New Algorithm for the Flexibility Index Problem of Quadratic Systems

Hao Jiang[a], Bingzhen Chen[a*], Ignacio E. Grossmann[b*]

[a] Institute of Process Systems Engineering, Dept. of Chemical Engineering,

Tsinghua University, Beijing 100084, P.R. China

[b] Center for Advanced Process Decision-making, Dept. of Chemical Engineering,

Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.

Abstract: A new flexibility index algorithm for systems under uncertainty and represented by quadratic inequalities is presented. Inspired by the outer-approximation algorithm for convex mixed-integer nonlinear programming, a similar iterative strategy is developed. The subproblem, which is a nonlinear program, is constructed by fixing the vertex directions since this class of systems is proved to have a vertex solution if the entries on the diagonal of the Hessian matrix are non-negative. By overestimating the nonlinear constraints, a linear min-max problem is formulated. By dualizing the inner maximization problem, and introducing new variables and constraints, the master problem is reformulated as a mixed-integer linear program. By iteratively solving the subproblem and master problem, the algorithm can be guaranteed to converge to the flexibility index. Numerical examples including a heat exchanger network, a process network, and a unit commitment problem are presented to illustrate the computational efficiency of the algorithm.

Topical heading: Process Systems Engineering

Keywords: flexibility index, quadratic systems, overestimation

[*] Correspondence concerning this article should be addressed to I. E. Grossmann at grossmann@cmu.edu and B. Chen at dcecbz@tsinghua.edu.cn

Introduction

Since Grossmann and Sargent addressed the optimum design with uncertain parameters in 1978,[1] the area of flexibility analysis has evolved for about 40 years in the community of process systems engineering (PSE). For a historical perspective on the evolution of this area and the relations between flexibility, resiliency and robust optimization, please refer to the review paper by Grossmann et al.,[2] and the paper by Zhang et al.[3] as a tribute to professor Roger Sargent, the pioneer in the PSE community.

The basic idea of flexibility analysis is to explicitly consider uncertainties in the evaluation and design of chemical processes. The uncertainties can be thermodynamic and kinetic parameters, concentration and temperature of inlet streams, demand and price of products, etc. If the design of a chemical process is fixed, the aim of flexibility analysis to determine whether the design can tolerate the uncertainties in a specified range, or to quantify the range of uncertainties that the design can tolerate. The former is known as flexibility test problem, and the latter is known as flexibility index problem.[4] Another related topic of flexibility analysis is to design a chemical process under uncertainty with flexibility constraints.[5]

The feasibility of a chemical process can be described by a set of inequalities:[4]

$$f_j(d,z,\theta) \le 0, \quad \forall j \in J \tag{1}$$

where $d \in \mathrm{R}^{n_d}$ are design variables, $z \in \mathrm{R}^{n_z}$ are control variables, and $\theta \in \mathrm{R}^{n_\theta}$ are uncertain parameters, $J = \{1,2,...,m\}$ is the set of inequalities.

For a fixed design d (this is always the case in this paper unless noted) and every realization of uncertain parameters θ in a region R, if a set of control variables can be found such that Eq.

(1) is satisfied, then this process is feasible in *R*, which is the feasible region projected in the θ-space. Let us define the following feasibility function:[4,6]

$$\psi(d,\theta) = \min_{z} \max_{j\in J} f_j(d,z,\theta) \tag{2}$$

which is equivalent to,

$$\begin{aligned} &\psi(d,\theta) = \min_{u,z} u \\ &\text{s.t. } f_j(d,z,\theta) \le u, \quad \forall j \in J \end{aligned} \tag{3}$$

Hence, the feasible region *R* can be expressed as,

$$R = \{\theta \mid \psi(d,\theta) \le 0\} \tag{4}$$

The feasible region *R* is difficult to obtain for high-dimensional systems, and the shape is quite different for different processes. Therefore, a hyperrectangle inscribed in *R* is introduced to simplify the evaluation and comparison of tolerance for uncertain parameters. The hyperrectangle is centered at the nominal point θ^N, and the lengths of its sides are proportional to the expected positive and negative deviations, denoted by $\Delta\theta^+$ and $\Delta\theta^-$, respectively. A non-negative scalar δ is used to define the hyperrectangle $T(\delta)$:

$$T(\delta) = \left\{\theta \mid \theta^N - \delta\Delta\theta^- \le \theta \le \theta^N + \delta\Delta\theta^+\right\} \tag{5}$$

