Balancing Market

Bob Pepin

1. Introduction

This work is motivated by a problem in energy markets, where an energy producer needs to decide on the optimal quantity and price of energy to bid into a market with auction-based clearing. When the amount of energy finally produced differs from the accepted bid quantity, any excess/shortfall of energy needs to be sold/bought at an imbalance settlement price. Denoting θ the bid price, η the bid quantity, x_1 the clearing price, x_2 the imbalance settlement price and q the final energy produced, the revenue of the producer is $\eta x_1 + (q - \eta)x_2$ if the bid is accepted ($\theta \leq x_1$) and qx_2 otherwise.

In this work, we will study a generalized version of the above problem. Fix two vectors $u, v \in \mathbb{R}^d$ and for $\theta \in \mathbb{R}$, $\eta \in [0, 1]$ define the value of a bid as

$$\begin{aligned} V_{\theta,\eta}(x, q) &= \begin{cases} q(u \cdot x), & x_1 < \theta, \\ \eta(v \cdot x) + (q - \eta)(u \cdot x), & x_1 \geq \theta. \end{cases} \\ &= \eta((v - u) \cdot x)\mathbb{1}\{x_1 \geq \theta\} + qu \cdot x \end{aligned}$$

Note that the energy market problem above corresponds to $u = (0, 1), v = (1, 0)$ where the bid is accepted if the bid price θ is below or at the clearing price x_1, $x_1 \geq \theta$.

The setting in this work is probabilistic, with integrable random variables X and Q, not necessarily independent. We will consider the problem of maximizing the value of a bid over θ and η for the expected value, for arbitrary convex risk measures and in the distributionally robust setting.

The main contributions are

- We show that we can always choose $\eta = 1$ in the optimal bid when the value of a bid is measured by an arbitrary convex risk measure.
- We provide a necessary condition satisfied by the optimal value θ when X has a probability density and the risk measure is expectation.
- We give an explicit expression for the optimal value θ^* when X is Gaussian and the risk measure is expectation.
- We completely solve the distributionally robust optimization problem for a discontinuous cost function that is piecewise linear on two domains in $\mathbb{R}^d$ separated by a hyperplane.

Most similar to the present work is the work on two-price imbalance settlement in power systems where any deviation between bid quantity and produced quantity is penalized, leading to a newsvendor problem [3]. In two-price imbalance settlement, the key variable to optimize is quantity and the bid price is usually not even included in the set of optimization variables. This contrasts with the situation in single-price imbalance settlement as studied in this work, where bid price is the key optimization variable and quantity is trivial. This is particularly relevant due to the ongoing transition of energy markets from dual-price to single-price schemes in a number of jurisdictions, including the Nordic countries.

Existing work on Wasserstein distributionally robust optimizations focuses on deriving tractable formulations of the dual for continuous cost functions[2]. To the author's knowledge, the present work is the first work studying a class of cost functions that are neither upper- nor lower-semicontinuous.

2. Optimal bid quantities

For the expected value, we note that we have the following expression for the expectation of the bid value:

$$\mathbb{E}[V_{\theta,\eta}(X, Q)] = \eta\mathbb{E}[(v - u) \cdot X; X_1 \geq \theta] + \mathbb{E}[Q(u \cdot X)].$$

From this it is clear that the bid value is affine in η and so attains its maximum either at $\eta = 0$ or $\eta = 1$. As the following theorem shows, this observation generalizes to arbitrary convex risk measures, a very general class which contains for example expectation, VaR or CVaR.

Theorem 1. *For any convex risk measure $\mathcal{R}$, any maximum (η^*, θ^*) of*

$$\mathcal{R}\left(V_{\theta,\eta}(X, Q)\right)$$

will have either $\eta^* = 0$ or $\eta^* = 1$.

