Dual approach for two-stage robust nonlinear optimization E-Companion

Frans J.C.T. de Ruiter
Operations Research and Logistics Group, Wageningen University and Research, Wageningen, The Netherlands

Jianzhe Zhen
Risk Analytics and Optimization Chair, École Polytechnique Fédérale de Lausanne, Switzerland

Dick den Hertog
Amsterdam Business School, University of Amsterdam, Amsterdam, The Netherlands

1. Proof for Theorem 1

Proof of Theorem 1. We adopt a similar proof strategy as for Theorem 1 in Bertsimas and de Ruiter (2016). We consider the inner infimum of (1) over y for a given $x \in \mathcal{X}$ and $\zeta \in \mathcal{U}$. Since Assumption 1 holds we can apply the Lagrangian principle to (1), and obtain the following equivalent reformulation:

$$\inf_{x\in\mathcal{X}} \sup_{\substack{\zeta\in\mathcal{U}\\ v\geq 0, w}} \inf_{y} \; f_0(x) + g_0(y) + \sum_{i=1}^{m_1} v_i \left(\zeta^\top F_{i\cdot}(x) + f_i(x) + g_i(y)\right) + w^\top \left(A(x)\zeta + By - b(x)\right).$$

We then use the definition of the conjugate functions and calculus rules for conjugate functions (specifically Rule 5 in Table 2 of Roos et al. (2020) for the conjugate of the sum of convex functions) to obtain the following inf-sup-sup reformulation:

$$\inf_{x\in\mathcal{X}} \sup_{\zeta\in\mathcal{U}} \sup_{(u,v,w)\in\mathcal{W}} \; f_0(x) + \sum_{i=1}^{m_1} v_i \left(\zeta^\top F_{i\cdot}(x) + f_i(x)\right) + w^\top \left(A(x)\zeta - b(x)\right) - \sum_{i=0}^{m_1} v_i \left(g_i\right)^* \left(\frac{u_i}{v_i}\right),$$

where $\mathcal{W} = \left\{(u,v,w): \; v \geq 0, \; v_0 = 1, \; \sum_{i=0}^{m_1} u_i = -B^\top w\right\}$. We can then switch the order of supremum such that the inner supremum is over $\zeta \in \mathcal{U}$. Since the inner supremum model is linear in ζ, we can apply strong duality for linear optimization to obtain the inf-sup-inf reformulation:

$$\inf_{x\in\mathcal{X}} \sup_{(u,v,w)\in\mathcal{W}} \inf_{\lambda} \left\{ \sum_{i=0}^{m_1} v_i f_i(x) + d^\top \lambda - w^\top b(x) - \sum_{i=0}^{m_1} v_i \left(g_i\right)^* \left(\frac{u_i}{v_i}\right) \;\right|$$

$$\sum_{k=1}^{p} D_{kj}\lambda_k \geq w^\top A_{\cdot j}(x) + \sum_{i=1}^{m_1} v_i F_{ij}(x), \quad j = 1, \ldots, n_\zeta \Bigg\}.$$

We then introduce epigraph variables z_i for every $v_i\,(g_i)^*\left(\frac{u_i}{v_i}\right)$, $i = 0, ..., m_1$, and finally obtain (6).

□

2. Example when conditions in Theorem 3 are violated

Consider the following problem:

$$\min_{(x_1,x_2)\in\mathcal{X}} \max_{(\zeta_1,\zeta_2)\in\mathcal{U}} \min_{y} \left\{-y \mid -1 + x_1\zeta_1 + x_2\zeta_2 + y^2 \leq 0\right\},$$

where $\mathcal{X} = \{(x_1, x_2) \geq 0 \mid x_1 + x_2 = 1\}$ and $\mathcal{U} = \{(\zeta_1, \zeta_2) \geq 0 \mid \zeta_1 + \zeta_2 \leq 1\}$. This problem satisfies the strong relatively complete recourse condition, because for all $(x_1, x_2) \in \mathcal{X}$ and $(\zeta_2, \zeta_2) \in \mathcal{U}$ the wait-and-see decision $y = 0$ is feasible. Using Theorem 1 we can obtain the dual version of this problem:

$$\min_{(x_1,x_2)\in\mathcal{X}} \max_{v\geq 0} \min_{\lambda\geq 0} \left\{-\tfrac{1}{4v} - v + \lambda \mid \lambda \geq vx_1,\ \lambda \geq vx_2\right\}.$$