The flexibility index *F* can be defined as the maximum value of δ such that $T(\delta) \subseteq R$, which means Eq. (1) is satisfied for every realization of θ in $T(\delta)$ with appropriate adjustment of *z*. This abstract definition can be converted to an equivalent mathematical programming problem:[4,6]

$$\begin{aligned} &F(d) = \max_{\delta\ge 0} \delta \\ &\text{s.t.} \quad \max_{\theta\in T(\delta)} \min_{z} \max_{j\in J} f_j(d,z,\theta) \le 0 \\ &T(\delta) = \left\{\theta \mid \theta^N - \delta\Delta\theta^- \le \theta \le \theta^N + \delta\Delta\theta^+\right\} \end{aligned} \tag{6}$$

Similar to the aforementioned flexibility index problem, the flexibility test problem is to determine whether by adjusting the control variables *z*, Eq. (1) holds for all $\theta \in T(1) = \{\theta \mid \theta^N - \Delta\theta^- \le \theta \le \theta^N + \Delta\theta^+\}$. This paper does not consider the flexibility test problem explicitly since it is equivalent to determine whether the flexibility index is greater than or equal to one, although one may take advantage of its simpler formulation to devise a more efficient algorithm.

The geometric interpretation of the flexibility index is to find the largest hyperrectangle inscribed in the feasible region *R*. According to the Properties 1 and 2 proved by Swaney and Grossmann,[4] if $f_j(d,z,\theta), \forall j \in J$, are continuous in *z* and θ, then the critical point (the solution θ^* of problem (6)) lies at the boundary of *R*. If the critical point corresponds to a vertex of the hyperrectangle, then problem (6) can be solved by direct search or implicit enumeration algorithms.[7] If there is a nonvertex critical point, there are also several algorithms. Grossmann and Floudas[8] proposed an active constraint strategy (active-set) in which the problem is reformulated as a mixed-integer linear programming (MILP) or mixed-integer nonlinear programming (MINLP) by applying the Karush-Kuhn-Tucker (KKT) conditions to problem (3). Ostrovsky et al.[9,10] used a branch and bound method to obtain nonvertex solutions for the flexibility function. Bansal et al.[11] solved the flexibility analysis problem of linear systems by parametric programming, and then generalized it to nonlinear systems.[12] Floudas et al.[13] proposed a global optimization algorithm to solve the flexibility test and flexibility index problems of nonconvex systems, in which the problem is convexified and solved to global optimality by the αBB algorithm to obtain a lower bound, and the original problem is solved

by the active-set method to obtain an upper bound. By branching on the space of θ and z, the gap is closed within a given tolerance.

As stated by Ierapetritou and coworkers,[14,15] the flexibility index has limitations to quantify the degree of uncertain parameters a process can tolerate. Sometimes, it is overly conservative since it underestimates the size of the feasible region. However, compared to other flexibility measures, e.g., the feasible convex hull ratio[14] and the simplicial approximation,[15] the flexibility index provides a good balance between the accuracy of approximation of the feasible region and ease of interpretation. Therefore, in this paper we use the traditional flexibility index to quantify the ability of a process to tolerate uncertain parameters.

In the area of optimum design under uncertainty with flexibility constraints, Grossmann and Halemane[16] proposed a projection-restriction strategy to exploit its mathematical structure. Pistikopoulos and Grossmann considered the optimal retrofit design problem to increase flexibility for linear systems[17] and nonlinear systems.[18,19] Pistikopoulos and Ierapetritou[20] formulated the convex process design problem as a two-stage stochastic model, and proposed a decomposition-based algorithm. Varvarezos et al.[21] presented a sensitivity based approach for the evaluation and design of flexible linear processes. These analysis and design methods have been widely used in synthesis of heat exchanger networks (HEN),[22,23] multiproduct plants,[24] and cryogenic air separation process.[25]

In this work, we consider a special class of systems, where all the inequalities are quadratic or linear in θ, and linear in z. Quadratic terms of uncertain parameters are common in chemical processes, such as the product of inlet temperature and heat capacity flowrate in HEN, the

product of flowrate and contaminant concentration in water networks, and the product of demand and price of product in planning problems. Although general algorithms can solve the flexibility index problem of these systems, one cannot expect they will take advantage of the special structure to accelerate the computation. Therefore, a more efficient iterative procedure, which is quite different from the existing algorithms, is proposed in this paper to solve the flexibility index problem for this class of systems.

