from Theorems 1 and 2. For ease of exposition, we first rewrite the convex optimization problem (P-W$'$) more concisely as

$$
\begin{array}{lll}
\inf & \varphi_0(\boldsymbol{x}, \{\boldsymbol{y}_{0\ell}, \nu_{0\ell}\}_\ell) & \\
\text{s.t.} & \varphi_i(\boldsymbol{x}, \{\boldsymbol{y}_{i\ell}, \nu_{i\ell}\}_\ell) \leq 0 & \forall i \in \mathcal{I} \\
& \psi_{ik}^+(\{y_{i\ell k}\}_\ell) \leq 0 & \forall i \in \mathcal{I}_0,\ \forall k \in \mathcal{K} \\
& \psi_{ik}^-(\{y_{i\ell k}\}_\ell) \leq 0 & \forall i \in \mathcal{I}_0,\ \forall k \in \mathcal{K} \\
& \boldsymbol{x},\ \boldsymbol{y}_{i\ell} \text{ free} & \forall i \in \mathcal{I}_0,\ \forall \ell \in \mathcal{L}_0 \\
& \nu_{i\ell} \text{ free} & \forall i \in \mathcal{I}_0,\ \forall \ell \in \mathcal{L},
\end{array}
\tag{B.10}
$$

where the extended real-valued functions φ_i for $i \in \mathcal{I}_0$ are defined through

$$
\varphi_i(\boldsymbol{x}, \{\boldsymbol{y}_{i\ell}, \nu_{i\ell}\}_\ell) = \begin{cases} (-f_i)^{*2}(\boldsymbol{x}, \boldsymbol{y}_{i0}) + \sum_{\ell \in \mathcal{L}} \nu_{i\ell} c_\ell^*(\boldsymbol{y}_{i\ell}/\nu_{i\ell}) & \text{if } \nu_{i\ell} \geq 0,\ \ell \in \mathcal{L}, \\ +\infty & \text{otherwise,} \end{cases}
$$

the linear equalities in (P-W$'$) are split into two sets of linear inequalities defined through

$$
\psi_{ik}^+(\{y_{i\ell k}\}_\ell) = \sum_{\ell \in \mathcal{L}_0} y_{i\ell k} \quad \text{and} \quad \psi_{ik}^-(\{y_{i\ell k}\}_\ell) = -\sum_{\ell \in \mathcal{L}_0} y_{i\ell k}, \qquad \forall i \in \mathcal{I}_0,\ \forall k \in \mathcal{K},
$$

and $y_{i\ell k}$ is the k-th element of the vector $\boldsymbol{y}_{i\ell}$ for every $k \in \mathcal{K} = \{1, \ldots, K\}$, where $K = d_{\boldsymbol{z}}$.

Note that (B.10) can be viewed as an instance of (P). Moreover, one can show that its objective and constraint functions satisfy assumption **(F)**, that is, one can show that φ_i, $\boldsymbol{\psi}_i^+$ and $\boldsymbol{\psi}_i^-$ are proper, closed and convex for every $i \in \mathcal{I}_0$. To see this, note first that the partial conjugate $(-f_i)^{*2}$ is proper, closed and convex by Proposition C.5 and by Theorem 12.2 of Rockafellar (1970), which apply because f_i obeys assumption **(RF)**. Similarly, the convex perspective $\nu_{i\ell} c_\ell^*(\boldsymbol{y}_{i\ell}/\nu_{i\ell})$ defined for $\nu_{i\ell} \geq 0$ is proper, closed and convex by Proposition C.2 and by Theorem 12.2 of Rockafellar (1970), which apply because c_ℓ obeys assumption **(C)**. Thus, φ_i constitutes a sum of proper, closed and convex functions with different arguments and is therefore also proper, closed and convex.[1] Finally, ψ_{ik}^+ and ψ_{ik}^- are linear functions and therefore proper, closed and convex.

If we interpret (B.10) as an instance of (P), denote the variables conjugate to $\boldsymbol{x}$ and $\nu_{i\ell}$ by $\boldsymbol{w}_i$ and $u_{i\ell}$, respectively, and denote the variables conjugate to $\boldsymbol{y}_{i\ell}$ by $\boldsymbol{r}_{i\ell}, \boldsymbol{r}_{i\ell}^+$ and $\boldsymbol{r}_{i\ell}^-$, then the corresponding instance of (D) can be represented as

$$
\begin{array}{lll}
\sup & \multicolumn{2}{l}{-\varphi_0^*(\boldsymbol{w}_0, \{\boldsymbol{r}_{0\ell}, u_{0\ell}\}_\ell) - \displaystyle\sum_{i\in\mathcal{I}} \left[\lambda_i \varphi_i^* \left(\frac{\boldsymbol{w}_i}{\lambda_i}, \frac{\{\boldsymbol{r}_{i\ell}\}_\ell}{\lambda_i}, \frac{\{u_{i\ell}\}_\ell}{\lambda_i} \right) + \right.} \\
& \multicolumn{2}{r}{\left. \displaystyle\sum_{k\in\mathcal{K}} v_{ik}^+ (\psi_{ik}^+)^* \left(\frac{\{r_{i\ell k}^+\}_\ell}{v_{ik}^+} \right) + \sum_{k\in\mathcal{K}} v_{ik}^- (\psi_{ik}^-)^* \left(\frac{\{r_{i\ell k}^-\}_\ell}{v_{ik}^-} \right) \right]} \\
\text{s.t.} & \displaystyle\sum_{i\in\mathcal{I}_0} \boldsymbol{w}_i = \boldsymbol{0} & \forall \ell \in \mathcal{L}_0 \\
& \boldsymbol{r}_{i\ell} + \boldsymbol{r}_{i\ell}^+ + \boldsymbol{r}_{i\ell}^- = \boldsymbol{0} & \forall i \in \mathcal{I}_0,\ \forall \ell \in \mathcal{L}_0 \\
& u_{i\ell} = 0 & \forall i \in \mathcal{I}_0,\ \forall \ell \in \mathcal{L} \\
& \boldsymbol{w}_i,\ r_{i\ell},\ \boldsymbol{r}_{i\ell}^+,\ \boldsymbol{r}_{i\ell}^- \text{ free} & \forall i \in \mathcal{I}_0,\ \forall \ell \in \mathcal{L}_0 \\
& \boldsymbol{\lambda},\ \boldsymbol{v}_i^+,\ \boldsymbol{v}_i^- \geq \boldsymbol{0} & \forall i \in \mathcal{I}_0,
\end{array}
\tag{B.11}
$$

where $r_{i\ell k}^+$ and $r_{i\ell k}^-$ are the k-th elements of the respective vectors $\boldsymbol{r}_{i\ell}^+$ and $\boldsymbol{r}_{i\ell}^-$ for every $k \in \mathcal{K}$, and $(\boldsymbol{\lambda}, \{\boldsymbol{v}_i^+, \boldsymbol{v}_i^-\}_i)$ are the dual variables associated with the three sets of inequalities in (B.10). Note that $u_{i\ell}$ is forced to 0 in (B.11) because its conjugate variable $\nu_{i\ell}$ only appears in the objective (if $i = 0$) or in the i-th constraint (if $i \in \mathcal{I}$) of the primal problem (B.10). The conjugate of φ_i,

[1] The fact that the summands do not share common arguments is crucial here. The sum of the two proper, closed and convex functions $\delta_{[0,1]}(x)$ and $\delta_{[2,3]}(x)$ in the common argument $x \in \mathbb{R}$, for example, is not proper.

$i \in \mathcal{I}_0$, can be calculated explicitly as

$$
\begin{aligned}
\varphi_i^*(\boldsymbol{w}_i, \{\boldsymbol{r}_{i\ell}, u_{i\ell}\}_\ell) &= \sup_{\boldsymbol{x}, \boldsymbol{y}_{i0}} \left\{ \boldsymbol{w}_i^\top \boldsymbol{x} + \boldsymbol{r}_{i0}^\top \boldsymbol{y}_{i0} - (-f_i)^{*2}(\boldsymbol{x}, \boldsymbol{y}_{i0}) \right\} \\
&\qquad + \sum_{\ell \in \mathcal{L}} \sup_{\substack{\boldsymbol{y}_{i\ell} \\ \nu_{i\ell} > 0}} \left\{ u_{i\ell}\nu_{i\ell} + \boldsymbol{r}_{i\ell}^\top \boldsymbol{y}_{i\ell} - \nu_{i\ell} c_\ell^* \left(\frac{\boldsymbol{y}_{i\ell}}{\nu_{i\ell}} \right) \right\} \\
&= f_i^{*1}(\boldsymbol{w}_i, \boldsymbol{r}_{i0}) + \sum_{\ell \in \mathcal{L}} \sup_{\nu_{i\ell} > 0} \left\{ u_{i\ell}\nu_{i\ell} + \nu_{i\ell} c_\ell \left(\boldsymbol{r}_{i\ell} \right) \right\} \\
&= \begin{cases} f_i^{*1}(\boldsymbol{w}_i, \boldsymbol{r}_{i0}) & \text{if } u_{i\ell} + c_\ell \left(\boldsymbol{r}_{i\ell} \right) \leq 0 \ \ \forall \ell \in \mathcal{L} \\ +\infty & \text{otherwise.} \end{cases}
\end{aligned}
$$

Note that we may restrict $\nu_{i\ell}$ to be strictly positive because the convex perspective of c_ℓ^* at $\nu_{i\ell} = 0$ is defined as the lower semicontinuous extension of the perspective for $\nu_{i\ell} > 0$; see (A.1). The second equality then follows from Proposition C.7, which applies because f_i satisfies assumption **(RF)**, and from Theorem 16.1 of Rockafellar (1970), which applies because c_ℓ satisfies assumption **(C)**.

Similarly, the conjugates of ψ_{ik}^+ and ψ_{ik}^- can be expressed as follows.

$$
(\psi_{ik}^+)^*(\{r_{i\ell k}^+\}_\ell) = \begin{cases} 0 & \text{if } r_{i\ell k}^+ = 1 \ \ \forall \ell \in \mathcal{L}_0 \\ +\infty & \text{otherwise} \end{cases}
$$

$$
(\psi_{ik}^-)^*(\{r_{i\ell k}^-\}_\ell) = \begin{cases} 0 & \text{if } r_{i\ell k}^- = -1 \ \ \forall \ell \in \mathcal{L}_0 \\ +\infty & \text{otherwise.} \end{cases}
$$

Substituting the formulas for φ_i^*, $(\psi_{ik}^+)^*$ and $(\psi_{ik}^-)^*$ into (B.11) with $\boldsymbol{z}_0 = \boldsymbol{v}_0^+ - \boldsymbol{v}_0^-$ and $\boldsymbol{v}_i = \boldsymbol{v}_i^+ - \boldsymbol{v}_i^-$, $i \in \mathcal{I}$, and eliminating the variables $\boldsymbol{r}_{i\ell}$, $\boldsymbol{r}_{i\ell}^+$, $\boldsymbol{r}_{i\ell}^-$ and $u_{i\ell}$, $i \in \mathcal{I}_0$ and $\ell \in \mathcal{L}_0$, finally yields (D-B′). □

Proof of Theorem 7. As for assertion *(i)*, assume that (P-W) admits a strict Slater point $\boldsymbol{x}^{\mathrm{S}}$ and that the uncertainty set $\mathcal{Z}$ is nonempty (by assumption) and compact (by assumption **(C)** and the assertion). Then the infima of (P-W) and (P-W′) coincide due to Proposition 4 *(iii)*. Moreover, since f_i is real-valued, the problem (B.7b) admits a strict Slater point $(\{\boldsymbol{y}_{i\ell}^{\mathrm{S}}, \nu_{i\ell}^{\mathrm{S}}\}_\ell)$ for fixed $\boldsymbol{x}^{\mathrm{S}}$ and for every $i \in \mathcal{I}_0$ due to Proposition C.6. We can combine these strict Slater points to a Slater point for problem (P-W′), and Theorem 6 *(ii)* implies that (P-W′) and (D-B′) satisfy strong duality, and (D-B′) is solvable. The claim then follows from Proposition 5 *(iii)*, which ensures that the suprema of (D-B) and (D-B′) coincide, and that (D-B) is solvable because (D-B′) is solvable.

As for assertion *(ii)*, assume that the feasible region of (P-W) is nonempty and bounded and that $\mathcal{Z}$ is bounded. Note that problem (P-W) can be represented more concisely as

$$\inf_{\boldsymbol{x}} \left\{ F_0(\boldsymbol{x}) \mid F_i(\boldsymbol{x}) \leq 0 \ \ \forall i \in \mathcal{I} \right\}, \tag{B.12}$$

where $F_i(\boldsymbol{x}) = \sup_{\boldsymbol{z}_i \in \mathcal{Z}} f_i(\boldsymbol{x}, \boldsymbol{z}_i)$ constitutes a pointwise maximum of convex functions and is therefore convex. Moreover, as $\mathcal{Z}$ is compact and $f_i(\boldsymbol{x}, \boldsymbol{z}_i)$ is continuous in $\boldsymbol{z}_i$ for every $\boldsymbol{x}$, $F_i(\boldsymbol{x})$ is indeed finite for every $\boldsymbol{x}$, *i.e.*, $\mathrm{dom}(F_i) = \mathbb{R}^{d_{\boldsymbol{x}}}$. As any finite-valued convex function is continuous, we may thus conclude that F_i is proper, closed and convex. The problem dual to (B.12) can be expressed as

$$\max_{\boldsymbol{\lambda} \geq \boldsymbol{0}, \boldsymbol{w}} \left\{ -F_0^* \left(\boldsymbol{w}_0 \right) - \sum_{i \in \mathcal{I}} \lambda_i F_i^*(\boldsymbol{w}_i / \lambda_i) \,\middle|\, \sum_{i \in \mathcal{I}_0} \boldsymbol{w}_i = \boldsymbol{0} \right\}. \tag{B.13}$$