For this small problem the optimal solution can be determined without heavy computations. The optimal objective value for these problems is $-\frac{1}{\sqrt{2}}$ and is obtained for here-and-now decision $x_1^* = x_2^* = \frac{1}{2}$, and wait-and-see decision $y^* = \frac{1}{\sqrt{2}}$ in the primal formulation and $\lambda^* = \frac{1}{2}v$ in the dual formulation. The worst-case objective value in the dual formulation is achieved for $v^* = \frac{1}{\sqrt{2}}$. Suppose we solve the sampled version of the problem for only this worst-case scenario v^*. In that case, the sampled model looks like:

$$\min_{(x_1,x_2)\in\mathcal{X},\bar{\lambda}\geq 0} \left\{-\tfrac{2}{2\sqrt{2}} - \tfrac{1}{\sqrt{2}} + \bar{\lambda} \mid \bar{\lambda} \geq \tfrac{1}{\sqrt{2}}x_1,\ \bar{\lambda} \geq \tfrac{1}{\sqrt{2}}x_2\right\},$$

which has optimal objective value of $-\frac{1}{\sqrt{2}}$. Hence, the lower bound that follows from the sampled version of the dual formulation is tight. If we now want to match a critical scenario ζ^* using (12) we get

$$\begin{aligned}
\zeta^* &\in \arg\max_{\zeta\in\mathcal{U}} \{v^*(x_1^*\zeta_1 + x_2^*\zeta_2)\} \\
&= \arg\max_{\zeta\in\mathcal{U}} \left\{v^*(\tfrac{1}{2}\zeta_1 + \tfrac{1}{2}\zeta_2)\right\} \\
&= \{\zeta_1, \zeta_2 \geq 0 \mid \zeta_1 + \zeta_2 = 1\}.
\end{aligned}$$

Notice that there is no unique maximizer to (12) for this problem. If we take extreme point $(\zeta_1^*, \zeta_2^*) = (1, 0)$, then the sampled version of the primal formulation is

$$\min_{(x_1,x_2)\in\mathcal{X},\bar{y}} \left\{-\bar{y} \mid -1 + x_1 + \bar{y}^2 \leq 0\right\},$$

which has optimal objective value of -1, which is strictly lower than the optimal solution to the original problem of $-\frac{1}{\sqrt{2}}$. Hence, in contrast to instances with only right-hand-side uncertainty, the sampled version of the dual formulation can give tighter lower bounds with left-hand-side uncertainty.

3. Example 1: distribution on a network with commitments

3.1. Problem formulation

This problem is adapted from Bertsimas and de Ruiter (2016). For the distribution on a network we determine the stock allocation x_i for location i, and the contracted transporting units z_{ij} from location i to location j, $i,j=1,...,N$, prior to knowing the realization of the demand at each location. The demand ζ is uncertain and assumed to be in a budget uncertainty set:

$$\mathcal{U}=\left\{\zeta\geq 0\ :\ \zeta\leq\hat{\zeta},\ e^\top\zeta\leq\Gamma\right\},$$

where $\hat{\zeta}_i\in\mathbb{R}_+$ denotes the maximum demand at location i, $i=1,...,N$, and $\Gamma\in\mathbb{R}_+$ denotes the maximum total demand. After we observe the realization of the demand we can transport stock y_{ij} from location i to location j at cost t_{ij} in order to meet all demand, $i,j=1,...,N$. The aim is to minimize the worst case total costs, which includes the storage costs (with unit costs c_i), the cost arising from shifting the products from one location to another (after the demands are realized), and the cost from violating the committed contract. A contract is violated if the transporting units y_{ij} differentiate from the committed units z_{ij}, $i,j=1,...N$. This distribution model can now be written as a specific instance of the primal problem as follows:

$$\inf_{\substack{x\in\mathcal{X}\\ z,\tau}}\sup_{\zeta\in\mathcal{U}}\inf_{y\geq 0}\left\{\sum_{i=1}^N c_ix_i+\tau\ \middle|\ \begin{array}{l}\sum_{i,j=1}^N t_{ij}y_{ij}+\frac{1}{2}\sum_{i,j=1}^N t_{ij}(y_{ij}-z_{ij})^2\leq\tau\\ \sum_{j=1}^N y_{ji}-\sum_{j=1}^N y_{ij}\geq\zeta_i-x_i\quad i=1,...,N\end{array}\right\},\tag{1}$$