The proposed algorithm in this paper is inspired by the work of Zhang et al,[3] in which a duality-based method is proposed to solve the flexibility analysis problem of linear systems that is much more efficient than the active-set method.[8] In this work, we extend this method to nonlinear systems, which are linearized, yielding a lower bound problem which can be efficiently solved by the duality-based method. An upper bound is easily obtained since the critical point of this special class of systems lies at a vertex as proved in the following section. By iteratively solving the upper bound and lower bound problems, the gap is closed without the need of branching. Although the proposed method is for a limited class of nonlinear systems, the efficient computational performance on several examples including HEN, process networks and unit commitment, shows that it is a promising method.

The remainder of this article is organized as follows. In the next section, the problem is stated mathematically and the property of vertex solution is proved. The new flexibility index algorithm is developed in the subsequent section. In section 4, we apply this new algorithm to four numerical examples, and compare it with the active-set method. Finally, the conclusions are presented in section 5.

Problem statement

The inequalities of this class of systems can be written as:

$$f_j = \theta^T Q_j \theta + a_j^T \theta + b_j^T z + c_j \le 0, \quad \forall j \in J \tag{7}$$

where the Hessian matrix Q_j is an $n_\theta \times n_\theta$ matrix, a_j and b_j are column vectors, and c_j are constants. Note that the design variables *d* are not treated as variables in Eq. (7) because they are included as part of Q_j, a_j, b_j, c_j if the design is fixed. If all the elements of Q_j are zero, then Eq. (7) becomes a linear system. If f_j, $\forall j \in J$, are convex, i.e., Q_j are positive semi-definite, the flexibility index problem of this system has a vertex solution. Below we derive a weaker condition, under which the vertex solution can also be guaranteed as shown in the following proposition.

Proposition 1. Let *I* denote the set of uncertain parameters and q_{jii} denote the diagonal element of Q_j. If $q_{jii} \ge 0$, $\forall i \in I$, $j \in J$, then the flexibility index problem where f_j is defined by (7) will have a vertex solution.

Proof. See Appendix A.

Note that, $q_{jii} \ge 0$, $\forall i \in I$, are merely necessary conditions for Q_j to be positive semi-definite. Therefore, it is not as restrictive as the condition that f_j is convex, and furthermore it may in fact correspond to a nonconvex function. The following parts will restrict the diagonal elements of Q_j, $j \in J$, to be non-negative to guarantee a vertex solution.

New algorithm for the flexibility index problem

Basic idea

To better understand the basic idea of this algorithm, the duality-based flexibility index

algorithm for linear systems is introduced first.

Consider the following linear inequality constraints:

$$f = A\theta + Bz + C \le 0 \tag{8}$$

for which the flexibility index problem has a vertex solution. Therefore, the flexibility index problem can be stated in an alternative way:

$$\begin{aligned} & F = \min_{x} \max_{z,\delta \ge 0} \delta \\ & \text{s.t. } A\theta + Bz + C \le 0 \\ & \quad \theta_i = \theta_i^N + \delta\left[x_i \Delta\theta_i^+ - (1 - x_i)\Delta\theta_i^- \right], \quad \forall i \in I \\ & \quad x \in \{0,1\}^{n_\theta} \end{aligned} \tag{9}$$

The interpretation of this formulation is that, for every vertex direction, i.e., every possible combination of *x*, the distance from the nominal point to the boundary of the feasible region is maximized. Among these distances, the shortest is chosen to make the hyperrectangle be totally inscribed in the feasible region.[26]