As the feasible region of (P-W) is nonempty and bounded, Theorem 2 *(ii)* ensures that strong duality holds and (P-W) is solvable, while Proposition C.6 implies that (B.13) admits a strict Slater point $(\boldsymbol{\lambda}^{\mathrm{S}}, \{\boldsymbol{w}_i^{\mathrm{S}}\}_i)$. It remains to be shown that (B.13) is equivalent to the dual best problem (D-B). To this end, we will show that

$$\lambda_i F_i^*(\boldsymbol{w}_i / \lambda_i) = \inf_{\boldsymbol{z}_i \in \mathcal{Z}} \lambda_i f_i^{*1}(\boldsymbol{w}_i / \lambda_i, \boldsymbol{z}_i) \tag{B.14}$$

for any fixed $\lambda_i \geq 0$ and $\boldsymbol{w}_i \in \mathbb{R}^{d_{\boldsymbol{x}}}$, $i \in \mathcal{I}$. Problem (D-B) is then obtained by substituting (B.14) into (B.13), and the assertion follows. To show (B.14), assume first that $\lambda_i > 0$. We then have that

$$\begin{aligned}
\lambda_i F_i^*(\boldsymbol{w}_i / \lambda_i) &= \lambda_i \sup_{\boldsymbol{x}} \left\{ \boldsymbol{x}^\top \boldsymbol{w}_i / \lambda_i - \sup_{\boldsymbol{z}_i \in \mathcal{Z}} f_i(\boldsymbol{x}, \boldsymbol{z}_i) \right\} \\
&= \inf_{\boldsymbol{z}_i \in \mathcal{Z}} \lambda_i \sup_{\boldsymbol{x}} \left\{ \boldsymbol{x}^\top \boldsymbol{w}_i / \lambda_i - f_i(\boldsymbol{x}, \boldsymbol{z}_i) \right\} \\
&= \inf_{\boldsymbol{z}_i \in \mathcal{Z}} \lambda_i f_i^{*1}(\boldsymbol{w}_i / \lambda_i, \boldsymbol{z}_i),
\end{aligned}$$

where the second equality follows from Sion's min-max theorem (Sion, 1958), which applies because $\mathcal{Z}$ is compact and f_i is a convex-concave saddle function that is continuous in each of its arguments, while the last equality follows from the definition of the partial conjugate. If $\lambda_i = 0$, on the other hand, then we have

$$0 F_i^*(\boldsymbol{w}_i / 0) \;=\; \delta^*_{\mathrm{dom}(F_i)}(\boldsymbol{w}_i) \;=\; \delta_{\{0\}}(\boldsymbol{w}_i) \;=\; \inf_{\boldsymbol{z}_i \in \mathcal{Z}} \delta^*_{\mathrm{dom}(f_i(\cdot, \boldsymbol{z}_i))}(\boldsymbol{w}_i) \;=\; \inf_{\boldsymbol{z}_i \in \mathcal{Z}} 0 f_i^{*1}(\boldsymbol{w}_i / 0, \boldsymbol{z}_i),$$

where the first equality follows from the definition of the convex perspective, while the second equality holds because $\mathrm{dom}(F_i) = \mathbb{R}^{d_{\boldsymbol{x}}}$. Similarly, the third equality holds because $\mathrm{dom}(f_i(\cdot, \boldsymbol{z}_i)) =$

$\mathbb{R}^{d_{\boldsymbol{x}}}$ for every $\boldsymbol{z}_i \in \mathbb{R}^{d_{\boldsymbol{z}}}$, while the last equality follows from the definition of the partial convex perspective. Thus, (B.14) holds for all $\boldsymbol{\lambda} \geq \mathbf{0}$ and $\boldsymbol{w}_i \in \mathbb{R}^{d_{\boldsymbol{x}}}$, $i \in \mathcal{I}$.

As for assertion *(iii)*, assume that (D-B) admits a strict Slater point $(\{\boldsymbol{w}_i^{\mathrm{S}}, \boldsymbol{z}_i^{\mathrm{S}}\}_i, \boldsymbol{\lambda}^{\mathrm{S}})$. Then the suprema of (D-B) and (D-B′) coincide due to Proposition 5 *(ii)*. Moreover, it is easy to verify that $(\{\boldsymbol{w}_i^{\mathrm{S}}\}_i, \boldsymbol{z}_0^{\mathrm{S}}, \boldsymbol{\lambda}^{\mathrm{S}}, \{\boldsymbol{v}_i^{\mathrm{S}}\}_i)$ is a strict Slater point for (D-B′) where $\boldsymbol{v}_i^{\mathrm{S}} = \boldsymbol{z}_i^{\mathrm{S}} \cdot \lambda_i^{\mathrm{S}}$ for $i \in \mathcal{I}$. Therefore, the problems (P-W′) and (D-B′) satisfy strong duality, and (P-W′) is solvable due to Theorem 6 *(ii)*. Finally, the infima of (P-W) and (P-W′) coincide, and (P-W) is solvable as (P-W′) is solvable due to Proposition 4 *(ii)*, which applies since any $\boldsymbol{z}_i^{\mathrm{S}}$ is a Slater point of $\mathcal{Z}$. The claim then follows. $\square$

Proof of Theorem 8. The statement trivially holds if either of the problems is infeasible. In the remainder of the proof we may thus assume that both (P-UQ) and (D-UQ) are feasible. Choose now an arbitrary $\mathbb{P}$ feasible in (P-UQ) and an arbitrary $(\alpha, \boldsymbol{\beta})$ feasible in (D-UQ). As $\mathbb{P}$ is feasible in (P-UQ), we have $\mathbb{E}_{\mathbb{P}}[h_j(\tilde{\boldsymbol{z}})] < +\infty$ for every $j \in \mathcal{J}$ and $\mathbb{E}_{\mathbb{P}}[g(\tilde{\boldsymbol{z}})] > -\infty$. Thanks to our conventions for infinite integrals, this ensures that $\mathbb{P}[\tilde{\boldsymbol{z}} \in \mathrm{dom}(h_j)] = 1$ for every $j \in \mathcal{J}$ and $\mathbb{P}[\tilde{\boldsymbol{z}} \in \mathrm{dom}(-g)] = 1$, respectively. This implies that $\mathbb{P}[\tilde{\boldsymbol{z}} \in \bar{\mathcal{S}}] = 1$. We then have

$$\mathbb{E}_{\mathbb{P}}[g(\tilde{\boldsymbol{z}})] \;\leq\; \mathbb{E}_{\mathbb{P}}\left[\alpha + \boldsymbol{h}(\tilde{\boldsymbol{z}})^\top \boldsymbol{\beta}\right] \;\leq\; \alpha + \boldsymbol{\mu}^\top \boldsymbol{\beta},$$

where the first inequality follows from the constraints in (D-UQ) and our insight that $\mathbb{P}[\tilde{\boldsymbol{z}} \in \bar{\mathcal{S}}] = 1$, and the second inequality follows from the constraints in (P-UQ) and the nonnegativity of $\boldsymbol{\beta}$. Thus, the objective value of $(\alpha, \boldsymbol{\beta})$ in (D-UQ) is non-inferior to the objective value of $\mathbb{P}$ in (P-UQ). As the primal and dual feasible solutions $\mathbb{P}$ and $(\alpha, \boldsymbol{\beta})$ were chosen arbitrarily, we may conclude that problem (D-UQ) indeed provides an upper bound on (P-UQ). $\square$

Proof of Proposition 9. We show that the robust constraint in (D-UQ) has the same feasible region as the I robust constraints in (AP-W). As $g(\boldsymbol{z}) = \max_{i \in \mathcal{I}} g_i(\boldsymbol{z})$, the robust constraint in (D-UQ) is equivalent to

$$\begin{aligned}
\max_{i \in \mathcal{I}} \sup_{\boldsymbol{z}_i \in \bar{\mathcal{S}}} \left\{ g_i(\boldsymbol{z}_i) - \alpha - \boldsymbol{h}(\boldsymbol{z}_i)^\top \boldsymbol{\beta} \right\} \leq 0 \quad &\iff \quad \max_{i \in \mathcal{I}} \sup_{\boldsymbol{z}_i \in \bar{\mathcal{S}}_i} \left\{ g_i(\boldsymbol{z}_i) - \alpha - \boldsymbol{h}(\boldsymbol{z}_i)^\top \boldsymbol{\beta} \right\} \leq 0 \\
&\iff \quad \max_{i \in \mathcal{I}} \sup_{(\boldsymbol{z}_i, \boldsymbol{u}_i, t_i) \in \mathcal{U}_i} \left\{ t_i - \alpha - \boldsymbol{u}_i^\top \boldsymbol{\beta} \right\} \leq 0,
\end{aligned}$$

where the first equivalence holds because $\bar{\mathcal{S}}_i \subseteq \bar{\mathcal{S}}$ and because $g_i(\boldsymbol{z}_i) = -\infty$ for all $\boldsymbol{z} \notin \bar{\mathcal{S}}_i$. The second equivalence follows from (3) and the fact that $\boldsymbol{u}_i = \boldsymbol{h}(\boldsymbol{z}_i)$ and $t_i = g_i(\boldsymbol{z}_i)$ maximize the inner supremum for any fixed admissible $\boldsymbol{z}_i$, $i \in \mathcal{I}$. The last inequality in the above expression is manifestly equivalent to the I robust constraints in (AP-W). $\square$

Proof of Proposition 10. As $\bar{\mathcal{S}}_i$ represents the projection of $\mathcal{U}_i$ onto $\mathbb{R}^{d_{\boldsymbol{z}}}$, problem (AD-B) is equivalent to

$$\begin{aligned} \sup \quad & \sum_{i \in \mathcal{I}} \lambda_i g_i(\boldsymbol{z}_i) \\ \text{s.t.} \quad & \sum_{i \in \mathcal{I}} \lambda_i = 1 \\ & \sum_{i \in \mathcal{I}} \lambda_i \boldsymbol{h}(\boldsymbol{z}_i) \leq \boldsymbol{\mu} \\ & \boldsymbol{z}_i \in \bar{\mathcal{S}}_i \qquad\qquad \forall i \in \mathcal{I} \\ & \boldsymbol{\lambda} \geq \boldsymbol{0}. \end{aligned} \tag{B.15}$$

Indeed, the equivalence between (AD-B) and (B.15) holds due to (3), which ensures $\boldsymbol{z}_i \in \bar{\mathcal{S}}_i$ if and only if there exist $\boldsymbol{u}_i \in \mathbb{R}^J$ and $t_i \in \mathbb{R}$ with $(\boldsymbol{z}_i, \boldsymbol{u}_i, t_i) \in \mathcal{U}_i$, and because for any fixed $(\{\boldsymbol{z}_i\}_i, \boldsymbol{\lambda})$ feasible in (B.15) it is optimal to set $\boldsymbol{u}_i = \boldsymbol{h}(\boldsymbol{z}_i)$ and $t_i = g_i(\boldsymbol{z}_i)$ for all $i \in \mathcal{I}$.

In the remainder of the proof we will show that (B.15) is equivalent to (FR). As $g(\boldsymbol{z}_i) \geq g_i(\boldsymbol{z}_i)$ for all $\boldsymbol{z}_i \in \bar{\mathcal{S}}_i$ and as $\bar{\mathcal{S}}_i \subseteq \bar{\mathcal{S}}$, it is clear that the optimal value of (FR) provides an upper bound on (B.15). To prove the converse inequality, select any $(\{\boldsymbol{z}_i\}_i, \boldsymbol{\lambda})$ feasible in (FR), and define

$$\hat{\mathcal{S}}_i = \{\boldsymbol{z} \in \bar{\mathcal{S}}_i \mid g_i(\boldsymbol{z}) \geq g_{i'}(\boldsymbol{z}) \;\; \forall i' \in \mathcal{I} : i' < i, \; g_i(\boldsymbol{z}) > g_{i'}(\boldsymbol{z}) \;\; \forall i' \in \mathcal{I} : i' > i\}$$

for all $i \in \mathcal{I}$. Note that these sets form a partition of $\bar{\mathcal{S}}$. By construction, we have $g_i(\boldsymbol{z}) = g(\boldsymbol{z}) > -\infty$ for all $\boldsymbol{z} \in \hat{\mathcal{S}}_i$. Next, define $\mathcal{I}_i = \{i' \in \mathcal{I} \mid \boldsymbol{z}_{i'} \in \hat{\mathcal{S}}_i\}$, and construct $(\{\boldsymbol{z}'_i\}_i, \boldsymbol{\lambda}')$ by setting $\lambda'_i = \sum_{i' \in \mathcal{I}_i} \lambda_{i'}$ and $\boldsymbol{z}'_i = \frac{1}{\lambda'_i} \sum_{i' \in \mathcal{I}_i} \lambda_{i'} \boldsymbol{z}_{i'}$ if $\lambda'_i > 0$. Otherwise, if $\lambda'_i = 0$, set $\boldsymbol{z}'_i$ to an arbitrary point in $\bar{\mathcal{S}}_i$, which is always possible because $\bar{\mathcal{S}}_i$ is nonempty due to assumption **(S)**. As both $-g_i$ and $\boldsymbol{h}$ are proper, closed and convex thanks to assumptions **(G)** and **(H)**, Jensen's inequality implies that

$$\sum_{i' \in \mathcal{I}_i} \lambda_{i'} g(\boldsymbol{z}_{i'}) \;\leq\; \lambda'_i g_i(\boldsymbol{z}'_i) \quad \text{and} \quad \lambda'_i \boldsymbol{h}(\boldsymbol{z}'_i) \;\leq\; \sum_{i' \in \mathcal{I}_i} \lambda_{i'} \boldsymbol{h}(\boldsymbol{z}_{i'}) \;\leq\; \lambda'_i \boldsymbol{\mu} \qquad \forall i \in \mathcal{I},$$

that is, $(\boldsymbol{\lambda}', \{\boldsymbol{z}_i'\}_i)$ is feasible in (B.15) and its objective value in (B.15) is larger or equal to that of $(\boldsymbol{\lambda}, \{\boldsymbol{z}_i\}_i)$ in (FR). Therefore, the optimal value of (FR) further provides a lower bound on the optimal value of (B.15). The above arguments also reveal that one can construct a feasible solution for (FR) from a feasible solution of (AD-B) and vice versa. Hence, the claim follows. $\square$

Proof of Theorem 11. In the absence of any regularity conditions, we have

$$\inf(\text{AP-W}') \;\geq\; \inf(\text{AP-W}) \;=\; \inf(\text{D-UQ}) \;\geq\; \sup(\text{P-UQ}) \;\geq\; \sup(\text{FR}) \;=\; \sup(\text{AD-B}) \quad (\text{B.16})$$

where the two equalities follow from Propositions 9 and 10, respectively, while the first inequality exploits Proposition 4 *(i)*, the second equality follows from the weak duality result established in Theorem 8, and the second inequality holds trivially because (FR) constitutes a restriction of (P-UQ). Proposition 5 *(i)* further implies that $\sup(\text{AD-B}) \leq \sup(\text{AD-B}')$. The relationships among the different problems are also summarized in Figure 2. It remains to be shown that either of the conditions in assertions *(i)* or *(ii)* imply the equivalence of (AD-B) and (AD-B$'$) as well as strong duality between (AP-W$'$) and (AD-B$'$).