where the quadratic terms in the first constraint captures the cost of contract violation, and $\mathcal{X}=\left\{x\in\mathbb{R}_+^N\mid e^\top x\geq\Gamma,\ x_i\leq K_i\quad i=1,...,N\right\}$. The set of linear constraints in (1) are the balance equations: we have to shift stock to and from location i such that the initial storage plus the net shift in stock still exceeds the demand at i. The constraints in $\mathcal{X}$ restrict the capacity of the stock at location i to at most K_i, $i=1,...,N$, as well as the total stock to be at least the maximum demand. The dualized formulation we obtain after consecutive dualization over the adjustable variables y and the uncertain parameters ζ is given below:

$$\inf_{\substack{x\in\mathcal{X}\\ z,\tau}}\sup_{(u,v,w)\in\mathcal{V}}\inf_{\lambda\geq 0}\left\{\sum_{i=1}^N c_ix_i+\tau\ \middle|\ \begin{array}{l}\sum_{i=1}^N(\hat{\zeta}\lambda_i-u_ix_i)+\Gamma\lambda_0\ ...\\ \qquad\qquad -\sum_{i,j=1}^N\left[(u_j-u_i-t_{ij}-v_{ij})z_{ij}+\frac{1}{2}w_{ij}\right]\leq\tau\\ \lambda_0+\lambda_i\geq u_i\qquad i=1,...,N\end{array}\right\},\tag{2}$$

where $\mathcal{V}=\{(u,v,w)\geq 0:\ (u_i-u_j+v_{ij}-t_{ij})^2\leq w_{ij}t_{ij}\quad\forall i,j=1,...N\}$. Note that in both problem formulations (1) and (2), the epigraphical auxiliary variable τ can be eliminated, then it can be verified that the resulting formulations satisfy (strongly relative) complete recourse.

3.2. Numerical setting

We choose $N \in \{5, 10, 20, 30, 40, 50, 60\}$ locations uniformly at random from $[0, 10]^2$. Let t_{ij}, the cost to transport one unit of demand from location i to j, be the Euclidean distance. The unit storage cost c_i are equal to 6 for $i = 1, ..., \lceil N/10 \rceil + 1$ warehouses and 10 for $i = \lceil N/10 \rceil + 1, ..., N$ stores. The individual maximum demand $\hat{\zeta}$ and the capacity K_i, $i = 1, ..., N$, of each location is set to 30 units. The total demand in the network is set to $20\sqrt{N}$. As an illustration, Figure 1 depicts a distribution on a network obtained from solving (2) with linear decision rules, which takes around 100s. All computations were carried out with MOSEK 8.0 (MOSEK ApS, 2017) on an Intel Core(TM) i5-4590 Windows computer running at 3.30GHz with 8GB of RAM. All modeling was done using the modeling package XProg (`http://xprog.weebly.com`). All the reported numbers in the tables are the average of 10 randomly generated instances.

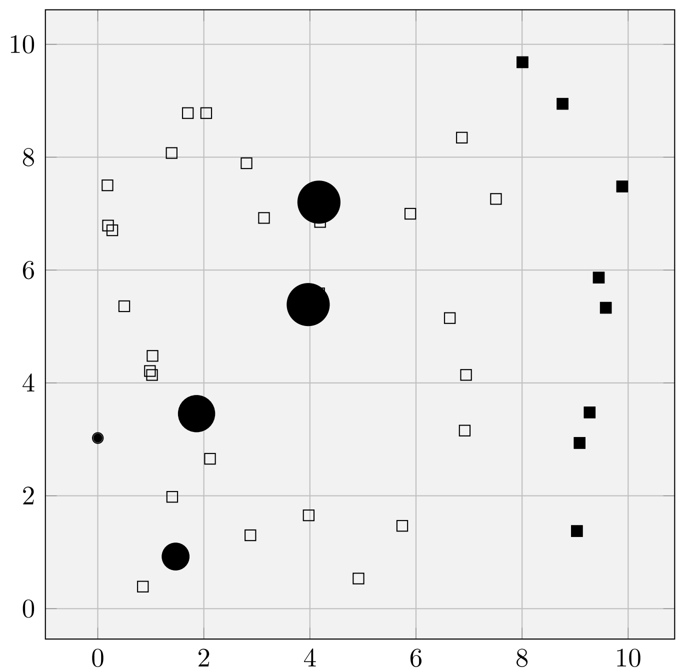

Figure 1 Stock allocation for $N = 40$ with 35 stores (squares) and 5 warehouses (circles) for one random instance. The filled dots have stock and the larger the dots are, the more stock is allocated.