Proof. To simplify notation, define two random variables

$$\begin{aligned} A_\theta &= ((v - u) \cdot X)\mathbb{1}\{X_1 \geq \theta\}, \\ B &= Q(u \cdot X) \end{aligned}$$

so that

$$V_{\theta,\eta}(X, Q) = \eta A_\theta + B.$$

Since $\mathcal{R}$ is a convex risk measure, there is an acceptance set $\mathcal{Q}$ of measures such that

$$\mathcal{R}(\eta A_\theta + B) = \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[\eta A_\theta + B]$$

so that by Sion's minimax theorem

$$\begin{aligned}\max_{\theta,\eta} \mathcal{R}(\eta A_\theta + B) &= \max_{\theta} \max_{\eta\in[0,1]} \sup_{\mathbb{Q}\in Q} \{\eta \mathbb{E}_{\mathbb{Q}}[A_\theta] + \mathbb{E}_{\mathbb{Q}}[B]\} \\ &= \max_{\theta} \sup_{\mathbb{Q}\in Q} \max_{\eta\in[0,1]} \{\eta \mathbb{E}_{\mathbb{Q}}[A_\theta] + \mathbb{E}_{\mathbb{Q}}[B]\}.\end{aligned}$$

Since the expression being maximized over η is affine, it attains its maximum at the boundaries of the domain and the conclusion follows. □

Since $V_{\theta_0,\eta=0} = \lim_{\theta\to\infty} V_{\theta,1}$ almost surely for any value of θ_0, we restrict ourselves to the case $\eta = 1$ for the rest of the paper and use the notation $V_\theta = V_{\theta,1}$.

3. Optimal bid prices

In this section, we derive a necessary condition for a bid price θ to be an optimum (minimum or maximum) of the expected revenue.

Theorem 2. *If X has a probability density, then any local optimum θ^* of $\mathbb{E}V_\theta$ satisfies*

$$E\left[u \cdot X | X_1 = \theta^*\right] = E\left[v \cdot X | X_1 = \theta^*\right].$$

Proof. Recall that $\mathbb{E}[V_\theta] = \mathbb{E}[(u - v) \cdot X; \theta \le X_1] + \mathbb{E}[v \cdot X]$ and denote $p(x_1, ..., x_d)$ the density of X. Then we can differentiate $\mathbb{E}V_\theta$ as follows:

$$\begin{aligned}\partial_\theta \mathbb{E}V_\theta &= \partial_\theta \mathbb{E}[(u - v) \cdot X; \theta \le X_1] \\ &= \partial_\theta \int_{x_1=\theta}^{\infty} \int_{x_2,\dots,x_d=-\infty}^{\infty} (u_1 - v_1)x_1 \\ &\quad + \sum_{i=2}^{d} (u_i - v_i)x_i \, p(x_1, \dots, x_d) \, dx_1 \, \dots \, dx_d \\ &= -\int_{x_2,\dots,x_d=-\infty}^{\infty} (u_1 - v_1)\theta \\ &\quad + \sum_{i=2}^{d} (u_i - v_i)x_i \, p(\theta, \dots, x_d) \, dx_2 \, \dots \, dx_d.\end{aligned}$$

Setting the derivative to zero and rearranging, we get that any local optimum θ^* satisfies

$$\begin{aligned}&-(u_1 - v_1)\theta^* \\ &= \frac{\int_{x_2,\dots,x_d=-\infty}^{\infty} \sum_{i=2}^{d} (u_i - v_i)x_i \, p(\theta, \dots, x_d) \, dx_2 \, \dots \, dx_d}{\int_{x_2,\dots,x_d=-\infty}^{\infty} p(\theta, \dots, x_d) \, dx_2 \, \dots \, dx_d} \\ &= \mathbb{E}\left[\sum_{i=2}^{d} (u_i - v_i)X_i \middle| X_1 = \theta^*\right]\end{aligned}$$

or equivalently by the properties of conditional expectation

$$\begin{aligned}&E\left[(u - v) \cdot X | X_1 = \theta^*\right] = 0 \\ \iff &E\left[u \cdot X | X_1 = \theta^*\right] = E\left[v \cdot X | X_1 = \theta^*\right].\end{aligned}$$

□

4. Examples

4.1. Many local maxima

The purpose of this section is to give an example that shows that $\mathbb{E}V_\theta$ can have an arbitrary number of local maxima and minima and that provides some insight into the single-price imbalance settlement problem.