For a given vertex direction *x*, and by substituting θ_i into Eq. (8), the inner maximization problem can then be replaced by its dual, which is a minimization problem and can be merged with the outer minimization problem. The obtained single-level problem is shown as follows:

$$\begin{aligned} & F = \min_{x,\lambda} \lambda^T(-A\theta^N - C) \\ & \text{s.t. } B^T\lambda = 0 \\ & \quad \sum_{j\in J}\sum_{i\in I}\left[x_i \Delta\theta_i^+ - (1 - x_i)\Delta\theta_i^- \right] a_{ji}\lambda_j \ge 1 \\ & \quad \lambda \ge 0 \\ & \quad x \in \{0,1\}^{n_\theta} \end{aligned} \tag{10}$$

Problem (10) is an MINLP since there are bilinear terms $x_i\lambda_j$ in the constraint. However, the bilinearity of binary and continuous variables can be eliminated by replacing $x_i\lambda_j$ by a new non-negative continuous variable $\overline{\lambda}_{ij}$ together with the following constraints:[27]

$$\begin{aligned}
&\bar{\lambda}_{ij} \geq \lambda_j - M_\lambda + M_\lambda x_i \\
&\bar{\lambda}_{ij} \leq \lambda_j \\
&\bar{\lambda}_{ij} \leq M_\lambda x_i
\end{aligned} \tag{11}$$

where M_λ is a big-M parameter. If $x_i = 0$, then $\bar{\lambda}_{ij} = 0$; if $x_i = 1$, then $\bar{\lambda}_{ij} = \lambda_j$, which is exactly equivalent to the result of $x_i \lambda_j$. Therefore, problem (10) is equivalent to the following MILP:

$$\begin{aligned}
&F = \min_{x,\lambda,\bar{\lambda}} \lambda^T (-A\theta^N - C) \\
&\text{s.t. } B^T \lambda = 0 \\
&\quad \sum_{j\in J}\sum_{i\in I} a_{ji}\left[-\Delta\theta_i^- \lambda_j + \left(\Delta\theta_i^+ + \Delta\theta_i^-\right)\bar{\lambda}_{ij}\right] \geq 1 \\
&\quad \bar{\lambda}_{ij} \geq \lambda_j - M_\lambda + M_\lambda x_i, \ \forall i \in I, \ j \in J \\
&\quad \bar{\lambda}_{ij} \leq \lambda_j, \qquad \forall i \in I, \ j \in J \\
&\quad \bar{\lambda}_{ij} \leq M_\lambda x_i, \quad \forall i \in I, \ j \in J \\
&\quad \lambda \geq 0, \bar{\lambda} \geq 0 \\
&\quad x \in \{0,1\}^{n_\theta}
\end{aligned} \tag{12}$$

However, this reformulation is not applicable to the flexibility index problem of quadratic systems as shown in the following since the inner maximization problem is nonlinear in δ.

$$\begin{aligned}
&F = \min_{x} \max_{\delta\geq 0, z} \delta \\
&\text{s.t. } \theta^T Q_j \theta + a_j^T \theta + b_j^T z + c_j \leq 0, \quad \forall j \in J \\
&\quad \theta_i = \theta_i^N + \delta\left[x_i \Delta\theta_i^+ - (1 - x_i)\Delta\theta_i^-\right], \quad \forall i \in I \\
&\quad x \in \{0,1\}^{n_\theta}
\end{aligned} \tag{13}$$

We are inspired by the outer-approximation (OA) algorithm for convex MINLP,[28] in which the nonlinear constraints are replaced by their supporting hyperplanes to construct a master MILP to provide a lower bound, and subsequently, the integer variables are fixed to the optimal solution of the master MILP to construct a nonlinear programming (NLP) subproblem to provide an upper bound. By iteratively solving the master problem and the subproblem, the algorithm will converge to the optimal solution. Similarly, we can develop the subproblem and

master problem for this nonlinear min-max problem, however, the NLP subproblem may have multiple local optima.

If all the binary variables *x* are fixed, meaning fixed vertex direction, problem (13) is an NLP, and provides an upper bound of the flexibility index, similar to the subproblem in the OA algorithm. However, the master problem is quite different from that used in the OA algorithm in two aspects: (a) the constraints are nonconvex in the min-max problem, OA will cut off some of the feasible region; (b) for the case where the constraints are convex, OA will yield an upper bound of the flexibility index instead of a lower bound since the inner problem is a maximization problem. On the contrary, if we use an overestimation of the function to replace OA, i.e., underestimation of the function, the master problem will yield a lower bound of the flexibility index.