3.3. Results on linear decision rules and bounds

For all cases we use linear decision rules to find solutions to the dualized model (2). For the smaller instance sizes, we also use the Fourier-Motzkin elimination of up to 10 adjustable variables. Since the model only has right-hand-side uncertainty, Theorem 3 states that we only have to use primal sampled scenarios to obtain a strong lower bound. The results are depicted in Table 1.

Table 1 Numerical results for distribution on a network with commitments. Objective values for the linear decision rule solution, Fourier-Motzkin up to 10 variable elimination and the sampled lower bound model using primal scenarios are depicted. The time refers to the computing time for the linear decision rule solution. All the numbers are the average of 10 randomly generated instances.

N	5	10	20	30	40	50	60
Linear decision rules	703	1029	1377	1606	1790	1962	2115
Fourier-Motzkin	635	944	1350	$*$	$*$	$*$	$*$
Lower bound	632	935	1272	1495	1681	1856	2004
Time linear decision rules (s)	< 0.1	0.3	14	31	118	337	665

We observe that in all cases the linear decision rules give good performance, with objective values within 10% of the lower bound for the smaller cases and within 5% for the larger cases. Furthermore, the nonlinear models with linear decision rules can all be solved within seconds for the smaller cases. For a larger number of locations the number of variables grows quadratically, which explains the computation time increases to several minutes for $N = 60$. The Fourier-Motzkin elimination only find a solution within one hour of computation time for $N \leq 20$, but has the potential to get the solution closer to the lower bound.

Remark 1. It follows from Theorem 2 that the infimum of (1) *with primal linear decision rules lower bounds that of* (2) *with dual linear decision rules. We propose to numerically evaluate the difference between the obtained infima from solving* (1) *and* (2) *with linear decision rules for* $N = 5$*. From Table 1 we observe that for* $N = 5$*, the average optimal value of* (2) *with dual linear decision rules is* 703*. By a vertex enumeration, we obtain the average optimal value of* (1) *with primal linear decision rules, i.e.,* 665*. The obtained average optimal values from solving* (1) *and* (2) *with linear decision rules are suboptimal because they are higher than that from Fourier-Motzkin elimination (i.e.,* 635*, see Table 1).*

4. Example 2: worst-case energy configuration of system with piecewise-linear springs

4.1. Problem formulation

The problem described in this section is adopted from Lobo et al. (1998). We consider a mechanical system that consists of N nodes at positions $x_1, \ldots, x_N \in \mathbb{R}^2$, with node i connected to node $i+1$, for $i = 1, \ldots, N-1$, by a nonlinear spring. The nodes x_1 and x_N are fixed at given values a and b, respectively. The tension in spring i is a nonlinear function of the distance between its endpoints, *i.e.*, $\|x_i - x_{i+1}\|_2$:

$$s\left(\|x_i - x_{i+1}\|_2 - l_i^0\right)_+ ,$$

where $z_+ = \sup\{z, 0\}$, $s \in \mathbb{R}_+$ is the stiffness of the springs, and $l_i^0 \in \mathbb{R}_+$ is the natural (no tension) length of spring i. In this model the springs can only produce positive tension (which would be the