Let $Q = 1$ be deterministic and fix $u = (0, 1), v = (1, 0)$ so that

$$\mathbb{E}V_\theta(X) = \mathbb{E}[X_2; X_1 < \theta] + \mathbb{E}[X_1; X_1 \ge \theta].$$

Suppose that X is distributed according to the empirical probability measure μ of the set of points $x^{(i)}, i = 1, \dots, N$ in $\mathbb{R}^2$, i.e. $\mu = N^{-1} \sum_{i=1}^{N} \delta_{x_i}$ where we set

$$x^{(i)} = \begin{cases} (i, 0), & i \text{ even}, \\ (i, N), & i \text{ odd}. \end{cases}$$

With this choice of measure, we have that

$$\mathbb{E}[X_2; X_1 < \theta] = N^{-1} \sum_{i=1,\dots,2k-1; i \text{ odd}} N = k \text{ for } \theta \in (2k - 1, 2k + 1]$$

and

$$\mathbb{E}[X_1; X_1 \ge \theta] = N^{-1} \sum_{i=k}^{N} i \text{ for } \theta \in (k - 1, k]$$

showing that $\mathbb{E}V_\theta$ is left-continuous with right limits and piecewise constant on intervals $(k - 1, k]$, $k = 1, \dots, N$. It is thus sufficient to study integer values of θ. We will show that $\mathbb{E}V_\theta$ has local minima at all odd values of θ and local maxima at all even values. From the preceding we have that

$$\mathbb{E}[X_2; X_1 < 2k + 1] = \mathbb{E}[X_2; X_1 < 2k] = k$$

and

$$\mathbb{E}[X_1; X_1 \ge k + 1] - \mathbb{E}[X_1; X_1 \ge k] = -k/N$$

from which it immediately follows that

$$\begin{aligned}\mathbb{E}V_{2k+1} - \mathbb{E}V_{2k} &= -2k/N < 0 \\ \mathbb{E}V_{2k} - \mathbb{E}V_{2k-1} &= 1 - (2k - 1)/N > 0\end{aligned}$$

which is what we wanted to show since $\mathbb{E}V_\theta$ is piecewise constant. Also note that $\mathbb{E}V_{2(k+1)} - \mathbb{E}V_{2k} = 1 - (4k + 1)/N$ so that for $N = 4K + 1$, $K \in \mathbb{N}$, $\mathbb{E}V_\theta$ admits a global maximum at $\theta = 2K$ since we already know that all maxima are located at even values of θ.

4.2. Unique local maximum in a Gaussian example

In this section, we study an example with a unique optimal value θ^* that can be explicitly computed. Fix $u = (0, 1)$ and $v = (1, 0)$ and suppose that X_1 and X_2 are jointly Gaussian with means μ_1, μ_2, variances σ_1^2, σ_2^2 and correlation coefficient ρ.

We have the following representation in terms of independent standard normal random variables Z_1, Z_2:

$$\begin{aligned}X_1 &= \sigma_1 Z_1 + \mu_1, \\ X_2 &= \sigma_2 \rho Z_1 + \sigma_2 \sqrt{(1 - \rho^2)} Z_2 + \mu_2.\end{aligned}$$

First, note that we have

$$\begin{aligned}
&\mathbb{E}[X_1 - X_2; X_1 \geq \theta] \\
&= \mathbb{E}\left[X_1 - X_2; Z_1 \geq \frac{\theta - \mu_1}{\sigma_1}\right] \\
&= (\mu_1 - \mu_2)\mathbb{P}\left(Z_1 \geq \frac{\theta - \mu_1}{\sigma_1}\right) + (\sigma_1 - \sigma_2\rho)\mathbb{E}\left[Z_1; Z_1 \geq \frac{\theta - \mu_1}{\sigma_1}\right] \\
&= (\mu_1 - \mu_2)(1 - F(g(\theta))) + (\sigma_1 - \sigma_2\rho)f(g(\theta))
\end{aligned}$$

with $g(\theta) = \frac{\theta - \mu_1}{\sigma_1}$ and F, f the Gaussian cumulative and probability density functions respectively. Differentiating yields

$$\begin{aligned}
&\partial_\theta \mathbb{E}[X_1 - X_2; X_1 \geq \theta] \\
&\quad = (-(\mu_1 - \mu_2) - (\sigma_1 - \sigma_2\rho)g(\theta))g'(\theta)f(\theta)
\end{aligned}$$

and evaluating the derivative at 0 shows that $\mathbb{E}[X_1 - X_2; X_1 \geq \theta]$ is increasing only if $\sigma_1 - \sigma_2\rho > 0$.