Here, we use a nonlinear convex function to compare OA and overestimation. Consider the following constraint:

$$f = \theta_1^2 - 2\theta_1 - \theta_2 + 2 \le 0 \tag{14}$$

The nominal point is $\theta^N = (1.5, 5)$, and the expected deviations are $\Delta\theta_1^{\pm} = 1$, $\Delta\theta_2^{\pm} = 2$, respectively.

Firstly, we apply OA at two points, $\theta^1 = (0, 2)$ and $\theta^2 = (3, 5)$. Equation (14) is therefore replaced by two linear inequalities, which are the tangents of *f* at the two points, as seen in Figure 1(a).

$$\begin{aligned} f_1 &= -2\theta_1 - \theta_2 + 2 \le 0 \\ f_2 &= 4\theta_1 - \theta_2 - 7 \le 0 \end{aligned} \tag{15}$$

For comparison, the range in which we overestimate the constraint is $0 \le \theta_1 \le 3$, $2 \le \theta_2 \le 8$.

Therefore, the constraint is overestimated by the following inequality, in which the nonlinear term is replaced by its secant, as seen in Figure 1(b).

$$f_3 = \theta_1 - \theta_2 + 2 \le 0 \tag{16}$$

The corresponding feasible regions are shaded in Figures 1(a) and 1(b), where the solid and dotted rectangles are the flexible regions for the original and linearized systems, respectively. The feasible region of the second case is derived from Eq. (16) together with the bounds of the uncertain parameters. It is clear that after applying OA to the constraint, the feasible region is enlarged. Therefore, we will obtain an upper bound of the flexibility index if we solve the flexibility index problem for the linearized system. On the contrary, if the constraint is overestimated, the feasible region is totally inscribed in the original one. Therefore, as expected, we obtain a lower bound of the flexibility index.

Details about the subproblem, master problem, properties of the master problem, and the algorithm are given in the following four subsections.

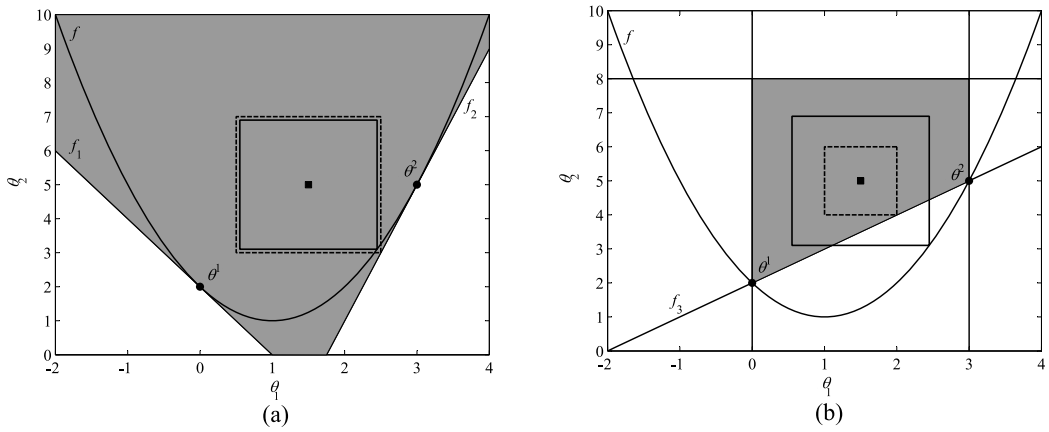

Figure 1. Feasible regions of the linearized systems. (a) OA and (b) overestimation.