As for assertion *(i)*, note first that the Slater point for (AD-B) can be used to construct a Slater point for (AD-B$'$) with $\boldsymbol{\lambda} > \mathbf{0}$. The suprema of (AD-B) and (AD-B$'$) then coincide thanks to Proposition 5 *(ii)* and Remark C.2. Theorem 6 *(ii)* further guarantees that the infimum of (AP-W$'$) coincides with the supremum of (AD-B$'$) and that (AP-W$'$) is solvable. This allows us to conclude that all problems in (B.16) have the same optimal value as (AD-B$'$). As (AD-B) admits a Slater point, finally, it is clear that the augmented support set $\mathcal{U}_i$ admits a Slater point for every $i \in \mathcal{I}$, and therefore Propositions 4 *(ii)* and 9 ensure that if $(\alpha^\star, \boldsymbol{\beta}^\star, \{\boldsymbol{y}_i^{(0)\star}\}_i, \{\boldsymbol{y}_{ij}^{(1)\star}\}_{ij}, \{\boldsymbol{y}_{i\ell}^{(2)\star}, \nu_{i\ell}^\star\}_{i\ell})$ solves (AP-W$'$), then $(\alpha^\star, \boldsymbol{\beta}^\star)$ solves (D-UQ).

As for assertion *(ii)*, note first that the infimum of (AP-W$'$) coincides with the supremum of (AD-B$'$) and that (AD-B$'$) is solvable. This is an immediate consequence of Theorem 6 *(ii)*, which applies because (AP-W$'$) admits a Slater point. To show that all problems in (B.16) have the same optimal value, it thus remains to prove that the suprema of (AD-B) and (AD-B$'$) coincide. To this end, fix any optimal solution $(\boldsymbol{\tau}^\star, \boldsymbol{\lambda}^\star, \{\boldsymbol{\omega}_i^\star, \boldsymbol{v}_i^\star\}_i)$ of (AD-B$'$), and assume without loss of generality that $\boldsymbol{\omega}_i^\star = \lambda_i^\star \boldsymbol{h}(\boldsymbol{v}_i^\star/\lambda_i^\star)$ and $\tau_i^\star = \lambda_i^\star g_i(\boldsymbol{v}_i^\star/\lambda_i^\star)$ for all $i \in \mathcal{I}$. We now show that this solution gives rise to an optimal solution $(\{\boldsymbol{z}_i^\star, \boldsymbol{u}_i^\star, t_i^\star\}_i, \boldsymbol{\lambda}^\star)$ of (AD-B) that attains the same optimal value. To this end, set $\boldsymbol{z}_i^\star = \boldsymbol{v}_i^\star/\lambda_i^\star$ if $\lambda_i^\star > 0$, and let $\boldsymbol{z}_i^\star$ be an arbitrary point in $\bar{\mathcal{S}}_i$ otherwise, $i \in \mathcal{I}$. If there is $i \in \mathcal{I}$ with $\lambda_i^\star = 0$, then Lemma C.8 *(i)* implies that $\boldsymbol{v}_i^\star$ is a recession direction for $\mathcal{S}$. As $\mathcal{S}$ is nonempty and bounded, this in turn implies that $\boldsymbol{v}_i^\star = \boldsymbol{0}$. Using the same reasoning as in the proof of Proposition 5 *(i)*, one can show that $\lambda_i^\star c_\ell(\boldsymbol{v}_i^\star/\lambda_i^\star) = \lambda_i^\star c_\ell(\boldsymbol{z}_i^\star)$, $\lambda_i^\star \boldsymbol{h}(\boldsymbol{v}_i^\star/\lambda_i^\star) = \lambda_i^\star \boldsymbol{h}(\boldsymbol{z}_i^\star)$ and $\lambda_i^\star g_i(\boldsymbol{v}_i^\star/\lambda_i^\star) = \lambda_i^\star g_i(\boldsymbol{z}_i^\star)$ for all $i \in \mathcal{I}$ and $\ell \in \mathcal{L}$. Setting $\boldsymbol{u}_i^\star = \boldsymbol{h}(\boldsymbol{z}_i^\star)$ and $t_i^\star = g_i(\boldsymbol{z}_i^\star)$ for all $i \in \mathcal{I}$, one readily verifies that $(\{\boldsymbol{z}_i^\star, \boldsymbol{u}_i^\star, t_i^\star\}_i, \boldsymbol{\lambda}^\star)$ is feasible in (AD-B). Moreover, since $\sum_{i\in\mathcal{I}} \tau_i^\star = \sum_{i\in\mathcal{I}} \lambda_i^\star t_i^\star$, this solution attains the same objective value as $(\boldsymbol{\tau}^\star, \boldsymbol{\lambda}^\star, \{\boldsymbol{\omega}_i^\star, \boldsymbol{v}_i^\star\}_i)$ in (AD-B′). Since (AD-B) bounds (AD-B′) from below by Proposition 5 *(i)*, $(\{\boldsymbol{z}_i^\star, \boldsymbol{u}_i^\star, t_i^\star\}_i, \boldsymbol{\lambda}^\star)$ must be optimal in (AD-B). The proof of Proposition 10 further implies that $(\{\boldsymbol{z}_i^\star\}_i, \boldsymbol{\lambda}^\star)$ solves (FR), which is a restriction of (P-UQ). As all problems in (B.16) share the same optimal value, the discrete distribution that assigns probability $\lambda_i^\star$ to the point $\boldsymbol{z}_i^\star = \boldsymbol{v}_i^\star/\lambda_i^\star$ for all $i \in \mathcal{I}$ with $\lambda_i^\star > 0$ indeed solves (P-UQ). $\square$

Proof of Proposition 12. Denote by $\mathbb{P}^{\mathrm{S}}$ the Slater distribution of (P-UQ) that exists by assumption. We will first argue that for each $i \in \mathcal{I}$ there exists a probability $\lambda_i > 0$ and a probability measure $\mathbb{P}_i^{\mathrm{S}}$ supported on $\bar{\mathcal{S}}_i$ such that $\mathbb{P}^{\mathrm{S}} = \sum_{i\in\mathcal{I}} \lambda_i \mathbb{P}_i^{\mathrm{S}}$. To see this, we define for every index set $\mathcal{I}' \subseteq \mathcal{I}$ the non-negative Borel measure $\rho_{\mathcal{I}'}$ obtained by restricting $\mathbb{P}^{\mathrm{S}}$ to $\bar{\mathcal{S}}_{\mathcal{I}'} = \{\boldsymbol{z} \in \bar{\mathcal{S}} \mid \boldsymbol{z} \in \bar{\mathcal{S}}_{i'}\ \forall i' \in \mathcal{I}',\ \boldsymbol{z} \notin \bar{\mathcal{S}}_{i'}\ \forall i' \in \mathcal{I}\backslash\mathcal{I}'\}$, that is, we set $\rho_{\mathcal{I}'}[\mathcal{B}] = \mathbb{P}^{\mathrm{S}}[\tilde{\boldsymbol{z}} \in \mathcal{B} \cap \bar{\mathcal{S}}_{\mathcal{I}'}]$ for every Borel set $\mathcal{B} \subseteq \mathbb{R}^{d_{\boldsymbol{z}}}$. By construction, we thus have $\mathbb{P}^{\mathrm{S}} = \sum_{\mathcal{I}'\subseteq\mathcal{I}} \rho_{\mathcal{I}'}$. Similarly, one readily verifies that $\sum_{\mathcal{I}'\subseteq\mathcal{I}: i\in\mathcal{I}'} \rho_{\mathcal{I}'}[\bar{\mathcal{S}}_i] = \mathbb{P}^{\mathrm{S}}[\tilde{\boldsymbol{z}} \in \bar{\mathcal{S}}_i] > 0$, which implies that for all $i \in \mathcal{I}$ there exists an index set $\mathcal{I}' \subseteq \mathcal{I}$ with $i \in \mathcal{I}'$ and $\rho_{\mathcal{I}'}[\bar{\mathcal{S}}_i] > 0$. Next, we define another family of non-negative Borel measures $\hat{\rho}_i = \sum_{\mathcal{I}'\subseteq\mathcal{I}: i\in\mathcal{I}'} \frac{1}{|\mathcal{I}'|} \rho_{\mathcal{I}'}$, $i \in \mathcal{I}$. Note that $\hat{\rho}_i$ is supported on $\bar{\mathcal{S}}_i$ and satisfies $\hat{\rho}_i[\bar{\mathcal{S}}_i] > 0$ for all $i \in \mathcal{I}$. In addition, we have $\mathbb{P}^{\mathrm{S}} = \sum_{i\in\mathcal{I}} \hat{\rho}_i$. Therefore, we can finally define $\lambda_i = \hat{\rho}_i[\bar{\mathcal{S}}_i]$ and $\mathbb{P}_i^{\mathrm{S}} = \frac{1}{\lambda_i}\hat{\rho}_i$ for all $i \in \mathcal{I}$. As desired, this construction ensures that $\lambda_i > 0$ and that the probability measure $\mathbb{P}_i^{\mathrm{S}}$ is supported on $\bar{\mathcal{S}}_i$ such that $\mathbb{P}^{\mathrm{S}} = \sum_{i\in\mathcal{I}} \lambda_i \mathbb{P}_i^{\mathrm{S}}$. Since $\mathbb{P}^{\mathrm{S}}$ is a probability measure, the last relation implies that $\sum_{i\in\mathcal{I}} \lambda_i = 1$. It is also clear that $\mathbb{P}_i^{\mathrm{S}}$ is absolutely continuous on $\mathbb{R}^{d_{\boldsymbol{z}}}$ for every $i \in \mathcal{I}$. Next, define $\boldsymbol{z}_i = \mathbb{E}_{\mathbb{P}_i^{\mathrm{S}}}[\tilde{\boldsymbol{z}}]$ for all $i \in \mathcal{I}$. As h_j is proper, closed and convex, we may then use Jensen's inequality to verify that

$$\sum_{i\in\mathcal{I}} \lambda_i h_j(\boldsymbol{z}_i) \;\leq\; \sum_{i\in\mathcal{I}} \lambda_i \mathbb{E}_{\mathbb{P}_i^{\mathrm{S}}}[h_j(\tilde{\boldsymbol{z}})] \;=\; \mathbb{E}_{\mathbb{P}^{\mathrm{S}}}[h_j(\tilde{\boldsymbol{z}})] \;\leq\; \mu_j$$

for every $j \in \mathcal{J}$, where the last inequality is strict whenever h_j is nonlinear because $\mathbb{P}^{\mathrm{S}}$ is a Slater distribution. Similarly, as c_ℓ is proper, closed and convex, Jensen's inequality implies that

$$c_\ell\left(\sum_{i\in\mathcal{I}}\lambda_i \boldsymbol{z}_i\right) \;\le\; \sum_{i\in\mathcal{I}}\lambda_i c_\ell(\boldsymbol{z}_i) \;\le\; \sum_{i\in\mathcal{I}}\lambda_i \mathbb{E}_{\mathbb{P}_i^{\mathrm{S}}}[c_\ell(\tilde{\boldsymbol{z}})] \;=\; \mathbb{E}_{\mathbb{P}^{\mathrm{S}}}[c_\ell(\tilde{\boldsymbol{z}})] \;\le\; 0$$

for every $\ell \in \mathcal{L}$, where the last inequality is strict whenever c_ℓ is nonlinear. We may thus conclude that $\sum_{i\in\mathcal{I}}\lambda_i \boldsymbol{z}_i$ is a Slater point for the support set $\mathcal{S}$, which in turn implies via Lemma C.11 that each $\boldsymbol{z}_i$, $i \in \mathcal{I}$, is a Slater point for $\mathcal{S}$ provided that $\boldsymbol{z}_i \in \mathrm{ri}(\mathcal{S})$. As $\mathbb{P}_i^{\mathrm{S}}$ is absolutely continuous on $\mathbb{R}^{d_{\boldsymbol{z}}}$ and supported on the convex set $\bar{\mathcal{S}}_i$, one can indeed prove that its mean $\boldsymbol{z}_i$ belongs even to the interior of $\bar{\mathcal{S}}_i \subseteq \mathcal{S}$. Otherwise, by the separating hyperplane theorem (Boyd and Vandenberghe, 2004, Section 2.5.1), there exist $\boldsymbol{a} \in \mathbb{R}^{d_{\boldsymbol{z}}}$, $\boldsymbol{a} \neq \boldsymbol{0}$, and $b \in \mathbb{R}$ such that $\boldsymbol{a}^\top \boldsymbol{z}_i \ge b$ and $\boldsymbol{a}^\top \boldsymbol{z} \le b$ for all $\boldsymbol{z} \in \bar{\mathcal{S}}_i$. These two inequalities imply via Theorem 1.6.6 (b) by Ash and Doléans-Dade (2000) that $\mathbb{P}_i^{\mathrm{S}}[\boldsymbol{a}^\top \tilde{\boldsymbol{z}} = b] = 1$, which, however, contradicts the absolute continuity of $\mathbb{P}_i^{\mathrm{S}}$ on $\mathbb{R}^{d_{\boldsymbol{z}}}$. We have thus shown that $\boldsymbol{z}_i$ belongs to the interior of $\bar{\mathcal{S}}_i$ and, as a consequence, in particular to the interior of $\mathrm{dom}(-g_i)$, the interior of $\mathrm{dom}(h_j)$ for every $j \in \mathcal{J}$ and the interior of $\mathcal{S}$. By Lemma C.11, $\boldsymbol{z}_i$ is thus a Slater point for $\mathcal{S}$. As $\boldsymbol{z}_i \in \bar{\mathcal{S}}_i$, we may finally select any $\boldsymbol{u}_i > \boldsymbol{h}(\boldsymbol{z}_i) \in \mathbb{R}^J$ and $t_i < g_i(\boldsymbol{z}_i) \in \mathbb{R}$ for every $i \in \mathcal{I}$. By construction, $(\{\boldsymbol{z}_i, \boldsymbol{u}_i, t_i\}_i, \boldsymbol{\lambda})$ constitutes a Slater point for (AD-B) that satisfies $\boldsymbol{\lambda} > \boldsymbol{0}$. This Slater point for (AD-B) can easily be converted to a Slater point for (AD-B′) that satisfies $\boldsymbol{\lambda} > \boldsymbol{0}$. □