If $\mathbb{E}V_\theta$ has a maximum, we can use the necessary condition for an optimal θ^* from the previous section to find

$$\begin{aligned}
&\mathbb{E}[X_2 | X_1 = \theta^*] = \theta^* \\
\iff &\frac{\sigma_2}{\sigma_1}\rho(\theta^* - \mu_1) + \mu_2 = \theta^* \\
\iff &\theta^* = \frac{\mu_2 - \frac{\sigma_2}{\sigma_1}\rho\mu_1}{1 - \frac{\sigma_2}{\sigma_1}\rho}.
\end{aligned}$$

5. Distributionally Robust Solutions

We now consider the distributionally robust version of the problem. In this section, we make the simplifying assumption that Q is fixed so that we can assimilate it into u and v. We assume that we are given an estimate of a joint distribution of X_1, X_2, represented by a measure μ, and we seek a solution to the following distributionally robust optimization problem, given a radius r of a Wasserstein ball around μ:

$$\sup_\theta \inf_{\nu: W_1(\nu,\mu) \leq r} \mathbb{E}_\nu[V_\theta(X_1, X_2)]. \tag{1}$$

Our main computational tool will be the strong duality formula for distributionally robust optimization from [1], Proposition 2 and Theorem 1: For any Borel measure μ on a Polish space (E, d) and function $\Psi \in L^1(\mu)$ it holds that

$$\sup_{\nu: W_d(\nu,\mu) \leq r} \mathbb{E}_\nu \Psi(X) = \min_{\lambda \geq 0}\left\{\lambda r - \mathbb{E}_\mu \Phi(\lambda, X)\right\}$$

with

$$\Phi(\lambda, x) = \inf_{y \in E}[\lambda d(x, y) - \Psi(y)].$$

Applied to our setting, after some sign manipulations to account for the infimum instead of the supremum, the duality reads

$$\inf_{\nu: W_1(\nu,\mu) \leq r} \mathbb{E}_\nu[V_\theta(X_1, X_2)] = \max_{\lambda \geq 0}\left\{\mathbb{E}_\mu \Phi_\theta(\lambda, X) - \lambda r\right\} \tag{2}$$

with

$$\Phi_\theta(\lambda, x) = \inf_{y \in \mathbb{R}^d}[\lambda|x - y| + V_\theta(y)].$$

We show below in Theorem 3 that $\lambda \mapsto \mathbb{E}[\Phi_\theta(\lambda, X)]$ is concave and give an explicit expression for its supergradient $\partial_\lambda \mathbb{E}[\Phi_\theta(\lambda, X)]$ in Theorem 4. The maximum on the right-hand side of problem (2) is therefore attained for all $\lambda^*_\theta(r)$ such that

$$r \in \partial_\lambda \mathbb{E}\left[\Phi_\theta(\lambda^*_\theta(r), X)\right] \tag{3}$$

and we have

$$\inf_{\nu: W_1(\nu,\mu) \leq r} \mathbb{E}_\nu\left[V_\theta(X_1, X_2)\right] = \mathbb{E}\left[\Phi_\theta(\lambda^*_\theta(r), X)\right] - \lambda^*_\theta(r)r.$$

Using the explicit expression of the supergradient, it is straightforward to compute $\lambda^*_\theta(r)$ for a given θ either by solving the concave problem on the right-hand side of (2) using numerical methods for convex optimization such as (stochastic) subgradient descent or by solving (3) using root-finding methods.

We thus have an efficient way to evaluate the inner infimum in the original problem (1). Since $\theta \mapsto \mathbb{E}V_\theta(X)$ can in general be arbitrary and its properties depend strongly on the distribution μ, the outer optimization needs to be tailored to specific use cases. Without making assumptions on μ, a simple line search over θ is often feasible in practice.

Lemma 1. *For any $u \in \mathbb{R}^d$, $\xi \in \mathbb{R}$, $\lambda > 0$ we have*

$$\inf_{x \in \mathbb{R}^d : x_1 = \xi}\{\lambda\|x\| + u \cdot x\} = \begin{cases} \sqrt{\lambda^2 - \sum_{i \neq 1} u_i^2}\,|\xi| + u_1\xi \\ \qquad \text{if } \lambda^2 \geq \sum_{i \neq 1} u_i^2, \\ -\infty \text{ otherwise}. \end{cases}$$

When finite, the infimum is attained at a point x^* with

$$x^*_1 = \xi, \qquad x^*_{i \neq 1} = \frac{-u_i|\xi|}{\sqrt{\lambda^2 - \sum_{i \neq 1} u_i^2}}.$$