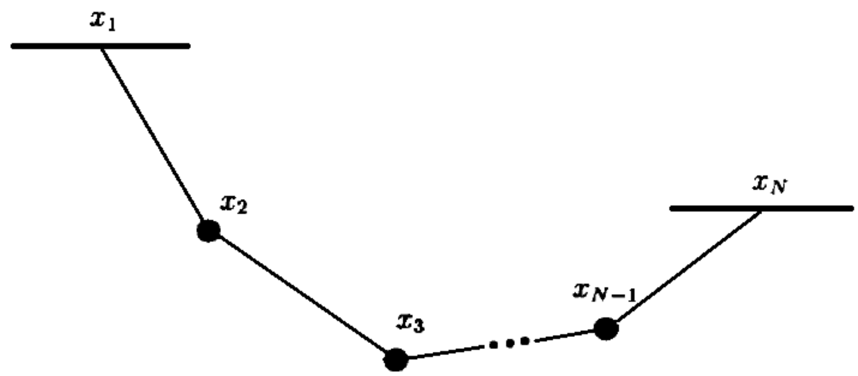

Figure 2 System of nodes (weights) connected by springs from Lobo et al. (1998). The first and last node positions, *i.e.*, x_1 and x_N, are fixed.

case if they buckled under compression). Each node has a mass of weight w attached to it. This is shown in Figure 2. The problem is to compute the equilibrium configuration of the system, *i.e.*, values of $x_1, \ldots, x_N$ such that the net force on each node is zero. This can be done by finding the minimum energy configuration, *i.e.*, solving a second-order cone optimization problem:

$$\begin{aligned} \inf_{x\geq 0} \; & w\sum_{i=1}^{N} x_{i2} + \frac{s}{2}\sum_{i=1}^{N-1}\left[\left(\|x_i - x_{i+1}\|_2 - l_i^0\right)_+\right]^2 \\ \text{s.t.} \; & x_1 = a, \quad x_N = b, \end{aligned} \tag{3}$$

where x_{i2} is the second element of the vector x_i. For more detailed description of this problem, we refer to the original paper Lobo et al. (1998). Suppose the length of the springs are uncertain. The uncertainty may arise due to variations in the production process. Of course other parameters, *e.g.*, weight(w), stiffness(s), initial location of x_1 and x_N, may also be uncertain. Here we focus on uncertainty in the length of the springs, *i.e.*, $l(\zeta) = l^0 - \zeta$ (because only positive tension is considered), and the uncertain parameter $\zeta \in \mathbb{R}^{N-1}$ resides in a budget uncertainty set:

$$\mathcal{U} = \left\{\zeta \geq 0 \;:\; \zeta \leq \hat{\zeta},\; e^\top\zeta \leq \Gamma\right\},$$

where $\hat{\zeta}_i \in \mathbb{R}_+$ denotes the maximum deviation from the nominal length l^0 of spring i, $i = 1, ..., N-1$, and $\Gamma \in \mathbb{R}_+$ denotes the maximum total deviation of the springs. The minimum energy configuration model becomes a robust optimization model:

$$\inf_{x\geq 0}\sup_{\zeta\in\mathcal{U}}\left\{ w\sum_{i=1}^{N} x_{i2} + \frac{s}{2}\sum_{i=1}^{N-1}\left[\left(\|x_i - x_{i+1}\|_2 - l_i(\zeta)\right)_+\right]^2 \;\middle|\; x_1 = a, \quad x_N = b\right\}, \tag{4}$$

which can be rewritten as a two-stage robust optimization problem:

$$\inf_{x\geq 0}\sup_{\zeta\in\mathcal{U}}\inf_{y\geq 0}\left\{ w\sum_{i=1}^{N} x_{i2} + \frac{s}{2}\sum_{i=1}^{N-1} y_i^2 \;\middle|\; \begin{array}{l} \|x_i - x_{i+1}\|_2 - l_i(\zeta) \leq y_i \;\; i = 1, ..., N-1 \\ x_1 = a, \quad x_N = b \end{array}\right\}. \tag{5}$$

It can be verified that models (4) and (5) are equivalent, that is, eliminating all the y_i's for $i = 1, \dots, N-1$ in (5) we obtain (4). We solve the dualized formulation of (5) via linear decision rules. Note that here the strong relatively complete recourse assumption is satisfied.