Proof of Proposition 13. Denote by $(\alpha^{\mathrm{S}}, \boldsymbol{\beta}^{\mathrm{S}})$ a strict Slater point of problem (D-UQ), which exists by assumption. By using similar arguments as in Proposition 9, one can show that $(\alpha^{\mathrm{S}}, \boldsymbol{\beta}^{\mathrm{S}})$ also constitutes a strict Slater point for (AP-W). If we fix $\alpha = \alpha^{\mathrm{S}}$ and $\boldsymbol{\beta} = \boldsymbol{\beta}^{\mathrm{S}}$, then the embedded maximization problem in the i-th constraint of (AP-W) is equivalent to

$$\begin{array}{lll} \sup & g_i(\boldsymbol{z}_i) - \alpha^{\mathrm{S}} - \displaystyle\sum_{j\in\mathcal{J}} h_j(\boldsymbol{z}_i)\beta_j^{\mathrm{S}} & \\ \text{s.t.} & c_\ell(\boldsymbol{z}_i) \le 0 & \forall \ell \in \mathcal{L} \\ & \boldsymbol{z}_i \text{ free} & \end{array} \tag{B.17}$$

because the strict inequality $\boldsymbol{\beta}^{\mathrm{S}} > \boldsymbol{0}$ implies that for every fixed $\boldsymbol{z}_i$ it is optimal to set $\boldsymbol{u}_i = \boldsymbol{h}(\boldsymbol{z}_i)$ and $t_i = g_i(\boldsymbol{z}_i)$. As $(\alpha^{\mathrm{S}}, \boldsymbol{\beta}^{\mathrm{S}})$ is a strict Slater point for (AP-W), the optimal value of (B.17) is strictly smaller than 0. Note also that (B.17) can be viewed as an instance of the minimization problem (P) that satisfies assumption **(F)** because the components g_i, $i \in \mathcal{I}$, of the disutility function satisfy **(G)**, the moment functions h_j, $j \in \mathcal{J}$, satisfy **(H)** and the constraint functions c_ℓ, $\ell \in \mathcal{L}$, of the support set satisfy **(C)**. The corresponding dual minimization problem is given by

$$\begin{array}{ll} \inf & (-g_i)^*\left(\boldsymbol{y}_i^{(0)}\right) + \sum_{j\in\mathcal{J}} \beta_j^{\mathrm{S}} h_j^*\left(\frac{\boldsymbol{y}_{ij}^{(1)}}{\beta_j^{\mathrm{S}}}\right) + \sum_{\ell\in\mathcal{L}} \nu_{i\ell} c_\ell^*\left(\frac{\boldsymbol{y}_{i\ell}^{(2)}}{\nu_{i\ell}}\right) - \alpha^{\mathrm{S}} \\ \text{s.t.} & \boldsymbol{y}_i^{(0)} + \sum_{j\in\mathcal{J}} \boldsymbol{y}_{ij}^{(1)} + \sum_{\ell\in\mathcal{L}} \boldsymbol{y}_{i\ell}^{(2)} = \mathbf{0} \\ & \boldsymbol{y}_i^{(0)}, \boldsymbol{y}_{ij}^{(1)}, \boldsymbol{y}_{i\ell}^{(2)} \text{ free}, \quad \nu_{i\ell} \geq 0 \qquad \forall j\in\mathcal{J},\ \forall \ell\in\mathcal{L}. \end{array} \tag{B.18}$$

Note that the feasible region of the primal problem (B.17) coincides with $\bar{\mathcal{S}}_i$ and is thus nonempty for every $i \in \mathcal{I}$ thanks to assumption **(S)**. In addition, it constitutes a subset of $\mathcal{S}$ and is thus bounded by assumption. Theorem 2 *(ii)* then implies that problems (B.17) and (B.18) share the same optimal value, which is strictly smaller than 0. In addition, Proposition C.6 implies that problem (B.18) admits a strict Slater point $(\boldsymbol{y}_i^{(0)\mathrm{S}}, \{\boldsymbol{y}_{ij}^{(1)\mathrm{S}}\}_j, \{\boldsymbol{y}_{i\ell}^{(2)\mathrm{S}}, \nu_{i\ell}^{\mathrm{S}}\}_\ell)$ for every $i \in \mathcal{I}$. As the infimum of (B.18) is strictly smaller than 0, we may assume without loss of generality that the objective function value of this strict Slater point in (B.18) is strictly negative, too. This is a direct consequence of Remark C.1. By construction, $(\alpha^{\mathrm{S}}, \boldsymbol{\beta}^{\mathrm{S}}, \{\boldsymbol{y}_i^{(0)\mathrm{S}}\}_i, \{\boldsymbol{y}_{ij}^{(1)\mathrm{S}}\}_{ij}, \{\boldsymbol{y}_{i\ell}^{(2)\mathrm{S}}, \nu_{i\ell}^{\mathrm{S}}\}_{i\ell})$ is thus a strict Slater point for (AP-W′). □

Proof of Corollary 14. Note first that the suprema of (P-UQ) and (AD-B′) coincide by virtue of Theorem 11 *(i)*, which applies because (AD-B′) admits a Slater point $(\boldsymbol{\tau}^{\mathrm{S}}, \boldsymbol{\lambda}^{\mathrm{S}}, \{\boldsymbol{\omega}_i^{\mathrm{S}}, \boldsymbol{v}_i^{\mathrm{S}}\}_i)$ with $\boldsymbol{\lambda}^{\mathrm{S}} > \mathbf{0}$. Next, select any tolerance $\epsilon > 0$ and any ϵ-optimal solution $(\boldsymbol{\tau}^{\epsilon}, \boldsymbol{\lambda}^{\epsilon}, \{\boldsymbol{\omega}_i^{\epsilon}, \boldsymbol{v}_i^{\epsilon}\}_i)$ of problem (AD-B′). If (AD-B′) is unbounded, then we adopt the standard convention that $(\boldsymbol{\tau}^{\epsilon}, \boldsymbol{\lambda}^{\epsilon}, \{\boldsymbol{\omega}_i^{\epsilon}, \boldsymbol{v}_i^{\epsilon}\}_i)$ is feasible in (AD-B′) and that its objective function value is larger than or equal to $1/\epsilon$. Next, define

$$(\boldsymbol{\tau}^{\theta}, \boldsymbol{\lambda}^{\theta}, \{\boldsymbol{\omega}_i^{\theta}, \boldsymbol{v}_i^{\theta}\}_i) = \theta \cdot (\boldsymbol{\tau}^{\mathrm{S}}, \boldsymbol{\lambda}^{\mathrm{S}}, \{\boldsymbol{\omega}_i^{\mathrm{S}}, \boldsymbol{v}_i^{\mathrm{S}}\}_i) + (1-\theta) \cdot (\boldsymbol{\tau}^{\epsilon}, \boldsymbol{\lambda}^{\epsilon}, \{\boldsymbol{\omega}_i^{\epsilon}, \boldsymbol{v}_i^{\epsilon}\}_i)$$

for any $\theta \in [0, 1]$, and note that this solution is feasible in (AD-B′) as it constitutes a convex combination of two feasible solutions. Note also that $\boldsymbol{\lambda}^{\theta} > \mathbf{0}$ whenever $\theta > 0$. As the objective function of (AD-B′) is linear and thus continuous, there exists $\theta \in (0, 1]$ such that $(\boldsymbol{\tau}^{\theta}, \boldsymbol{\lambda}^{\theta}, \{\boldsymbol{\omega}_i^{\theta}, \boldsymbol{v}_i^{\theta}\}_i)$ represents a 2ϵ-optimal solution of (AD-B′). Next, fix such a θ, and define $\mathbb{P}$ as the discrete distribution that assigns probability $\lambda_i^{\theta} > 0$ to $\boldsymbol{v}_i^{\theta}/\lambda_i^{\theta}$ for all $i \in \mathcal{I}$. As $(\boldsymbol{\tau}^{\theta}, \boldsymbol{\lambda}^{\theta}, \{\boldsymbol{\omega}_i^{\theta}, \boldsymbol{v}_i^{\theta}\}_i)$ is feasible in (AD-B$'$), we can readily verify that $\mathbb{P}$ is supported on $\mathcal{S}$ and satisfies

$$\mathbb{E}_{\mathbb{P}}[\boldsymbol{h}(\tilde{\boldsymbol{z}})] \;=\; \sum_{i\in\mathcal{I}} \lambda_i^\theta\, \boldsymbol{h}(\boldsymbol{v}_i^\theta/\lambda_i^\theta) \;\leq\; \sum_{i\in\mathcal{I}} \boldsymbol{\omega}_i^\theta \;\leq\; \boldsymbol{\mu},$$

which implies that $\mathbb{P} \in \mathscr{P}$. Similarly, the objective function value of $\mathbb{P}$ in (AD-B$'$) satisfies

$$\mathbb{E}_{\mathbb{P}}[g(\tilde{\boldsymbol{z}})] \;=\; \sum_{i\in\mathcal{I}} \lambda_i^\theta\, g(\boldsymbol{v}_i^\theta/\lambda_i^\theta) \;\geq\; \sum_{i\in\mathcal{I}} \lambda_i^\theta\, g_i(\boldsymbol{v}_i^\theta/\lambda_i^\theta) \;\geq\; \sum_{i\in\mathcal{I}} \tau_i^\theta.$$

The last expression non-inferior to sup (AD-B$'$) $- 2\epsilon$ if the supremum of (AD-B$'$) is finite and non-inferior to $1/2\epsilon$ otherwise. As the suprema of (P-UQ) and (AD-B$'$) match, the above reasoning implies that $\mathbb{P}$ constitutes a 2ϵ-optimal solution of the original uncertainty quantification problem (P-UQ). As $\epsilon > 0$ was chosen arbitrarily, we can thus construct feasible discrete distributions with I atoms whose objective function values are arbitrarily close to the supremum of (P-UQ). $\square$

Proof of Lemma 15. For every $\boldsymbol{\lambda} \in \mathcal{C}^*\backslash\{\boldsymbol{0}\}$ we have that $\mathrm{dom}(\boldsymbol{\lambda}^\top \boldsymbol{f}) = \mathrm{dom}(\boldsymbol{f})$ and $\boldsymbol{f}(\boldsymbol{x}) \succ_{\mathcal{C}} -\boldsymbol{\infty}_{\mathcal{C}}$ if and only if $\boldsymbol{\lambda}^\top \boldsymbol{f}(\boldsymbol{x}) > -\infty$. This implies that $\boldsymbol{f}$ is proper if and only if $\boldsymbol{\lambda}^\top \boldsymbol{f}$ is proper for every $\boldsymbol{\lambda} \in \mathcal{C}^*\backslash\{\boldsymbol{0}\}$. Also, it implies that $\mathrm{dom}(\boldsymbol{f})$ is convex if and only if $\mathrm{dom}(\boldsymbol{\lambda}^\top \boldsymbol{f})$ is convex for every $\boldsymbol{\lambda} \in \mathcal{C}^*\backslash\{\boldsymbol{0}\}$. Next, select any $\boldsymbol{x}, \boldsymbol{x}' \in \mathrm{dom}(\boldsymbol{f})$ and $\theta \in [0,1]$. By the definition of the dual cone $\mathcal{C}^*$, we then have

$$\begin{aligned} & \theta \boldsymbol{f}(\boldsymbol{x}) + (1-\theta)\boldsymbol{f}(\boldsymbol{x}') - \boldsymbol{f}(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{x}') \in \mathcal{C} \\ \iff\quad & \theta \boldsymbol{\lambda}^\top \boldsymbol{f}(\boldsymbol{x}) + (1-\theta)\boldsymbol{\lambda}^\top \boldsymbol{f}(\boldsymbol{x}') - \boldsymbol{\lambda}^\top \boldsymbol{f}(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{x}') \geq 0 \quad \forall \boldsymbol{\lambda} \in \mathcal{C}^*\backslash\{\boldsymbol{0}\}, \end{aligned}$$

where the reverse implication holds because $\mathcal{C}$ is proper and convex, which implies that $\mathcal{C}^{**} = \mathcal{C}$. Thus, $\boldsymbol{f}$ is $\mathcal{C}$-convex if and only if the scalarized function $\boldsymbol{\lambda}^\top \boldsymbol{f}$ is convex for every $\boldsymbol{\lambda} \in \mathcal{C}^*\backslash\{\boldsymbol{0}\}$. $\square$

Proof of Proposition 16. Assume first that $\boldsymbol{0} \in \mathrm{dom}(\boldsymbol{f})$. As $\boldsymbol{f}$ is proper, star $\mathcal{C}$-lower semicontinuous and $\mathcal{C}$-convex, Lemma 15 implies that $\boldsymbol{\lambda}^\top \boldsymbol{f}$ is proper, closed and convex for all $\boldsymbol{\lambda} \in \mathcal{C}^*\backslash\{\boldsymbol{0}\}$. As $\boldsymbol{0} \in \mathrm{dom}(\boldsymbol{\lambda}^\top \boldsymbol{f})$, Corollary 8.5.2 and Theorem 13.3 by Rockafellar (1970) imply

$$\lim_{t\downarrow 0} t\boldsymbol{\lambda}^\top \boldsymbol{f}(\boldsymbol{x}/t) \;=\; \delta^*_{\mathrm{dom}((\boldsymbol{\lambda}^\top \boldsymbol{f})^*)}(\boldsymbol{x}) \tag{B.19}$$

for all $\boldsymbol{\lambda} \in \mathcal{C}^*\backslash\{\boldsymbol{0}\}$. Note then that the dual cone $\mathcal{C}^*$ inherits properness from $\mathcal{C}$ (Ben-Tal and Nemirovski, 2001, Corollary 1.4.1). This means that $\mathcal{C}^*$ is solid and thus contains a basis $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_{d_{\mathcal{C}}}$

of $\mathbb{R}^{d_\mathcal{C}}$. Defining the invertible matrix $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_{d_\mathcal{C}})^\top \in \mathbb{R}^{d_\mathcal{C} \times d_\mathcal{C}}$, we conclude from (B.19) that

$$\lim_{t \downarrow 0} t \boldsymbol{\Lambda} \boldsymbol{f}(\boldsymbol{x}/t) = \boldsymbol{b}(\boldsymbol{x}), \quad \text{where} \quad \boldsymbol{b}(\boldsymbol{x}) = \left(\delta^*_{\mathrm{dom}((\boldsymbol{\lambda}_1^\top \boldsymbol{f})^*)}(\boldsymbol{x}), \ldots, \delta^*_{\mathrm{dom}((\boldsymbol{\lambda}_{d_\mathcal{C}}^\top \boldsymbol{f})^*)}(\boldsymbol{x}) \right)^\top \in \mathbb{R}^{d_\mathcal{C}},$$

which ensures that $\lim_{t \downarrow 0} t \boldsymbol{f}(\boldsymbol{x}/t) = \boldsymbol{\Lambda}^{-1} \boldsymbol{b}(\boldsymbol{x})$ exists. We may thus define the function $\underline{\boldsymbol{f}}$ through

$$\underline{\boldsymbol{f}}(\boldsymbol{x}, t) = \begin{cases} t \boldsymbol{f}(\boldsymbol{x}/t) & \text{if } t > 0, \\ \lim_{t \downarrow 0} t \boldsymbol{f}(\boldsymbol{x}/t) & \text{if } t = 0. \end{cases}$$

By construction, we have $\boldsymbol{\lambda}^\top \underline{\boldsymbol{f}}(\boldsymbol{x}, 0) = \lim_{t \downarrow 0} t \boldsymbol{\lambda}^\top \boldsymbol{f}(\boldsymbol{x}/t) = \delta^*_{\mathrm{dom}((\boldsymbol{\lambda}^\top \boldsymbol{f})^*)}(\boldsymbol{x})$ for every $\boldsymbol{\lambda} \in \mathcal{C}^* \backslash \{\boldsymbol{0}\}$, and thus $\underline{\boldsymbol{f}}$ satisfies property *(iii)*. This in turn implies that $\boldsymbol{\lambda}^\top \underline{\boldsymbol{f}}$ coincides with the convex perspective of $\boldsymbol{\lambda}^\top \boldsymbol{f}$ for every $\boldsymbol{\lambda} \in \mathcal{C}^* \backslash \{\boldsymbol{0}\}$, which is proper, closed and convex by Proposition C.2. Hence, $\underline{\boldsymbol{f}}$ is star $\mathcal{C}$-lower semicontinuous by definition as well as proper and $\mathcal{C}$-convex by Lemma 15. The function $\underline{\boldsymbol{f}}$ consequently satisfies property *(i)*. Property *(ii)* holds by construction.