Proof. First assume that $\lambda^2 \geq \sum_{i \neq 1} u_i^2$ and let $f(x) = \lambda\|x\| + u \cdot x$. For $\xi = 0$, note that by Cauchy-Schwarz we have for any x such that $x_1 = 0$ that $f(x) \geq (\lambda - \sqrt{\sum_{i \neq 1} u_i^2})|x| \geq 0$ and the lower bound is attained at 0, consistent with our formula for x^*. For $\xi \neq 0$, $|x| > 0$ and f is convex and continuously differentiable with partial derivatives $\partial_{i \neq 1} f(x) = \lambda\frac{x_i}{|x|} + u_i$. By substitution, we can verify that $\partial_{i \neq 1} f(x^*_i) = 0$ so that the minimum is attained at x^* with value $f(x^*)$ as given in the statement. Next, define $\hat{u} \in \mathbb{R}^d$ such that $\hat{u}_1 = 0, \hat{u}_{i \neq 1} = u_i$. When $\lambda^2 < \sum_{i \neq 1} u_i^2 = |\hat{u}|^2$, we have at any x by Cauchy-Schwarz that $-\hat{u} \cdot \nabla f(x) \leq (\lambda - |\hat{u}|)|\hat{u}| < 0$ which shows that the lower bound is $-\infty$. □

Lemma 2. *For $\lambda \geq 0$, $k, a \in \mathbb{R}$ we have*

$$\inf_{x \leq a}\{\lambda|x| + kx\} = \begin{cases} -\infty \text{ if } \lambda < k, \\ (k - \lambda)a \text{ if } \lambda \geq k, a \leq 0, \\ \min\{0, (k + \lambda)a\} \text{ if } \lambda \geq k, a \geq 0. \end{cases}$$

The same result holds with strict inequality with the infimum over $\{x < a\}$.

Proof. Let $f(x) = \lambda|x| + kx$. For $a \leq 0$, on $(-\infty, a]$ we have $f(x) = (k-\lambda)x$ which is linear and thus is either unbounded below or attains its minimum at a, depending on the sign of $k-\lambda$. For $a \geq 0$, f is continuous and piecewise linear on $(-\infty, 0]$ and $[0, a]$, so that when $k - \lambda \leq 0$ the minimum is attained in $\{0, a\}$ with value $\min\{f(0), f(a)\} = \min\{0, (k+\lambda)a\}$. The result for strict inequality follows from the definition of the infimum as the greatest lower bound and continuity of the minimum in a. ☐

Let

$$\Phi_\theta(\lambda, x) = \inf_{y\in\mathbb{R}^d}\{\lambda|y-x| + V_\theta(y)\} \tag{4}$$

and for $w \in \mathbb{R}^d$ define $g(\lambda, w) = \sqrt{\lambda^2 - \sum_{i\neq 1} w_i^2}$.

Theorem 3. *For $\lambda \geq 0, x \in \mathbb{R}^d$ we have $\Phi_\theta(\lambda, x) > -\infty$ if and only if $g(\lambda, u) - u_1 \geq 0$ and $g(\lambda, v) + v_1 \geq 0$. In that case, for each x fixed, $\lambda \mapsto \Phi_\theta(\lambda, x)$ is concave, increasing, continuous and piecewise differentiable. For each λ fixed, $x \mapsto \Phi_\theta(\lambda, x)$ is continuous and piecewise affine. Furthermore, Φ_θ admits the following explicit representation:*

$$\begin{aligned}
\Phi_\theta(\lambda, x) &= \begin{cases} \min\left(\Phi_\theta^1(\lambda, x), \Phi_\theta^3(\lambda, x)\right), & x_1 \leq \theta, \\ \min\left(\Phi_\theta^2(\lambda, x), \Phi_\theta^4(\lambda, x)\right), & x_1 \geq \theta \end{cases} \\
&= \begin{cases} \Phi_\theta^1(\lambda, x), & x \in \bar{\Gamma}_\theta^1(\lambda), \\ \Phi_\theta^2(\lambda, x), & x \in \bar{\Gamma}_\theta^2(\lambda), \\ \Phi_\theta^3(\lambda, x), & x \in \bar{\Gamma}_\theta^3(\lambda), \\ \Phi_\theta^4(\lambda, x), & x \in \bar{\Gamma}_\theta^4(\lambda). \end{cases}
\end{aligned}$$

with

$$\begin{aligned}
\Phi_\theta^1(\lambda, x) &= (g(\lambda, v) + v_1)(\theta - x_1) + v \cdot x, \\
\Phi_\theta^2(\lambda, x) &= (g(\lambda, u) - u_1)(x_1 - \theta) + u \cdot x, \\
\Phi_\theta^3(\lambda, x) &= \min(0, g(\lambda, u) + u_1)(\theta - x_1) + u \cdot x, \\
\Phi_\theta^4(\lambda, x) &= \min(0, g(\lambda, v) - v_1)(x_1 - \theta) + v \cdot x.
\end{aligned}$$