Subproblem

The subproblem, which will provide an upper bound, is the inner maximization problem with

fixed *x*, as shown below. In difference to the subproblem in the OA algorithm, this subproblem is always feasible as long as the nominal point is feasible, and this is always the case.

$$\begin{aligned} &\delta^*(x) = \max_{\delta \geq 0, z} \delta \\ &\text{s.t. } f_j = \sum_i q_{jii}\theta_i^2 + \sum_i \sum_{k=i+1} \left(q_{jik} + q_{jki}\right)\theta_i\theta_k + a_j^T\theta + b_j^T z + c_j \leq 0, \quad \forall j \in J \\ &\theta_i = \theta_i^N + \delta\left[x_i\Delta\theta_i^+ - \left(1-x_i\right)\Delta\theta_i^-\right], \quad \forall i \in I \end{aligned} \tag{S}$$

As required by the assumption of the alternative formulation of the flexibility index, and given that problem (S) is nonconvex, if (S) has multiple local solutions, the smallest one is chosen to guarantee the feasibility of the entire path from the nominal point to the boundary. For some systems, it can be difficult to solve because the constraints are nonconvex in δ, and we need to guarantee the smallest local maximum to be obtained. But for this quadratic system, although nonconvex, we can add some extra constraints to make the solution of (S) to be unique and satisfy the requirement. The following derives the extra constraints and demonstrates how to obtain the smallest maximum with a local NLP solver.

By substituting θ_i into f_j, f_j becomes a function of *x*, δ and *z*, and hence can be denoted by $f_j\left(x,\delta,z\right)$. For fixed *x*, f_j is quadratic and/or linear in δ and linear in *z*. Hence, (S) can be rewritten in compact form as:

$$\begin{aligned} &\delta^*(x) = \max_{\delta \geq 0, z} \delta \\ &\text{s.t. } f_j = q_j'\delta^2 + p_j\delta + b_j^T z + c_j' \leq 0, \quad \forall j \in J \end{aligned} \tag{17}$$

where the coefficients q_j', p_j, and c_j' are functions of *x*:

$$
\begin{aligned}
q'_j(x) &= \sum_i q_{jii}\left[x_i\Delta\theta_i^+ - (1-x_i)\Delta\theta_i^-\right]^2 \\
&+ \sum_i \sum_{k=i+1} \left(q_{jik}+q_{jki}\right)\left[x_i\Delta\theta_i^+ - (1-x_i)\Delta\theta_i^-\right]\left[x_k\Delta\theta_k^+ - (1-x_k)\Delta\theta_k^-\right] \\
p_j(x) &= \sum_i 2q_{jii}\theta_i^N\left[x_i\Delta\theta_i^+ - (1-x_i)\Delta\theta_i^-\right] \\
&+ \sum_i \sum_{k=i+1} \left(q_{jik}+q_{jki}\right)\left\{\theta_k^N\left[x_i\Delta\theta_i^+ - (1-x_i)\Delta\theta_i^-\right] + \theta_i^N\left[x_k\Delta\theta_k^+ - (1-x_k)\Delta\theta_k^-\right]\right\} \\
&+ \sum_i a_{ji}\left[x_i\Delta\theta_i^+ - (1-x_i)\Delta\theta_i^-\right] \\
c'_j(x) &= \sum_i q_{jii}\left(\theta_i^N\right)^2 + \sum_i \sum_{k=i+1} \left(q_{jik}+q_{jki}\right)\theta_i^N\theta_k^N + \sum_i a_{ji}\theta_i^N + c_j
\end{aligned}
\tag{18}
$$