If $\boldsymbol{0} \notin \mathrm{dom}(\boldsymbol{f})$, then Corollary 8.5.2 by Rockafellar (1970) is no longer applicable. As $\boldsymbol{f}$ is proper by assumption, however, there exists some point $\boldsymbol{x}_0 \in \mathrm{dom}(\boldsymbol{f})$. Next, define $\boldsymbol{g} : \mathbb{R}^{d_{\boldsymbol{x}}} \to \overline{\mathbb{R}}^{d_\mathcal{C}}$ through $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x} - \boldsymbol{x}_0)$ for all $\boldsymbol{x} \in \mathbb{R}^{d_{\boldsymbol{x}}}$, and note that $\boldsymbol{g}$ is proper, closed and convex and that $\boldsymbol{0} \in \mathrm{dom}(\boldsymbol{g})$. By the first part of the proof, we may thus conclude that there exists a function $\underline{\boldsymbol{g}} : \mathbb{R}^{d_{\boldsymbol{x}}} \times \mathbb{R}_+ \to \overline{\mathbb{R}}^{d_\mathcal{C}}$ that satisfies properties *(i)–(iii)*. Next, define the function $\underline{\boldsymbol{f}}$ through $\underline{\boldsymbol{f}}(\boldsymbol{x}, t) = \underline{\boldsymbol{g}}(\boldsymbol{x} + t\boldsymbol{x}_0, t)$. It is clear that $\underline{\boldsymbol{f}}$ inherits perperness, star $\mathcal{C}$-lower semicontinuity and $\mathcal{C}$-convexity from $\underline{\boldsymbol{g}}$ and thus satisfies property *(i)*. By construction, we further have for every $t > 0$ that

$$\underline{\boldsymbol{f}}(\boldsymbol{x}, t0) \;=\; \underline{\boldsymbol{g}}(\boldsymbol{x} + t\boldsymbol{x}_0, t) \;=\; t\boldsymbol{g}(\boldsymbol{x}/t + \boldsymbol{x}_0) \;=\; t\boldsymbol{f}(\boldsymbol{x}/t),$$

where the second equality holds because $\underline{\boldsymbol{g}}$ satisfies property *(ii)*, and the third equality follows from the definition of $\boldsymbol{g}$. This shows that $\underline{\boldsymbol{f}}$ satisfies property *(ii)*. Finally, we also have

$$\boldsymbol{\lambda}^\top \underline{\boldsymbol{f}}(\boldsymbol{x}, 0) \;=\; \boldsymbol{\lambda}^\top \underline{\boldsymbol{g}}(\boldsymbol{x}, 0) \;=\; \delta^*_{\mathrm{dom}((\boldsymbol{\lambda}^\top \boldsymbol{g})^*)}(\boldsymbol{x}) \;=\; \delta^*_{\mathrm{dom}((\boldsymbol{\lambda}^\top \boldsymbol{f})^*)}(\boldsymbol{x}),$$

for every $\boldsymbol{\lambda} \in \mathcal{C}^* \backslash \{\boldsymbol{0}\}$, where the second equality holds because $\underline{\boldsymbol{g}}$ satisfies property *(iii)*, and the third equality follows from the observation that $(\boldsymbol{\lambda}^\top \boldsymbol{g})^*(\boldsymbol{w}) = (\boldsymbol{\lambda}^\top \boldsymbol{f})^*(\boldsymbol{w}) - \boldsymbol{w}^\top \boldsymbol{x}_0$ for all $\boldsymbol{w} \in \mathbb{R}^{d_{\boldsymbol{x}}}$. This reasoning shows that $\underline{\boldsymbol{f}}$ also satisfies property *(iii)*. The uniqueness of $\underline{\boldsymbol{f}}$ is a direct consequence of property *(iii)* and the observation that the proper cone $\mathcal{C}^*$ is solid. □

Proof of Proposition 19. We first show that the negative objective function of problem (5) is proper, closed and convex. Indeed, the convex perspective $-\lambda_{ik} g_i(\hat{\boldsymbol{z}}_k + \boldsymbol{v}_{ik}/\lambda_{ik})$ defined for $\lambda_{ik} \geq 0$

is proper, closed and convex by Proposition C.2 and by Theorem 12.2 of Rockafellar (1970), which apply because g_i obeys assumption **(G)**. Thus, the negative objective function of (5) constitutes a sum of proper, closed and convex functions with different arguments and is therefore also proper, closed and convex. As c_ℓ obeys assumption **(C)** and d obeys assumption **(D)**, similar arguments can be used to show that the feasible region of (5) is closed. To prove that the feasible region is also bounded, note that the first constraint group in (5) forces the non-negative variables $\{\lambda_{ik}\}_i$ to reside within a bounded simplex for every $k \in \mathcal{K}$. In addition, by Lemma C.10 there exists a constant $\delta > 0$ such that $d(\hat{\boldsymbol{z}}_k + \boldsymbol{z}, \hat{\boldsymbol{z}}_k) \geq \delta \|\boldsymbol{z}\|_2 - 1$ for all $\boldsymbol{z} \in \mathbb{R}^{d_{\boldsymbol{z}}}$ and $k \in \mathcal{K}$, and thus we have

$$\sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{I}_k} \lambda_{ik} d\left(\hat{\boldsymbol{z}}_k + \frac{\boldsymbol{v}_{ik}}{\lambda_{ik}}, \hat{\boldsymbol{z}}_k\right) \leq \epsilon \quad \Longrightarrow \quad \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{I}_k} \|\boldsymbol{v}_{ik}\|_2 \leq \frac{1+\epsilon}{\delta},$$

where we used the elementary identity $\sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{I}_k} \lambda_{ik} = 1$. Thus, the last constraint group in (5) forces the variables $\{\boldsymbol{v}_{ik}\}_{ik}$ to reside within a bounded set, as well. In conclusion, we have shown that the objective function of (5) is upper semicontinuous and that the feasible region is both closed and bounded and therefore compact. Thus, problem (5) is indeed solvable. $\square$

Proof of Proposition 20. Under the assumptions of the proposition, problem (5) is solvable and has the same optimal value as (OT). Even though (5) is reminiscent of a restriction of (OT) that evaluates the worst-case expected disutility over all I-point distributions in $\mathbb{B}_\epsilon(\hat{\mathbb{P}})$, the solvability of (5) does *not* imply that (OT) admits a maximizer, see, *e.g.*, Mohajerin Esfahani and Kuhn (2018, Example 2) or Kuhn et al. (2019, Example 4).

By construction, $\mathcal{I}_k^+$, $\mathcal{I}_k^0 := \{i \in \mathcal{I}_k \mid \lambda_{ik}^\star = 0,\ \boldsymbol{v}_{ik}^\star = \boldsymbol{0}\}$ and $\mathcal{I}_k^\infty$ form a partition of $\mathcal{I}_k$ for every $k \in \mathcal{K}$. Note that if $\lambda_{ik}^\star = 0$, then the constraints $\lambda_{ik} c_\ell(\hat{\boldsymbol{z}}_k + \boldsymbol{v}_{ik}/\lambda_{ik}) \leq 0$, $\ell \in \mathcal{L}$, of problem (5) imply via Lemma C.8 *(ii)* that $\boldsymbol{v}_{ik}^\star$ is a recession direction of the support set $\mathcal{S}$. In particular, if $\mathcal{S}$ is bounded, this implies that $\boldsymbol{v}_{ik}^\star = \boldsymbol{0}$. We may thus conclude that $\mathcal{I}_k^\infty = \emptyset$ whenever $\mathcal{S}$ is bounded. The converse implication does not hold in general.

Assume first that $\mathcal{I}_k^\infty = \emptyset$ for every $k \in \mathcal{K}$. To see that $\mathbb{P}^\star$ defined in (6) is optimal in (OT), observe that the constraints of (5) imply that $\mathbb{P}^\star \in \mathbb{B}_\epsilon(\hat{\mathbb{P}})$ and that the expected disutility $\mathbb{E}_{\mathbb{P}^\star}[g(\tilde{\boldsymbol{z}})]$ is at least as large as the optimal value $\sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{I}_k} \lambda_{ik}^\star g_i(\hat{\boldsymbol{z}}_k + \boldsymbol{v}_{ik}^\star/\lambda_{ik}^\star)$ of (5). However, as the suprema of (OT) and (5) match, it is clear that $\mathbb{P}^\star$ must be optimal in (OT).

Assume now that $\mathcal{I}_k^\infty \neq \emptyset$ for some $k \in \mathcal{K}$. To see that the discrete distributions defined in (7) are asymptotically optimal, we first show that $\mathbb{P}_n \in \mathbb{B}_\epsilon(\hat{\mathbb{P}})$ whenever $n \geq |\mathcal{I}_k^\infty|$ for every

$k \in \mathcal{K}$. Indeed, in this case it is easy to see that $\lambda_{ik}(n) \geq 0$ for all $i \in \mathcal{I}_k^+ \cup \mathcal{I}_k^\infty$ and $k \in \mathcal{K}$ and that $\sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}_k^+\cup\mathcal{I}_k^\infty} \lambda_{ik}(n) = 1$ because $\sum_{i\in\mathcal{I}_k^+} \lambda_{ik}^\star = \hat{p}_k$ for every $k \in \mathcal{K}$. In addition, note that $\boldsymbol{z}_{ik}(n) \in \mathcal{S}$ for every $i \in \mathcal{I}_k^+ \cup \mathcal{I}_k^\infty$ and $k \in \mathcal{K}$ thanks to the constraints of problem (5) and because $\boldsymbol{v}_{ik}^\star$ is a recession direction of $\mathcal{S}$ whenever $i \in \mathcal{I}_k^\infty$. In summary, these insights imply that $\mathbb{P}_n \in \mathscr{P}_0(\mathcal{S})$. Finally, moving mass $\lambda_{ik}(n)$ from $\hat{\boldsymbol{z}}_k$ to $\boldsymbol{z}_{ik}(n)$ for every $i \in \mathcal{I}_k^+ \cup \mathcal{I}_k^\infty$ and $k \in \mathcal{K}$ incurs a total cost of

$$\begin{aligned}
&\sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}_k^+\cup\mathcal{I}_k^\infty} \lambda_{ik}(n) d(\boldsymbol{z}_{ik}(n), \hat{\boldsymbol{z}}_k) \\
&\quad = \sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}_k^+} \lambda_{ik}^\star \left(1 - \frac{|\mathcal{I}_k^\infty|}{n}\right) d\left(\hat{\boldsymbol{z}}_k + \frac{\boldsymbol{v}_{ik}^\star}{\lambda_{ik}^\star}, \hat{\boldsymbol{z}}_k\right) + \sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}_k^\infty} \frac{\hat{p}_k}{n} d\left(\hat{\boldsymbol{z}}_k + n\frac{\boldsymbol{v}_{ik}^\star}{\hat{p}_k}, \hat{\boldsymbol{z}}_k\right) \\
&\quad \leq \sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}_k^+} \lambda_{ik}^\star d\left(\hat{\boldsymbol{z}}_k + \frac{\boldsymbol{v}_{ik}^\star}{\lambda_{ik}^\star}, \hat{\boldsymbol{z}}_k\right) + \sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}_k^\infty} \lim_{n\to\infty} \frac{\hat{p}_k}{n} d\left(\hat{\boldsymbol{z}}_k + n\frac{\boldsymbol{v}_{ik}^\star}{\hat{p}_k}, \hat{\boldsymbol{z}}_k\right) \\
&\quad = \sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}_k} \lambda_{ik}^\star d\left(\hat{\boldsymbol{z}}_k + \frac{\boldsymbol{v}_{ik}^\star}{\lambda_{ik}^\star}, \hat{\boldsymbol{z}}_k\right) \;\leq\; \epsilon,
\end{aligned}$$

where the first equality follows from the definitions of $\lambda_{ik}(n)$ and $\boldsymbol{z}_{ik}(n)$, and the first inequality holds because the transportation cost $d(\boldsymbol{z}, \boldsymbol{z}')$ is non-negative and convex in $\boldsymbol{z}$, which implies that both terms in the second line are non-decreasing in n. The second equality in the above expression exploits our definition of the convex perspective for $\lambda_{ik}^\star = 0$, and the second inequality follows from the constraints of problem (5). Similarly, by using our conventions for the convex perspective, one can show that the asymptotic expected disutility $\lim_{n\to\infty} \mathbb{E}_{\mathbb{P}_n}[g(\tilde{\boldsymbol{z}})]$ is at least as large as the optimal value $\sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}_k} \lambda_{ik}^\star g_i(\hat{\boldsymbol{z}}_k + \boldsymbol{v}_{ik}^\star/\lambda_{ik}^\star)$ of (5). However, as the suprema of (OT) and (5) match, it is clear that the distributions $\mathbb{P}_n$, $n \in \mathbb{N}$, must be asymptotically optimal in (OT). □

C Auxiliary Results

Proposition C.1 (Properties of Conjugate Functions)**.** *The conjugate of a function f is closed and convex. Moreover, if f is closed and convex, then $f^{**} = f$. Finally, a convex function f is proper if and only if f^* is proper.*

Proof. The conjugate f^* is a pointwise supremum of affine functions, and hence it is closed and convex. The other claims follow from Rockafellar (1970, Theorem 12.2). □

Proposition C.2 (Properties of Convex Perspective Functions)**.** *If $f : \mathbb{R}^{d_{\boldsymbol{x}}} \to \overline{\mathbb{R}}$ is proper, closed and convex, then its convex perspective is also proper, closed and convex.*

Proof. Convexity and properness follow from page 35 of Rockafellar (1970), while closedness follows from page 67 and Theorem 13.3 of Rockafellar (1970). ☐

Proposition C.3 (Conjugates of Perspective Functions)**.** *If $f : \mathbb{R}^{d_{\boldsymbol{x}}} \to \overline{\mathbb{R}}$ is proper and convex, then for any $t > 0$ we have*

$$h^*(\boldsymbol{w}) = \begin{cases} tf^*(\boldsymbol{w}/t) & \text{if } h(\boldsymbol{x}) = tf(\boldsymbol{x}), \\ tf^*(\boldsymbol{w}) & \text{if } h(\boldsymbol{x}) = tf(\boldsymbol{x}/t). \end{cases}$$

Proof. The claim follows from Theorem 16.1 of Rockafellar (1970). ☐

Proposition C.4 (Conjugates of Sums)**.** *If $g_1, \ldots, g_J : \mathbb{R}^{d_{\boldsymbol{x}}} \to \overline{\mathbb{R}}$ are proper convex functions, then*

$$\left(\sum_{j \in \mathcal{J}} g_j \right)^* (\boldsymbol{w}) \;\leq\; \inf_{\{\boldsymbol{w}_j\}_{j \in \mathcal{J}}} \left\{ \sum_{j \in \mathcal{J}} g_j^*(\boldsymbol{w}_j) \;\Big|\; \sum_{j \in \mathcal{J}} \boldsymbol{w}_j = \boldsymbol{w} \right\} \qquad \forall \boldsymbol{w} \in \mathbb{R}^{d_{\boldsymbol{x}}}. \tag{C.1}$$

If $\cap_{j \in \mathcal{J}} \mathrm{ri}(\mathrm{dom}(g_j)) \neq \emptyset$, then the inequality is tight, and the minimum is attained for every $\boldsymbol{w}$.