The spaces $\Gamma_\theta^i(\lambda)$, $i = 1, 2, 3, 4$, are affine convex cones arising from the intersection of two halfspaces

$$\begin{aligned}
\Gamma_\theta^1 &= \{x : x_1 < \theta, \Phi_\theta^1 < \Phi_\theta^3\} \\
&= \{x : x_1 < \theta, (v - u + b_1(\lambda)e_1) \cdot x < b_1(\lambda)\theta\} \\
\Gamma_\theta^2 &= \{x : x_1 > \theta, \Phi_\theta^2 < \Phi_\theta^4\} \\
&= \{x : x_1 > \theta, (u - v + b_2(\lambda)e_1) \cdot x < b_2(\lambda)\theta\} \\
\Gamma_\theta^3 &= \{x : x_1 < \theta, \Phi_\theta^1 > \Phi_\theta^3\} \\
&= \{x : x_1 < \theta, (v - u + b_1(\lambda)e_1) \cdot x > b_1(\lambda)\theta\} \\
\Gamma_\theta^4 &= \{x : x_1 < \theta, \Phi_\theta^2 > \Phi_\theta^4\} \\
&= \{x : x_1 > \theta, (u - v + b_2(\lambda)e_1) \cdot x > b_2(\lambda)\theta\}
\end{aligned}$$

with

$$\begin{aligned}
b_1(\lambda) &= \min(0, g(\lambda, u) + u_1) - (g(\lambda, v) + v_1) \\
b_2(\lambda) &= \min(0, g(\lambda, v) - v_1) - (g(\lambda, u) - u_1).
\end{aligned}$$

Proof. On $\{y : y_1 \leq \theta\} = \{y : y_1 - x_1 \leq \theta - x_1\}$ we have by definition of V_θ and the successive application of Lemma 1 and Lemma 2

$$\begin{aligned}
&\inf_{y:y_1<\theta}\{\lambda|y-x| + V_\theta(y)\} \\
&\quad = \inf_{\xi<\theta-x_1}\ \inf_{y_1-x_1=\xi}\{\lambda|y-x| + u\cdot(y-x)\} + u\cdot x \\
&\quad = \inf_{\xi<\theta-x_1}\{g(\lambda, u)|\xi| + u_1\xi\} + u\cdot x \\
&\quad = \begin{cases} (g(\lambda, u) - u_1)(x_1 - \theta) + u\cdot x \text{ if } \theta \leq x_1, \\ \min\{0, g(\lambda, u) + u_1\}(\theta - x_1) + u\cdot x \text{ if } \theta \geq x_1. \end{cases}
\end{aligned}$$

Similarly, we have on $\{y : y_1 \geq \theta\} = \{y : x_1 - y_1 \leq x_1 - \theta\}$

$$\begin{aligned}
&\inf_{y:y_1\geq\theta}\{\lambda|y-x| + V_\theta(y)\} \\
&\quad = \inf_{\xi\leq x_1-\theta}\ \inf_{x_1-y_1=\xi}\{\lambda|x-y| - v\cdot(x-y)\} + v\cdot x \\
&\quad = \inf_{\xi\leq x_1-\theta}\{g(\lambda, v)|\xi| - v_1\xi\} + v\cdot x \\
&\quad = \begin{cases} (g(\lambda, v) + v_1)(\theta - x_1) + v\cdot x \text{ if } \theta \geq x_1, \\ \min\{0, g(\lambda, v) - v_1\}(x_1 - \theta) + v\cdot x \text{ if } \theta \leq x_1. \end{cases}
\end{aligned}$$

The expressions for Φ_θ now follow by collecting the cases for $x_1 \leq \theta$ and $x_1 \geq \theta$. Since g is concave in λ, so is each of Φ_θ^i in their corresponding halfspace, so that Φ_θ is a minimum of concave functions and therefore concave. The necessary and sufficient conditions for $\Phi_\theta > -\infty$ follow from the corresponding conditions in Lemma 1 and Lemma 2. The expressions for Γ_θ^i in terms of halfspaces follow immediately from the definitions by rearranging terms. ☐