4.2. Numerical setting

We consider $N \in \{15, 20, 30, 45, 60, 100\}$ nodes that are connecting $N-1$ springs. The nodes x_1 and x_N are fixed at given values $a = (0, 90)$ and $b = (100, 50)$, respectively. The natural (no tension) nominal length is $l_i^0 = 1 + \epsilon_i$, where ϵ_i is a random number drawn from a uniform distribution $U(0, 4)$, $i = 1, \dots, N-1$, and the stiffness of the springs is $s = 2$. Each node has a mass of weight $w = \frac{1}{10}$ attached to it. The upper-bound $\hat{\zeta}_i$ is set at $15\% l_i^0$ for $i = 1, \dots, N-1$, and $\Gamma = \frac{1}{2} e^\top \hat{\zeta}$. The computations is carried out with MOSEK 8.1 (MOSEK ApS, 2017) on an Intel(R) Xeon(R) E3-1241 v3 Windows computer running at 3.50GHz with 16GB of RAM. All modeling was done using the modeling package XProg (`http://xprog.weebly.com`).

4.3. Results

Figure 3 illustrates the static and robust locations of the nodes for $N = 45$, which shows that in order to minimize energy configuration under length uncertainty, in the solution from linear decision rules consecutive nodes are placed closer to each other than in the solution from static decision rules. Figure 4 depicts the robust locations obtained from solving the dualized model of (5) with linear decision rules. It shows that as N increases, the curvature of the connection between x_1 and x_N becomes severer; if N is large enough, *i.e.*, $N = 100$, then there are too many nodes with positive weights, all the useless nodes will simply be closely placed on the ground.

Since the infimum obtained from linear decision rule coincides with its lower bound (LB-P) in Table 2, which implies that the approximations from solving the dualized model of (5) via linear decision rules are tight. For small N, we observe that the infimum from primal static decision rule values are larger than that from solving the dualized model of (5) via linear decision rules, which means that the approximated solutions obtained via static decision rules are suboptimal. Since the robust problem (5) becomes easier to solve as N becomes larger, the infimum from primal static decision rule values becomes closer to that from solving the dualized model of (5) with linear decision rules.

5. Distribution on a network without commitments

Consider the linear variant of Problem (1) considered in §3, that is, distribution on a network *without* commitments, which can be written as the following two-stage robust linear optimization problem:

$$\inf_{x \in \mathcal{X}} \sup_{\zeta \in \mathcal{U}} \inf_{y \geq 0} \left\{ \sum_{i=1}^{N} c_i x_i + \sum_{i,j=1}^{N} t_{ij} y_{ij} \;\middle|\; \sum_{j=1}^{N} y_{ji} - \sum_{j=1}^{N} y_{ij} \geq \zeta_i - x_i \quad i = 1, ..., N \right\}. \tag{6}$$

Table 2 Equilibrium of system with $N-1$ piecewise-linear springs for $N \in \{15, 20, 30, 45, 65, 100\}$. Objective values for the primal static decision rule solution, the linear decision rule solution and the sampled lower bound model using primal scenarios are depicted. The time refers to the computing time for the linear decision rule solution.

N	15	20	30	45	65	100
Static decision rule	535.8	344.5	254.1	213.7	189.1	180.7
Linear decision rule	507.6	320.2	239.0	202.8	183.5	180.7
Lower bound	507.6	320.2	239.0	202.8	183.5	180.7
Time(s)	0.05	0.06	0.13	0.42	1.50	4.97

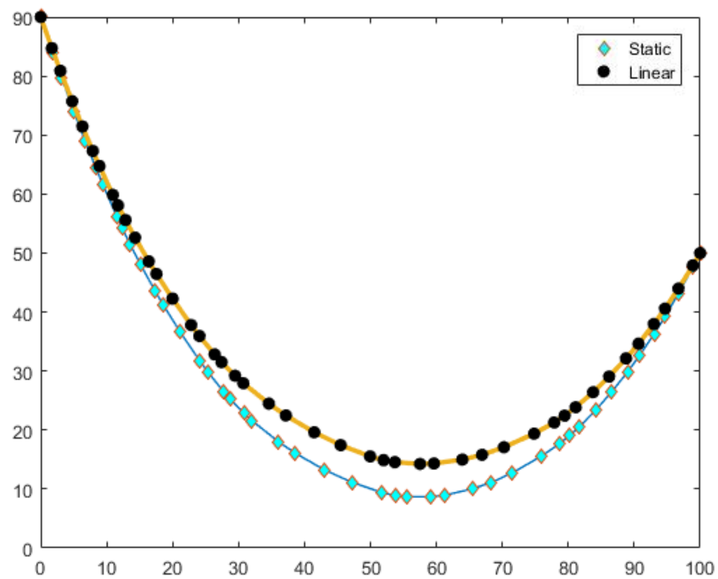

Figure 3 System of nodes (weights) connected by 44 springs for $N = 45$. The diamonds and dots represent the robust locations of the nodes from solving (5) and its dualized formulation with static decision rules and linear decision rules, respectively.