Proof. For any $j \in \mathcal{J}$, we denote by $\mathrm{cl}(g_j)$ the *closure* of the function g_j, that is, the largest closed function that resides underneath g_j. By the definition of the conjugate we have

$$\begin{aligned} \left(\sum_{j \in \mathcal{J}} g_j \right)^* (\boldsymbol{w}) \;&=\; \sup_{\boldsymbol{x}} \left\{ \boldsymbol{w}^\top \boldsymbol{x} - \sum_{j \in \mathcal{J}} g_j(\boldsymbol{x}) \right\} \;\leq\; \sup_{\boldsymbol{x}} \left\{ \boldsymbol{w}^\top \boldsymbol{x} - \sum_{j \in \mathcal{J}} \mathrm{cl}(g_j)(\boldsymbol{x}) \right\} \\ &=\; \left(\sum_{j \in \mathcal{J}} \mathrm{cl}(g_j) \right)^* (\boldsymbol{w}) \;\leq\; \inf_{\{\boldsymbol{w}_j\}_{j \in \mathcal{J}}} \left\{ \sum_{j \in \mathcal{J}} g_j^*(\boldsymbol{w}_j) \;\Big|\; \sum_{j \in \mathcal{J}} \boldsymbol{w}_j = \boldsymbol{w} \right\}, \end{aligned}$$

where the first inequality holds because $g_j(\boldsymbol{x}) \geq \mathrm{cl}(g_j)(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^{d_{\boldsymbol{x}}}$, $j \in \mathcal{J}$, while the second inequality follows from Theorem 16.4 of Rockafellar (1970). If $\cap_{j \in \mathcal{J}} \mathrm{ri}(\mathrm{dom}(g_j)) \neq \emptyset$, then Theorem 16.4 of Rockafellar (1970) further implies that both inequalities become equalities. ☐

Proposition C.4 asserts that the conjugate of a sum of proper convex functions (the left-hand side of (C.1)) provides a lower bound on the infimal convolution of the conjugates of these functions (the right-hand side of (C.1)). This lower bound becomes tight if the relative interiors of the domains of the convex functions have a point in common. Under the same condition one can show that the epigraph of the infimal convolution coincides with the Minkowski sum of the epigraphs of the conjugate functions g_j^*, $j \in \mathcal{J}$, see, *e.g.*, Rockafellar and Wets (2009, Exercise 1.28 (a)).

Example C.1 (Conjugates of Sums)**.** *The inequality in* (C.1) *can be strict. To see this, assume that $\boldsymbol{x} \in \mathbb{R}^2$ and $J = 2$, and define $g_1(\boldsymbol{x}) = \delta_{\{1\}}(x_1)$ and $g_2(\boldsymbol{x}) = \delta_{\{-1\}}(x_1)$. Note that both g_1 and g_2 are proper and convex. In fact, they are even closed. As the domains of g_1 and g_2 have an empty intersection, we may conclude that $(g_1 + g_2)^*(\boldsymbol{w}) = -\infty$ for every $\boldsymbol{w} \in \mathbb{R}^2$. A direct calculation shows that $g_1^*(\boldsymbol{w}) = w_1 + \delta_{\{0\}}(w_2)$ and $g_2^*(\boldsymbol{w}) = -w_1 + \delta_{\{0\}}(w_2)$, which in turn implies that*

$$\inf \left\{ g_1^*(\boldsymbol{w}_1) + g_2^*(\boldsymbol{w}_2) \;\middle|\; \boldsymbol{w}_1 + \boldsymbol{w}_2 = \boldsymbol{w} \right\} = \begin{cases} -\infty & \text{if } w_2 = 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Thus, the gap between the left and the right-hand side in (C.1) *amounts to ∞ unless $w_2 = 0$.*

Proposition C.5 (Properties of Partial Conjugates)**.** *For any function $f : \mathbb{R}^{d_{\boldsymbol{x}}} \times \mathbb{R}^{d_{\boldsymbol{z}}} \to \overline{\mathbb{R}}$, the partial conjugate f^{*1} is closed and convex (jointly in both arguments) if $-f$ is closed and convex in its second argument. Similarly, the partial conjugate f^{*2} is closed and convex if $-f$ is closed and convex in its first argument.*

Proof. For any fixed $\boldsymbol{x} \in \mathbb{R}^{d_{\boldsymbol{x}}}$, the functions $\boldsymbol{w}^\top \boldsymbol{x}$ and $-f(\boldsymbol{x}, \boldsymbol{z})$ are closed and convex in $\boldsymbol{w}$ and $\boldsymbol{z}$, respectively. Thus, the partial conjugate $f^{*1}(\boldsymbol{w}, \boldsymbol{z}) = \sup_{\boldsymbol{x} \in \mathbb{R}^{d_{\boldsymbol{x}}}} \left\{ \boldsymbol{w}^\top \boldsymbol{x} - f(\boldsymbol{x}, \boldsymbol{z}) \right\}$ is closed and convex jointly in $\boldsymbol{w}$ and $\boldsymbol{z}$ as a pointwise supremum of closed and convex functions. A similar argument can be made for the partial conjugate f^{*2}. □

Below we provide an example of two mutually dual convex optimization problems with a strictly positive duality gap. We also showcase that the presence of a positive duality gap critically depends on the representation of these problems, that is, simple equivalent reformulations of the primal (dual) may change the dual (primal) and eliminate the duality gap.

Example C.2 (Representation-Dependence of Duality Results)**.** *Consider an instance of problem* (P) *adapted from Exercise 5.21 by Boyd and Vandenberghe (2004) with two decision variables x_1 and x_2, a convex objective function $f_0(\boldsymbol{x}) = e^{-x_1}$ and a single convex constraint function defined through $f_1(\boldsymbol{x}) = x_1^2/x_2$ if $x_2 \geq 0$ and $f_1(\boldsymbol{x}) = \infty$ otherwise. The fraction x_1^2/x_2 should be interpreted as the convex perspective of x_1^2 whenever $x_2 \geq 0$. Thus, f_0 and f_1 are both proper and closed. Note* *that this instance of* (P) *violates both conditions in Theorem 2. Moreover, any feasible solution satisfies* $x_1 = 0$ *and thus attains the optimal value* 1 *of* (P)*. A direct calculation reveals that*

$$f_0^*(\boldsymbol{w}_0) = \begin{cases} -w_{01}\log(-w_{01}) + w_{01} & \text{if } w_{01} \leq 0 \textit{ and } w_{02} = 0, \\ +\infty & \text{otherwise,} \end{cases}$$

where we use the standard convention that $0\log(0) = 0$*, and*

$$f_1^*(\boldsymbol{w}_1) = \begin{cases} 0 & \text{if } \frac{1}{4}w_{11}^2 + w_{12} \leq 0, \\ +\infty & \text{otherwise.} \end{cases}$$

The dual problem (D) *can therefore be expressed as*

$$\begin{array}{ll} \sup & w_{01}\log(-w_{01}) - w_{01} \\ \text{s.t.} & \boldsymbol{w}_0 + \boldsymbol{w}_1 = \boldsymbol{0} \\ & \dfrac{w_{11}^2}{4\lambda_1} + w_{12} \leq 0, \quad w_{01} \leq 0, \quad w_{02} = 0 \\ & \boldsymbol{w}_0, \boldsymbol{w}_1 \text{ free} \\ & \lambda_1 \geq 0. \end{array}$$

Note that this instance of (D) *violates both conditions in Theorem 2. Any feasible solution satisfies* $\boldsymbol{w}_0 = \boldsymbol{w}_1 = \boldsymbol{0}$ *and thus attains the optimal value* 0 *of* (D)*. We conclude that the duality gap amounts to* 1*. Furthermore, Lemma B.1 allows to recover the primal problem* (P) *by dualizing* (D)*. However, all of these conclusions break down if we simplify* (P) *or* (D) *by eliminating redundant constraints and variables. For example, as any primal feasible solution satisfies* $x_1 = 0$*, problem* (P) *is equivalent to the linear program* $\inf\{1 \mid x_2 \geq 0\}$*, which admits a Slater point and therefore has a strong dual that is no longer equivalent to* (D)*. Similarly, as any dual feasible solution satisfies* $\boldsymbol{w}_0 = \boldsymbol{w}_1 = \boldsymbol{0}$*, problem* (D) *is equivalent to the linear program* $\sup\{0 \mid \lambda_1 \geq 0\}$*, which admits a Slater point and therefore has a strong dual that is no longer equivalent to* (P)*. Therefore, the dual of the dual may not be equivalent to the primal if one simplifies the dual problem.*

Based on the data of the primal problem (P), it is often difficult to verify whether the dual problem (D) admits a Slater point. However, a dual Slater point is guaranteed to exist whenever the primal feasible region is nonempty and bounded. This result plays an important role in Section 3, where we need to verify that certain dualized embedded optimization problems admit Slater points in order to invoke strong duality for the outer optimization problems.

Proposition C.6 (Sufficient Condition for a Dual Slater Point)**.** *If the feasible region of* (P) *is nonempty and bounded, then* (D) *admits a strict Slater point.*

The next example shows that the reverse implication of Proposition C.6 does not hold in general.

Example C.3 (Unbounded Dual Feasible Region for Primal with a Slater Point)**.** *Consider an instance of problem* (P) *with* $I = 3$, $f_0(x) = x$, $f_1(x) = x - 1$, $f_2(x) = 1 - x$ *and* $f_3(x) = -x$. *It can be easily verified that* (P) *admits a Slater point. We can readily verify that the corresponding dual problem* (D) *can be expressed as*

$$\begin{array}{ll} \sup & -\lambda_1 + \lambda_2 \\ \text{s.t.} & w_0 + w_1 + w_2 + w_3 = 0, \quad w_0 = 1 \\ & w_1 = \lambda_1, \quad w_2 = -\lambda_2, \quad w_3 \leq 0 \\ & w_0, w_1, w_2 \text{ free}, \quad \boldsymbol{\lambda} \geq \mathbf{0}. \end{array}$$

Clearly, the feasible region of (D) *is unbounded.*

The following remark is useful in Section 3 when we wish to replace embedded optimization problems with their duals without changing the feasible region of the outer optimization problem.

Remark C.1 (Strict Inequalities)**.** *If* (P) *admits a strict Slater point* $\boldsymbol{x}^{\mathrm{S}}$, *then any feasible solution* $\boldsymbol{x}$ *of* (P) *can be expressed as the limit of a sequence of strict Slater points* $\boldsymbol{x}_n = \frac{1}{n}\boldsymbol{x}^{\mathrm{S}} + (1 - \frac{1}{n})\boldsymbol{x}$, $n \in \mathbb{N}$, *and its objective function value satisfies* $f_0(\boldsymbol{x}) = \liminf_{n\to\infty} f_0(\boldsymbol{x}_n)$. *Indeed, we have*

$$f_0(\boldsymbol{x}) \;\leq\; \liminf_{n\to\infty} f_0(\boldsymbol{x}_n) \;\leq\; \liminf_{n\to\infty} \left\{ \frac{1}{n} f_0(\boldsymbol{x}^{\mathrm{S}}) + \left(1 - \frac{1}{n}\right) f_0(\boldsymbol{x}) \right\} \;=\; f_0(\boldsymbol{x}),$$

where the two inequalities follow from the closedness and the convexity of f_0, *respectively. Therefore, replacing weak inequalities by strict inequalities in* (P) *does not change the infimum of* (P) *if* (P) *admits a strict Slater point.*

Example C.4 (Non-Convexity of Problem (D-B))**.** *Consider an instance of problem* (P-W) *with* $I = d_{\boldsymbol{x}} = d_{\boldsymbol{z}} = 1$, $\mathcal{Z} = \mathbb{R}$, $f_0(x, z_0) = xz_0$ *and* $f_1(x, z_1) = \frac{1}{2}x^2 + z_1$. *Then, we can readily compute*

$$f_0^{*1}(w_0, z_0) = \begin{cases} 0 & \text{if } w_0 = z_0 \\ \infty & \text{if } w_0 \neq z_0 \end{cases} \qquad \text{and} \qquad f_1^{*1}(w_1, z_1) = \frac{w_1^2}{2} - z_1,$$

which results in the following instance of problem (D-B)*.*

$$
\begin{array}{ll}
\sup & -\dfrac{w_1^2}{2\lambda_1} + \lambda_1 z_1 \\
\text{s.t.} & w_0 + w_1 = 0, \quad w_0 = z_0 \\
& w_0, w_1, z_0, z_1 \text{ free}, \quad \lambda_1 \geq 0
\end{array}
$$

Using the format $(w_0, w_1, z_0, z_1, \lambda_1)$ *to denote solutions of* (D-B)*, it is easy to verify that both* $(1, -1, 1, 2, \frac{1}{2})$ *and* $(0, 0, 0, 0, 0)$ *are feasible in* (D-B) *with the same objective value* 0*. Even though their equally weighted convex combination* $(\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, 1, \frac{1}{4})$ *is also feasible, its objective value amounts to* $-\frac{1}{4} < \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 0$*. Therefore, the instance of* (D-B) *at hand is non-convex.*