Theorem 4. *The supergradient of the concave function $\lambda \mapsto \mathbb{E}[\Phi_\theta(\lambda, X)]$ can be written*

$$\begin{aligned}
\partial_\lambda \mathbb{E}[\Phi_\theta(\lambda, X)] = \Big\{ &\frac{\lambda}{g(\lambda, v)}\mathbb{E}\left[\theta - X_1; \Gamma_\theta^1(\lambda)\right] \\
&+ \frac{\lambda}{g(\lambda, u)}\mathbb{E}\left[X_1 - \theta; \Gamma_\theta^2(\lambda)\right] \\
&+ \frac{\lambda}{g(\lambda, u)}\mathbb{E}\left[\theta - X_1; \Gamma_\theta^3(\lambda); \lambda < |u|\right] \\
&+ \frac{\lambda}{g(\lambda, v)}\mathbb{E}\left[X_1 - \theta; \Gamma_\theta^4(\lambda); \lambda < |v|\right] \\
&+ \beta; \beta \in I_\theta(\lambda) \Big\}
\end{aligned}$$

where the interval I_θ is the convex hull of the supergradients at the boundaries of the open sets Γ_θ^i and at the points of disconti-

nuity $\lambda = |u|$ *and* $\lambda = |v|$*:*

$$
\begin{aligned}
I_\theta(\lambda) = \Bigg[&\lambda \min\left(\frac{1}{g(\lambda,u)}, \frac{1}{g(\lambda,v)}\right)\mathbb{E}\left[\theta - X_1; \bar{\Gamma}^1_\theta \cap \bar{\Gamma}^3_\theta; \lambda < |u|\right] \\
&+ \lambda \min\left(\frac{1}{g(\lambda,u)}, \frac{1}{g(\lambda,v)}\right)\mathbb{E}\left[X_1 - \theta; \bar{\Gamma}^2_\theta \cap \bar{\Gamma}^4_\theta; \lambda < |v|\right], \\
&\lambda \max\left(\frac{1}{g(\lambda,u)}, \frac{1}{g(\lambda,v)}\right)\mathbb{E}\left[\theta - X_1; \bar{\Gamma}^1_\theta \cap \bar{\Gamma}^3_\theta\right] \\
&+ \lambda \max\left(\frac{1}{g(\lambda,u)}, \frac{1}{g(\lambda,v)}\right)\mathbb{E}\left[X_1 - \theta; \bar{\Gamma}^2_\theta \cap \bar{\Gamma}^4_\theta\right] \\
&+ \frac{\lambda}{g(\lambda,u)}\mathbb{E}\left[\theta - X_1; \Gamma^3_\theta; \lambda = |u|\right] \\
&+ \frac{\lambda}{g(\lambda,v)}\mathbb{E}\left[X_1 - \theta; \Gamma^4_\theta; \lambda = |v|\right]\Bigg].
\end{aligned}
$$

Proof. By linearity, the supergradient commutes with expectation. Since $\partial_\lambda g(\lambda, w) = \lambda / g(\lambda, w)$, $w \in \mathbb{R}^d$ we have immediately that

$$
\begin{aligned}
\partial_\lambda \Phi^1_\theta &= \left\{\frac{\lambda}{g(\lambda,v)}(\theta - x_1)\right\} \\
\partial_\lambda \Phi^2_\theta &= \left\{\frac{\lambda}{g(\lambda,u)}(x_1 - \theta)\right\}
\end{aligned}
$$

yielding the supergradient of V_θ on Γ^1_θ and Γ^2_θ. Since Φ^3_θ and Φ^4_θ are not differentiable at $\lambda = |u|$ and $\lambda = |v|$ respectively, we now discuss the supergradients of all terms involving Φ^3_θ for different values of λ. The reasoning for Φ^4_θ is completely analogous and therefore omitted.