We choose $N \in \{10, 20, 30, 40, 50\}$ locations uniformly at random from $[0, 10]^2$. Let t_{ij}, the cost to transport one unit of demand from location i to j, be the Euclidean distance. The unit storage cost c_i are equal to 10 for $i = 1, ..., N$ stores. The individual maximum demand $\hat{\zeta}$ and the capacity K_i, $i = 1, ..., N$, of each location is set to 20 units. The total demand in the network is set to $20\sqrt{N}$. Note that here we consider the exact same problem setting as in Bertsimas and de Ruiter (2016), and compare the numerical performance of the lower bounding scheme proposed in §4 with the primal-dual lower bounding scheme proposed in Bertsimas and de Ruiter (2016). Table 3 reports the numerical results. One can observe that the optimality gaps obtained from our method almost halve the ones obtained from using the technique of Bertsimas and de Ruiter (2016), where the

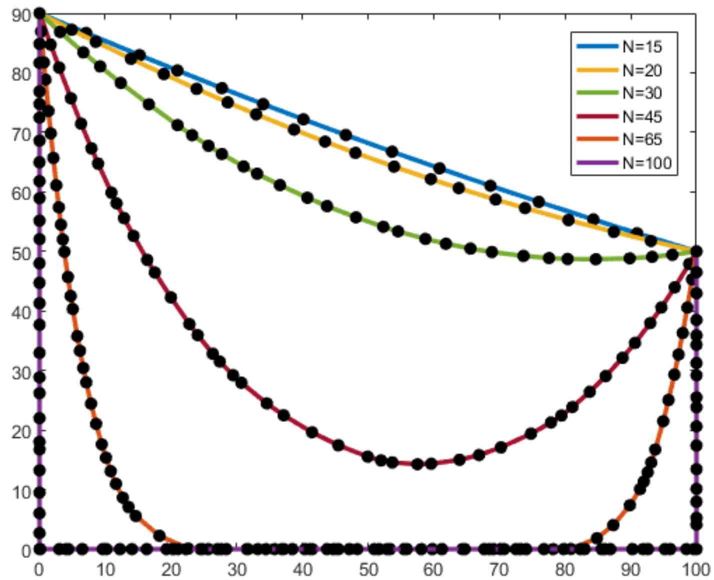

Figure 4 System of nodes (weights) connected by $N-1$ springs for $N \in \{15, 20, 30, 45, 65, 100\}$. The dots represent the robust location of the nodes.

optimality gap is computed via:

$$p\% = \frac{v(\text{DL}) - v(\text{LB})}{v(\text{DL})} \times 100\%,$$

where $v(\cdot)$ denotes the optimal value of the corresponding problems, *e.g.*, $v(\text{DL})$ is the optimal value obtained from solving the dualized model of (6) with linear decision rules, while LB$\in$\{LB-P, LB-BR\}.

Table 3 Lot-sizing problem with $N \in \{10, 20, 30, 40, 50\}$. LB-P and LB-BR denotes the approximated optimality gap using the primal scenarios (see §4) and using the primal and dual scenarios in Bertsimas and de Ruiter (2016). All the numbers are the average of 10 randomly generated instances.

N	10	20	30	40	50
LB-P	6.0%	5.7%	6.0%	5.0%	5.2%
LB-BR	13.9%	12.9%	10.4%	11.2%	10.7%

References

D. Bertsimas and F. de Ruiter. Duality in two-stage adaptive linear optimization: faster computation and stronger bounds. *INFORMS Journal on Computing*, 28(3):500–511, 2016.

M. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra and its Applications*, 284(1-3):193–228, 1998.

MOSEK ApS. *The MOSEK optimization toolbox for MATLAB manual. Version 8.1*, 2017. URL `http://docs.mosek.com/8.1/toolbox.pdf`.

K. Roos, M. Balvert, B. Gorissen, and D. den Hertog. A universal and structured way to derive dual optimization problem formulations. *INFORMS Journal on Optimization*, 2(4):229–255, 2020.