Proposition C.7 (Conjugates of Partial Conjugates)**.** *If* $f : \mathbb{R}^{d_{\boldsymbol{x}}} \times \mathbb{R}^{d_{\boldsymbol{z}}} \to \overline{\mathbb{R}}$ *is closed and convex in its first argument, and* $-f$ *is closed and convex in its second argument, then* $(f^{*1})^* = (-f)^{*2}$ *and* $((-f)^{*2})^* = f^{*1}$*.*

Proof. The conjugate of f^{*1} with respect to both of its arguments is given by

$$
\begin{aligned}
(f^{*1})^*(\boldsymbol{x}, \boldsymbol{y}) &= \sup_{\boldsymbol{w}, \boldsymbol{z}} \left\{ \boldsymbol{x}^\top \boldsymbol{w} + \boldsymbol{y}^\top \boldsymbol{z} - f^{*1}(\boldsymbol{w}, \boldsymbol{z}) \right\} \\
&= \sup_{\boldsymbol{z}} \left\{ \boldsymbol{y}^\top \boldsymbol{z} + \sup_{\boldsymbol{w}} \left\{ \boldsymbol{x}^\top \boldsymbol{w} - f^{*1}(\boldsymbol{w}, \boldsymbol{z}) \right\} \right\} \\
&= \sup_{\boldsymbol{z}} \left\{ \boldsymbol{y}^\top \boldsymbol{z} + f(\boldsymbol{x}, \boldsymbol{z}) \right\} \;=\; (-f)^{*2}(\boldsymbol{x}, \boldsymbol{y}),
\end{aligned}
$$

where the third equality holds because f is closed and convex in its first argument, which implies that $(f^{*1})^{*1} = f$; see Proposition C.1. This establishes that $(f^{*1})^* = (-f)^{*2}$. Since f^{*1} is jointly closed and convex in both of its arguments due to Proposition C.5, Proposition C.1 further implies that $((-f)^{*2})^* = (f^{*1})^{**} = f^{*1}$. □

Lemma C.8 (Recession Directions)**.** *The following statements hold.*

(i) A vector $\boldsymbol{v} \in \mathbb{R}^{d_{\boldsymbol{z}}}$ *is a recession direction for the function* c_ℓ *if and only if* $0c_\ell(\boldsymbol{v}/0) \leq 0$*.*

(ii) A vector $\boldsymbol{v} \in \mathbb{R}^{d_{\boldsymbol{z}}}$ *is a recession direction for the set* $\mathcal{Z}$ *if and only if* $0c_\ell(\boldsymbol{v}/0) \leq 0$ *for all* $\ell \in \mathcal{L}$*.*

Proof. The result follows from Theorem 8.6 by Rockafellar (1970). To keep this paper self-contained, however, we provide an alternative proof using our notation. As for assertion *(i)*, assume first that

$\boldsymbol{v}$ is a recession direction for c_ℓ. Thus, for any $\boldsymbol{z} \in \mathbb{R}^{d_{\boldsymbol{z}}}$ with $c_\ell(\boldsymbol{z}) \leq 0$ we have

$$
\begin{aligned}
0 \;\geq\; \frac{1}{t} c_\ell(\boldsymbol{z} + t\boldsymbol{v}) \qquad \forall t > 0 \qquad \Longrightarrow \qquad & 0 \;\geq\; s c_\ell \left(\frac{s\boldsymbol{z} + \boldsymbol{v}}{s} \right) \qquad \forall s > 0 \\
\Longrightarrow \qquad & 0 \;\geq\; \liminf_{s \downarrow 0} s c_\ell \left(\frac{s\boldsymbol{z} + \boldsymbol{v}}{s} \right) \;\geq\; 0 c_\ell(\boldsymbol{v}/0).
\end{aligned}
$$

Assume next that $0c_\ell(\boldsymbol{v}/0) \leq 0$, and fix any $\boldsymbol{z} \in \mathbb{R}^{d_{\boldsymbol{z}}}$ with $c_\ell(\boldsymbol{z}) \leq 0$. Thus, we have

$$
\begin{aligned}
c_\ell(\boldsymbol{z} + t\boldsymbol{v}) \;&=\; \left(\tfrac{1}{2} \cdot 2 + \tfrac{1}{2} \cdot 0\right) c_\ell \left(\tfrac{\frac{1}{2} \cdot 2\boldsymbol{z} + \frac{1}{2} \cdot 2t\boldsymbol{v}}{\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 0} \right) && \forall t > 0 \\
&\leq\; \tfrac{1}{2} \left[2 c_\ell \left(\tfrac{2\boldsymbol{z}}{2} \right) \right] + \tfrac{1}{2} \left[0 c_\ell \left(\tfrac{2t\boldsymbol{v}}{0} \right) \right] = c_\ell(\boldsymbol{z}) + 0c_\ell(t\boldsymbol{v}/0) \leq 0 && \forall t > 0,
\end{aligned}
$$

where the first equality is trivial because $\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 0 = 1$, and the inequality follows from the convexity of the convex perspective established in Proposition C.2. The second equality exploits the properness of c_l^* and the positive homogeneity of support functions of nonempty sets. The last inequality holds because $c_\ell(\boldsymbol{z}) \leq 0$ by assumption and because $0c_\ell(\boldsymbol{v}/0) \leq 0$ implies that $0c_\ell(t\boldsymbol{v}/0) \leq 0$. As the above reasoning applies for any $t > 0$, we conclude that $\boldsymbol{v}$ is indeed a recession direction for c_ℓ.

Assertion *(ii)* follows from assertion *(i)* and the observation that $\boldsymbol{v}$ is a recession direction for $\mathcal{Z}$ if and only if $\mathcal{Z}$ is nonempty and $\boldsymbol{v}$ is a recession direction for every c_ℓ, $\ell \in \mathcal{L}$. □

The following three remarks discuss various generalizations of the main theorems of Section 3.

Remark C.2 (Equivalence of (D-B) and (D-B′) without a Strict Slater Point)**.** *Proposition 5* (ii) *remains valid if* (D-B) *admits a feasible solution* $(\{\boldsymbol{w}_i, \boldsymbol{z}_i\}_i, \boldsymbol{\lambda})$ *with* $\boldsymbol{\lambda} > \mathbf{0}$ *instead of a strict Slater point. Indeed, the only property of a strict Slater point needed in the proof is that* $\boldsymbol{\lambda} > \mathbf{0}$.

Remark C.3 (Strong Duality for (P-W) and (D-B) without a Strict Slater Point)**.** *Theorem 7 (iii) remains valid if* (D-B) *admits a Slater point* $(\{\boldsymbol{w}_i, \boldsymbol{z}_i\}_i, \boldsymbol{\lambda})$ *with* $\boldsymbol{\lambda} > \mathbf{0}$ *instead of a strict Slater point. Similarly as for Proposition 5 (ii), the only property of* $(\{\boldsymbol{w}_i, \boldsymbol{z}_i\}_i, \boldsymbol{\lambda})$ *required in the proof, beyond it being a Slater point, is that* $\boldsymbol{\lambda} > \mathbf{0}$.

Remark C.4 (Heterogeneous Uncertainty Sets)**.** *All results of Section 3 extend in a straightforward manner to situations in which the objective function and the constraints of problem* (P-W) *are equipped with individual uncertainty sets* $\mathcal{Z}_i$, $i \in \mathcal{I}_0$, *all of which satisfy assumption* ***(C)****.*

Example C.5 (Random Matrix Theory)**.** *The techniques developed in Section 5 allow us to analyze the spectral properties of random matrices governed by an ambiguous distribution. For example,* *they enable us to compute the worst-case conditional value-at-risk (CVaR) at level* $\varepsilon \in (0,1)$ *of the (negative) largest eigenvalue* $-\lambda_{\max}(\tilde{\boldsymbol{Z}})$ *of a random matrix* $\tilde{\boldsymbol{Z}}$ *in the proper convex cone* $\mathbb{S}_+^{d_{\boldsymbol{Z}}}$ *of positive semidefinite matrices within* $\mathbb{R}^{d_{\boldsymbol{Z}} \times d_{\boldsymbol{Z}}}$*. We assume that the distribution of* $\tilde{\boldsymbol{Z}}$ *belongs to*

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{S}_+^{d_{\boldsymbol{Z}}}) \;\middle|\; \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{Z}}] \preceq_{\mathbb{S}_+^{d_{\boldsymbol{Z}}}} \overline{\boldsymbol{M}},\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{Z}}^{-1}] \preceq_{\mathbb{S}_+^{d_{\boldsymbol{Z}}}} \underline{\boldsymbol{M}} \right\}.$$

Here, $\boldsymbol{Z}^{-1}$ *is a shorthand for the function* $\boldsymbol{F}(\boldsymbol{Z}) = \boldsymbol{Z}^{-1}$ *if* $\boldsymbol{Z} \succ_{\mathbb{S}_+^{d_{\boldsymbol{Z}}}} \boldsymbol{0}$ *and* $\boldsymbol{F}(\boldsymbol{Z}) = +\boldsymbol{\infty}_{\mathbb{S}_+^{d_{\boldsymbol{Z}}}}$ *otherwise. This function is proper,* $\mathbb{S}_+^{d_{\boldsymbol{Z}}}$*-convex and star* $\mathbb{S}_+^{d_{\boldsymbol{Z}}}$*-lower semicontinuous; see also Example A.2 in Appendix A. In addition,* $\lambda_{\max}(\boldsymbol{Z})$ *is proper, closed and convex in the usual sense. By using Jensen's inequality, one can verify that* $\mathcal{P}$ *is nonempty if and only if the generalized moment bounds* $\overline{\boldsymbol{M}}, \underline{\boldsymbol{M}} \in \mathbb{S}_+^{d_{\boldsymbol{Z}}}$ *satisfy* $\underline{\boldsymbol{M}}^{-1} \preceq_{\mathbb{S}_+^{d_{\boldsymbol{Z}}}} \overline{\boldsymbol{M}}$*. By the definition of the CVaR due to Rockafellar and Uryasev (2000) and by Sion's minimax theorem (Sion, 1958), the worst-case CVaR of* $-\lambda_{\max}(\tilde{\boldsymbol{Z}})$ *satisfies*

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-CVaR}_{\varepsilon}(-\lambda_{\max}(\tilde{\boldsymbol{Z}})) = \inf_{x \in \mathbb{R}} x + \frac{1}{\varepsilon} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\max\{-\lambda_{\max}(\tilde{\boldsymbol{Z}}) - x, 0\} \right].$$

The worst-case expectation in the above expression constitutes an instance of the generalized uncertainty quantification problem (P-UQ$_{\mathrm{g}}$) *that satisfies the conditions* **(**$\boldsymbol{C}_{\mathrm{g}}$**)**, **(**$\boldsymbol{G}_{\mathrm{g}}$**)**, **(**$\boldsymbol{H}_{\mathrm{g}}$**)** *and* **(**$\boldsymbol{S}_{\mathrm{g}}$**)**. *By Theorem 17, it can be reformulated as a tractable convex minimization problem, and thus the worst-case CVaR can be computed efficiently. We emphasize that this instance of* (P-UQ$_{\mathrm{g}}$) *is beyond the reach of existing methods in distributionally robust optimization.*

Lemma C.9 (Conjugates of Powers of Norms)**.** *Assume that* $\|\cdot\|$ *and* $\|\cdot\|_*$ *are mutually dual norms on* $\mathbb{R}^{d_{\boldsymbol{z}}}$ *and that* $p, q \in [1, +\infty]$ *satisfy* $\frac{1}{p} + \frac{1}{q} = 1$*. Then, the following statements hold.*

(i) The conjugate of $h(\boldsymbol{z}) = \frac{1}{p}\|\boldsymbol{z}\|^p$ *is given by* $h^*(\boldsymbol{y}) = \frac{1}{q}\|\boldsymbol{y}\|_*^q$*. Here, we interpret* $\frac{1}{p}\|\boldsymbol{z}\|^p$ *as the indicator function of the closed unit ball around* $\boldsymbol{0}$ *with respect to* $\|\cdot\|$ *if* $p = +\infty$ *and* $\frac{1}{q}\|\boldsymbol{y}\|_*^q$ *as the indicator function of the closed unit ball around* $\boldsymbol{0}$ *with respect to* $\|\cdot\|_*$ *if* $q = +\infty$*.*

(ii) The first partial conjugate of $d(\boldsymbol{z}, \boldsymbol{z}') = \|\boldsymbol{z} - \boldsymbol{z}'\|^p$ *is given by* $d^{*1}(\boldsymbol{y}, \boldsymbol{z}') = \boldsymbol{y}^\top \boldsymbol{z}' + \varphi(q)\,\|\boldsymbol{y}\|_*^q$*, where* $\varphi(q) = (q-1)^{(q-1)}/q^q$*. Here, we interpret* $\|\boldsymbol{z} - \boldsymbol{z}'\|^p$ *as the indicator function of the closed unit ball around* $\boldsymbol{z}'$ *with respect to* $\|\cdot\|$ *if* $p = +\infty$ *and* $\varphi(q)\,\|\boldsymbol{y}\|_*^q$ *as the indicator function of the closed unit ball around* $\boldsymbol{0}$ *with respect to* $\|\cdot\|_*$ *if* $q = +\infty$*.*

Proof. Assume first that $p \in (1, +\infty)$. For any fixed $\boldsymbol{z}, \boldsymbol{y} \in \mathbb{R}^{d_{\boldsymbol{z}}}$ we then have

$$\boldsymbol{z}^\top \boldsymbol{y} - \frac{1}{p}\|\boldsymbol{z}\|^p \;\leq\; \|\boldsymbol{z}\|\|\boldsymbol{y}\|_* - \frac{1}{p}\|\boldsymbol{z}\|^p \;\leq\; \max_{t \geq 0}\; t\|\boldsymbol{y}\|_* - \frac{1}{p}t^p \;=\; \frac{1}{q}\|\boldsymbol{y}\|_*^q,$$

where the first inequality follows from the definition of the dual norm, and the equality holds because the maximum over $t \geq 0$ is attained at $t^\star = \|\boldsymbol{y}\|_*^{q-1}$. Both inequalities in the above expression are tight if we set $\boldsymbol{z}$ to $\boldsymbol{z}^\star = (\|\boldsymbol{y}\|_*^{q-1}/\|\boldsymbol{y}\|)\boldsymbol{y}$. Indeed, the first inequality is tight because $\boldsymbol{z}^\star$ is parallel to $\boldsymbol{y}$, and the second one is tight because $\|\boldsymbol{z}^\star\| = t^\star$. Therefore, we have

$$h^*(\boldsymbol{y}) \;=\; \sup_{\boldsymbol{z}} \left\{ \boldsymbol{z}^\top \boldsymbol{y} - \frac{1}{p}\|\boldsymbol{z}\|^p \right\} \;=\; \frac{1}{q}\|\boldsymbol{y}\|_*^q.$$

Standard limit arguments show the claim for $p \in \{1, +\infty\}$, and thus assertion *(i)* follows.