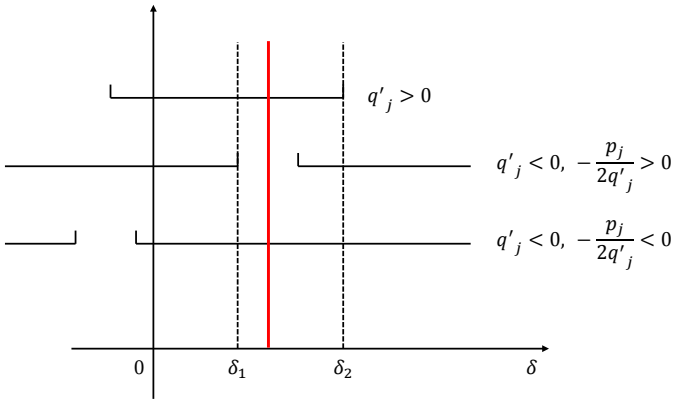

Figure 2. The range of δ for different constraints

The range of δ for every inequality in (17) can be derived analytically since all the inequalities are quadratic. Assume δ_j^L and δ_j^U are roots of $f_j = 0$ and $\delta_j^L \le \delta_j^U$, the range of δ for the *j*th inequality is $\delta_j^L \le \delta \le \delta_j^U$ if $q'_j > 0$, and $\delta \le \delta_j^L$ or $\delta \ge \delta_j^U$ if $q'_j < 0$. These ranges can be divided into three typical cases as shown in Figure 2. The horizontal lines denote the feasible regions of the three cases, the dash lines denote the two local maxima, and the red thick line denotes the extra constraint to exclude the undesired maxima. The first case corresponds to constraints with positive q'_j which are convex, and the feasible region is a continuous interval. The second case corresponds to constraints with negative q'_j and right-hand half plane symmetry axis, then the feasible region is two disjunctive half open intervals

and will lead to multiple local optima. The last case is similar to the second one but with left-hand half plane symmetry axis, and will not lead to multiple local optima. The linear constraints ($q'_j = 0$) are neglected in the figure. It is clear from Figure 2 that there are two maxima of δ, among which δ_1 is a local maximum and δ_2 is the global maximum. As required by the assumption of the alternative formulation of the flexibility index problem, we need to guarantee that the solution of (17) is the smallest maximum δ_1. The only reason that problem (17) has multiple local maxima lies in the constraints with negative q'_j and right-hand half plane symmetry axis. Therefore, for those constraints we can add some extra constraints as shown below to make sure δ does not cross the intermediate infeasible region. The vertical red thick line in Figure 2 gives a graphical illustration.

$$\delta \le -\frac{p_j}{2q'_j}, \quad \forall j \in J,\ q'_j < 0, -\frac{p_j}{2q'_j} > 0 \tag{19}$$

Note that if there exists some *j* such that constraint (19) is active after solving the problem, then the active constraint should be removed and the problem needs to be re-solved. This is because if constraint (19) is active, the corresponding f_j is always less than zero and the intermediate infeasible interval vanishes. Therefore, f_j will not lead to multiple optima.

Master problem

As mentioned in the basic idea section, the nonlinear constraints need to be overestimated to derive the master problem. Therefore, we need to derive a linear or piecewise linear function f_j^l (the superscript l means linear), such that $f_j^l \ge f_j$ in the hyperrectangle $T(F^U) = \{\theta \mid \theta^L \le \theta \le \theta^U\}$, where F^U denotes the upper bound of the flexibility index *F*, and θ^L, θ^U are given by

$$\begin{aligned}\theta^L &= \theta^N - F^U \Delta\theta^- \\ \theta^U &= \theta^N + F^U \Delta\theta^+\end{aligned} \tag{20}$$

In this hyperrectangle, the convex quadratic term in (13) can be overestimated by its secant:

$$\theta_i^2 \le \left(\theta_i^L\right)^2 + \frac{\left(\theta_i^U\right)^2 - \left(\theta_i^L\right)^2}{\theta_i^U - \theta_i^L}\left(\theta_i - \theta_i^L\right) = \left(\theta_i^L + \theta_i^U\right)\theta_i - \theta_i^L\theta_i^U \tag{21}$$

The overestimation of the bilinear term is based on the McCormick relaxation as shown below:[29,30]

$$\theta_i\theta_k \le \min(\theta_k^L\theta_i + \theta_i^U\theta_k - \theta_i^U\theta_k^L,\ \ \theta_k^U\theta_i + \theta_i^L\theta_k - \theta_i^L\theta_k^U) \tag{22a}$$

$$\theta_i\theta_k \ge \max(\theta_k^L\theta_i + \theta_i^L\theta_k - \theta_i^L\theta_k^L,\ \ \theta_k^U\theta_i + \theta_i^U\theta_k - \theta_i^U\theta_k^U) \tag{22b}$$

By keeping Eq. (22a) unchanged and multiplying -1 to Eq. (22b), we get,

$$\theta_i\theta_k \le w_{ik}^p = \min(\theta_k^L\theta_i + \theta_i^U\theta_k - \theta_i^U\theta_k^L,\ \ \theta_k^U\theta_i + \theta_i^L\theta_k - \theta_i^L\theta_k^U) \tag{23a}$$

$$-\theta_i\theta_k \le w_{ik}^n = \min(-\theta_k^L\theta_i - \theta_i^L\theta_k + \theta_i^L\theta_k^L,\ \ -\theta_k^U\theta_i - \theta_i^U\theta_k + \theta_i^U\theta_k^U) \tag{23b}$$