- $\lambda < |u|$:

$$\partial_\lambda \Phi^3_\theta = \left\{\frac{\lambda}{g(\lambda,u)}(\theta - x_1)\right\},$$

$$
\begin{aligned}
&\mathrm{Co}(\partial_\lambda \Phi^3_\theta \cup \partial_\lambda \Phi^1_\theta) \\
&= \left[\min(\partial_\lambda \Phi^3_\theta \cup \partial_\lambda \Phi^1_\theta), \max(\partial_\lambda \Phi^3_\theta \cup \partial_\lambda \Phi^1_\theta)\right] \\
&= \lambda(\theta - x_1)\left[\min\left(\frac{1}{g(\lambda,u)}, \frac{1}{g(\lambda,v)}\right), \max\left(\frac{1}{g(\lambda,u)}, \frac{1}{g(\lambda,v)}\right)\right].
\end{aligned}
$$

- $\lambda = |u|$:

$$\partial_\lambda \Phi^3_\theta = \left[0, \frac{\lambda}{g(\lambda,u)}(\theta - x_1)\right],$$

$$
\begin{aligned}
&\mathrm{Co}(\partial_\lambda \Phi^3_\theta \cup \partial_\lambda \Phi^1_\theta) \\
&= \left[\min(\partial_\lambda \Phi^3_\theta \cup \partial_\lambda \Phi^1_\theta), \max(\partial_\lambda \Phi^3_\theta \cup \partial_\lambda \Phi^1_\theta)\right] \\
&= \lambda(\theta - x_1)\left[0, \max\left(\frac{1}{g(\lambda,u)}, \frac{1}{g(\lambda,v)}\right)\right].
\end{aligned}
$$

- $\lambda > |u|$:

$$\partial_\lambda \Phi^3_\theta = 0,$$

$$
\begin{aligned}
&\mathrm{Co}(\partial_\lambda \Phi^3_\theta \cup \partial_\lambda \Phi^1_\theta) \\
&= \left[\min(\partial_\lambda \Phi^3_\theta \cup \partial_\lambda \Phi^1_\theta), \max(\partial_\lambda \Phi^3_\theta \cup \partial_\lambda \Phi^1_\theta)\right] \\
&= \lambda(\theta - x_1)\left[0, \max\left(\frac{1}{g(\lambda,u)}, \frac{1}{g(\lambda,v)}\right)\right].
\end{aligned}
$$

The result now follows by writing the left endpoint of $\mathrm{Co}(\partial_\lambda \Phi^3_\theta \cup \partial_\lambda \Phi^1_\theta)$ as

$$\lambda(\theta - x_1)\mathbb{1}\{\lambda < |u|\} \min\left(\frac{1}{g(\lambda,u)}, \frac{1}{g(\lambda,v)}\right)$$

and assimilating the indicator function into the expectation. The term in $\partial_\lambda \Phi^3_\theta$ is only a non-singleton for $\lambda = |u|$, so we again assimilate the indicator function into the expectation and include this case into I_θ. □

Corollary 1.

$$
\begin{aligned}
\partial_\lambda \mathbb{E}[\Phi_\theta(\lambda, X)] \ni &\frac{\lambda}{g(\lambda,v)}\mathbb{E}\left[\theta - X_1; \bar{\Gamma}^1_\theta(\lambda)\right] \\
&+ \frac{\lambda}{g(\lambda,u)}\mathbb{E}\left[X_1 - \theta; \bar{\Gamma}^2_\theta(\lambda)\right] \\
&+ \frac{\lambda}{g(\lambda,u)}\mathbb{E}\left[\theta - X_1; \Gamma^3_\theta(\lambda); \lambda < |u|\right] \\
&+ \frac{\lambda}{g(\lambda,v)}\mathbb{E}\left[X_1 - \theta; \Gamma^4_\theta(\lambda); \lambda < |v|\right]
\end{aligned}
$$

References

[1] Rui Gao and Anton J. Kleywegt. *Distributionally Robust Stochastic Optimization with Wasserstein Distance*. Apr. 30, 2022. DOI: 10.48550/arXiv.1604.02199. arXiv: 1604.02199[math]. URL: http://arxiv.org/abs/1604.02199 (visited on 04/16/2023).

[2] Peyman Mohajerin Esfahani and Daniel Kuhn. "Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations". In: *Mathematical Programming* 171.1 (Sept. 1, 2018), pp. 115–166. ISSN: 1436-4646. DOI: 10.1007/s10107-017-1172-1. URL: https://doi.org/10.1007/s10107-017-1172-1 (visited on 10/10/2023).

[3] Pierre Pinson. "Distributionally Robust Trading Strategies for Renewable Energy Producers". In: *Policy and Regulation IEEE Transactions on Energy Markets* 1.1 (Mar. 2023). Conference Name: Policy and Regulation IEEE Transactions on Energy Markets, pp. 37–47. ISSN: 2771-9626. DOI: 10.1109/TEMPR.2023.3241232. URL: https://ieeexplore.ieee.org/document/10032566 (visited on 10/10/2023).