Assume now again that $p \in (1, +\infty)$. By the definition of partial conjugates, we then have

$$\begin{aligned} d^{*1}(\boldsymbol{y}, \boldsymbol{z}') \;&=\; \sup_{\boldsymbol{z}} \left\{ \boldsymbol{y}^\top \boldsymbol{z} - \|\boldsymbol{z} - \boldsymbol{z}'\|^p \right\} \;=\; \boldsymbol{y}^\top \boldsymbol{z}' + p \sup_{\boldsymbol{z}} \left\{ \left(\frac{\boldsymbol{y}}{p}\right)^\top \boldsymbol{z} - \frac{1}{p}\|\boldsymbol{z}\|^p \right\} \\ &=\; \boldsymbol{y}^\top \boldsymbol{z}' + \frac{p}{q}\left\|\frac{\boldsymbol{y}}{p}\right\|_*^q \;=\; \boldsymbol{y}^\top \boldsymbol{z}' + \varphi(q)\, \|\boldsymbol{y}\|_*^q, \end{aligned}$$

where the second and the third equality follow from the variable substitution $\boldsymbol{z} \leftarrow \boldsymbol{z} - \boldsymbol{z}'$ and from assertion *(i)*, respectively, while the last equality follows from elementary algebra. Standard limit arguments can again be used to prove the claim for $p \in \{1, +\infty\}$, and thus assertion *(ii)* follows. $\square$

Lemma C.10 (Growth of Non-Negative Convex Functions)**.** *If $f : \mathbb{R}^{d_{\boldsymbol{z}}} \to [0, +\infty]$ is closed and convex with $f(\boldsymbol{z}) = 0$ if and only if $\boldsymbol{z} = \boldsymbol{0}$, then there is $\delta > 0$ with $f(\boldsymbol{z}) \geq \delta\|\boldsymbol{z}\|_2 - 1$ for all $\boldsymbol{z} \in \mathbb{R}^{d_{\boldsymbol{z}}}$.*

Proof. Assume for the sake of argument that there exists no $\delta > 0$ with the advertised properties. In this case, for every $n \in \mathbb{N}$ there exists $\boldsymbol{z}_n \in \mathbb{R}^{d_{\boldsymbol{z}}}$ such that $f(\boldsymbol{z}_n) < \frac{1}{n}\|\boldsymbol{z}_n\|_2 - 1$. As the unit sphere in $\mathbb{R}^{d_{\boldsymbol{z}}}$ is compact, there further exists a subsequence $\boldsymbol{z}_{n_k}$, $k \in \mathbb{N}$, and a vector $\boldsymbol{v} \in \mathbb{R}^{d_{\boldsymbol{z}}}$ that satisfies $\lim_{k\to\infty} \boldsymbol{z}_{n_k}/\|\boldsymbol{z}_{n_k}\|_2 = \boldsymbol{v}$. By construction, we thus have $\|\boldsymbol{v}\|_2 = 1$ and

$$f(\boldsymbol{v}) \;=\; \liminf_{k\to\infty} f\left(\frac{\boldsymbol{z}_{n_k}}{\|\boldsymbol{z}_{n_k}\|_2}\right) \;\leq\; \liminf_{k\to\infty} \left(1 - \frac{1}{\|\boldsymbol{z}_{n_k}\|_2}\right) f(\boldsymbol{0}) + \frac{1}{\|\boldsymbol{z}_{n_k}\|_2} f(\boldsymbol{z}_{n_k}) \;=\; 0,$$

where the first equality and the inequality follow from the lower semicontinuity and the convexity of f, respectively, while the second equality holds because $f(\boldsymbol{0}) = 0$ and because $f(\boldsymbol{z}_{n_k})/\|\boldsymbol{z}_{n_k}\|_2 \leq 1/n_k$ by the construction of $\boldsymbol{z}_{n_k}$. As f is non-negative, the above reasoning implies that $f(\boldsymbol{v}) = 0$, which in turn implies that $\boldsymbol{v} = \boldsymbol{0}$. However, this conclusion contradicts our earlier observation that $\|\boldsymbol{v}\|_2 = 1$. Hence, our hypothesis was false, and the claim follows. $\square$

Lemma C.11 (Slater Points)**.** *Assume that $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^{d_{\boldsymbol{x}}} \mid f_i(\boldsymbol{x}) \leq 0 \;\; \forall i \in \mathcal{I},\; h_j(\boldsymbol{x}) = 0 \;\; \forall j \in \mathcal{J}\}$ is a convex set defined in terms of convex inequality constraint functions f_i, $i \in \mathcal{I}$, and affine equality constraint functions h_j, $j \in \mathcal{J}$. If $\mathcal{X}$ admits a Slater point, then any $\boldsymbol{x} \in \mathrm{ri}(\mathcal{X})$ is a Slater point.*

Proof. Let $\boldsymbol{x}^{\mathrm{S}}$ be a Slater point for $\mathcal{X}$, which exists by assumption. Select any $\boldsymbol{x} \in \mathrm{ri}(\mathcal{X})$, and assume that $\boldsymbol{x} \neq \boldsymbol{x}^{\mathrm{S}}$ for otherwise the claim is trivial. As both $\boldsymbol{x}$ and $\boldsymbol{x}^{\mathrm{S}}$ are elements of the convex set $\mathcal{X}$, all points of the form $\theta\boldsymbol{x} + (1-\theta)\boldsymbol{x}^{\mathrm{S}}$ for some $\theta \in \mathbb{R}$ belong to the affine hull of $\mathcal{X}$. In addition, as $\boldsymbol{x} \in \mathrm{ri}(\mathcal{X})$, we may thus conclude that there exists $\varepsilon > 0$ such that $\theta\boldsymbol{x} + (1-\theta)\boldsymbol{x}^{\mathrm{S}} \in \mathcal{X}$ for all $\theta \in [0, 1+\varepsilon]$. Setting $\theta = 1+\varepsilon$, we thus find $f_i((1+\varepsilon)\boldsymbol{x} - \varepsilon\boldsymbol{x}^{\mathrm{S}}) \leq 0$ and consequently

$$f_i(\boldsymbol{x}) \;=\; f_i\left(\frac{1}{1+\varepsilon}\left((1+\varepsilon)\boldsymbol{x} - \varepsilon\boldsymbol{x}^{\mathrm{S}}\right) + \frac{\varepsilon}{1+\varepsilon}\boldsymbol{x}^{\mathrm{S}}\right) \;\leq\; \frac{1}{1+\varepsilon} f_i\left((1+\varepsilon)\boldsymbol{x} - \varepsilon\boldsymbol{x}^{\mathrm{S}}\right) + \frac{\varepsilon}{1+\varepsilon} f_i\left(\boldsymbol{x}^{\mathrm{S}}\right) \;<\; 0$$

for all $i \in \mathcal{I}$ such that f_i is nonlinear, where the first inequality exploits the convexity of f_i, and the second inequality holds because $f_i(\boldsymbol{x}^{\mathrm{S}}) < 0$ by the definition of a Slater point.

By using similar arguments as in the first part of the proof, one can show that there exists $\varepsilon > 0$ such that $(1+\varepsilon)\boldsymbol{x} - \varepsilon\boldsymbol{x}^{\mathrm{S}} \in \mathrm{dom}(f_i)$ for all $i \in \mathcal{I}$ and $(1+\varepsilon)\boldsymbol{x} - \varepsilon\boldsymbol{x}^{\mathrm{S}} \in \mathrm{dom}(h_j)$ for all $j \in \mathcal{J}$. As $\boldsymbol{x}^{\mathrm{S}} \in \mathrm{ri}(\mathrm{dom}(f_i))$ for all $i \in \mathcal{I}$ and $\boldsymbol{x}^{\mathrm{S}} \in \mathrm{ri}(\mathrm{dom}(h_j)$ for all $j \in \mathcal{J}$ by the definition of a Slater point, the line segment principle by Bertsekas (2009, Proposition 1.3.1) then implies that the point $\boldsymbol{x}$ on the line segment between $\boldsymbol{x}^{\mathrm{S}}$ and $\theta\boldsymbol{x} + (1-\theta)\boldsymbol{x}^{\mathrm{S}}$ belongs to $\mathrm{ri}(\mathrm{dom}(f_i))$ for all $i \in \mathcal{I}$ and to $\mathrm{ri}(\mathrm{dom}(h_j)$ for all $j \in \mathcal{J}$. Thus, $\boldsymbol{x}$ is indeed a Slater point. $\square$

Remark C.5 (Solvability of (OT) under Superlinear Transportation Costs)**.** *Assume as usual that the finite convex program* (AD-B$'_{\mathrm{OT}}$) *admits a Slater point with $\lambda_{ik} > 0$ for all $i \in \mathcal{I}_k$ and $k \in \mathcal{K}$ and that the transportation cost $d(\boldsymbol{z}, \hat{\boldsymbol{z}}_k)$ grows superlinearly in $\boldsymbol{z}$ for every $k \in \mathcal{K}$. If $(\{\lambda^\star_{ik}, \boldsymbol{v}^\star_{ik}\}_{ik})$ is a maximizer of* (5) *and $\mathcal{I}^\infty_k \neq \emptyset$ for some $k \in \mathcal{K}$, then we have*

$$\lambda^\star_{ik} d\left(\hat{\boldsymbol{z}}_k + \frac{\boldsymbol{v}^\star_{ik}}{\lambda^\star_{ik}}, \hat{\boldsymbol{z}}_k\right) \;=\; 0 d\left(\hat{\boldsymbol{z}}_k + \frac{\boldsymbol{v}^\star_{ik}}{0}, \hat{\boldsymbol{z}}_k\right) \;=\; \lim_{\lambda_{ik}\downarrow 0} \lambda_{ik} d\left(\hat{\boldsymbol{z}}_k + \frac{\boldsymbol{v}^\star_{ik}}{\lambda_{ik}}, \hat{\boldsymbol{z}}_k\right) \;=\; \infty$$

for every $i \in \mathcal{I}^\infty_k$, where the second equality follows from our conventions about perspective functions, while the third equality holds because the transportation cost grows superlinearly in the first argument. Thus, $(\{\lambda^\star_{ik}, \boldsymbol{v}^\star_{ik}\}_{ik})$ violates the transportation budget constraint, which contradicts our assumption that it is a maximizer of the feasible problem (5)*. Hence, $\mathcal{I}^\infty_k$ must be empty for every $k \in \mathcal{K}$, which in turn implies via the above discussion that* (OT) *is solvable.*

Example C.6 (Shaping the Transportation Cost)**.** *The transportation cost function $d(\boldsymbol{z}, \boldsymbol{z}')$ can be used to incorporate structural distributional information into the uncertainty quantification problem* (OT)*. For example, if it is known that $\tilde{\boldsymbol{z}}$ is supported on the non-negative orthant $\mathbb{R}^{d_{\boldsymbol{z}}}_+$ and is unlikely to have small components, then one can set $d(\boldsymbol{z}, \boldsymbol{z}') = \sum_{n=1}^{d_{\boldsymbol{z}}} (z_n - z'_n)^2 / z_n$ if $z_n > 0$*

*for every $n = 1, \dots, d_{\boldsymbol{z}}$ and $\mathrm{d}(\boldsymbol{z}, \boldsymbol{z}') = +\infty$ otherwise. This transportation cost function satisfies condition **(D)**. In addition, $\mathrm{d}(\boldsymbol{z}, \boldsymbol{z}')$ tends to $+\infty$ as $\boldsymbol{z}$ approaches the boundary of $\mathbb{R}_+^{d_{\boldsymbol{z}}}$. Thus, it is expensive to move probability mass to areas of the support set that are expected to have a low probability. One can show that the first partial conjugate of this transportation cost function is given by $\mathrm{d}^{*1}(\boldsymbol{y}, \boldsymbol{z}') = \sum_{n=1}^{d_{\boldsymbol{z}}} 2z_n'(1 - \sqrt{1 - y_n})$ if $y_n \leq 1$ for every $n = 1, \dots, d_{\boldsymbol{z}}$ and by $\mathrm{d}^{*1}(\boldsymbol{y}, \boldsymbol{z}') = +\infty$ otherwise. As another example, if it is known that the atoms of the nominal distribution represent random samples from the unknown true distribution that are corrupted by isotropic noise with variance γ^2, then one can set the transportation cost to the Huber loss function $\mathrm{d}(\boldsymbol{z}, \boldsymbol{z}') = \frac{1}{2}\|\boldsymbol{z} - \boldsymbol{z}'\|_2^2$ if $\|\boldsymbol{z} - \boldsymbol{z}'\|_2 \leq \gamma$ and $\mathrm{d}(\boldsymbol{z}, \boldsymbol{z}') = \gamma\|\boldsymbol{z} - \boldsymbol{z}'\|_2 - \frac{\gamma^2}{2}$ otherwise. This transportation cost function satisfies condition **(D)**. In addition, it ensures that the cost of moving probability mass over short distances $\leq \gamma$ is small but increases linearly over longer transportation distances. One can show that the first partial conjugate of this transportation cost function is given by $\mathrm{d}^{*1}(\boldsymbol{y}, \boldsymbol{z}') = \boldsymbol{y}^\top \boldsymbol{z}' + \frac{1}{2}\|\boldsymbol{y}\|_2^2$ if $\|\boldsymbol{y}\|_2 \leq \gamma$ and by $\mathrm{d}^{*1}(\boldsymbol{y}, \boldsymbol{z}') = +\infty$ otherwise. The results of this section imply that the uncertainty quantification problem* (OT) *can be reformulated as a finite convex minimization problem of the form* (AP-W$'_{\mathrm{OT}}$) *or* (AD-B$'_{\mathrm{OT}}$) *under either of these transportation cost functions. These reformulations are new and beyond the scope of existing methods of distributionally robust optimization.*

References

R. Ash and C. A. Doléans-Dade. *Probability and Measure Theory.* Academic Press, 2000.

A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization.* SIAM, 2001.

D. Bertsekas. *Convex Optimization Theory.* Athena Scientific, 2009.

S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

D. Kuhn, P. Mohajerin Esfahani, V.A. Nguyen, and S. Shafieezadeh-Abadeh. Wasserstein distributionally robust optimization: Theory and applications in machine learning. *INFORMS TutORials in Operations Research*, pages 130–169, 2019.

P. Mohajerin Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1-2):1–52, 2018.

R. Rockafellar. *Convex Analysis.* Princeton University Press, 1970.

R. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.

R. Rockafellar and R. Wets. *Variational Analysis.* Springer, 2009.

M. Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958.