Therefore, w_{ik}^p and w_{ik}^n are the overestimations of $\theta_i\theta_k$ and $-\theta_i\theta_k$, respectively. By replacing the positive and negative bilinear terms in (13) together with their signs by w_{ik}^p and w_{ik}^n, respectively, the master problem for (13) with piecewise linear constraints is shown in (M1). The upper bound of δ is explicitly added to make the problem tighter, although this constraint is not essential.

$$
\begin{aligned}
F^L = & \min_{x} \max_{\delta \geq 0, z, w^p, w^n} \delta \\
\text{s.t. } & f_j^l = \sum_i \left[q_{jii} \left(\theta_i^L + \theta_i^U \right) + a_{ji} \right] \theta_i + \sum_{(i,k) \in I^p} \left(q_{jik} + q_{jki} \right) w_{ik}^p \\
& \quad - \sum_{(i,k) \in I^n} \left(q_{jik} + q_{jki} \right) w_{ik}^n + b_j^T z - \sum_i q_{jii} \theta_i^L \theta_i^U + c_j \leq 0, \qquad \forall j \in J \\
& w_{ik}^p = \min\left(\theta_k^U \theta_i + \theta_i^L \theta_k - \theta_i^L \theta_k^U,\ \theta_k^L \theta_i + \theta_i^U \theta_k - \theta_i^U \theta_k^L \right), \qquad \forall (i,k) \in I^p \\
& w_{ik}^n = \min\left(-\theta_k^L \theta_i - \theta_i^L \theta_k + \theta_i^L \theta_k^L,\ -\theta_k^U \theta_i - \theta_i^U \theta_k + \theta_i^U \theta_k^U \right), \quad \forall (i,k) \in I^n \\
& \theta_i = \theta_i^N + \delta \left[x_i \Delta \theta_i^+ - \left(1 - x_i \right) \Delta \theta_i^- \right], \qquad \forall i \in I \\
& \delta \leq F^U \\
& x \in \{0,1\}^{n_\theta}
\end{aligned}
\tag{M1}
$$

The sets I^p and I^n are defined by $I^p = \{(i,k) \mid i,k \in I, i < k, \exists j \in J, q_{jik} + q_{jki} > 0\}$ and $I^n = \{(i,k) \mid i,k \in I, i < k, \exists j \in J, q_{jik} + q_{jki} < 0\}$, respectively. By substituting w_{ik}^p and w_{ik}^n into f_j^l, f_j^l becomes a function of θ and z. Since θ is a function of x and δ, f_j^l can also be a function of x, δ and z. Note that this is different from the traditional use of McCormick relaxation, in which every bilinear term is replaced by w, and w is bounded by the following inequalities:

$$w_{ik} \leq \theta_k^U \theta_i + \theta_i^L \theta_k - \theta_i^L \theta_k^U, \quad \forall i,k \in I, i < k \tag{24a}$$

$$w_{ik} \leq \theta_k^L \theta_i + \theta_i^U \theta_k - \theta_i^U \theta_k^L, \quad \forall i,k \in I, i < k \tag{24b}$$

$$w_{ik} \geq \theta_k^L \theta_i + \theta_i^L \theta_k - \theta_i^L \theta_k^L, \quad \forall i,k \in I, i < k \tag{24c}$$

$$w_{ik} \geq \theta_k^U \theta_i + \theta_i^U \theta_k - \theta_i^U \theta_k^U, \quad \forall i,k \in I, i < k \tag{24d}$$

The feasible region is enlarged after using the traditional McCormick relaxation. Therefore, it will provide an upper bound of the flexibility index instead of a lower bound.

The inner maximization problem of (M1) includes min operators of two linear functions. A conventional technique is to introduce binary variables, and then use disjunctions to represent the choice of the two linear functions. However, this will lead to a mixed-integer bilevel liner

programming with binary variables in both levels, which is much harder to solve than bilevel linear programming with binary variables only in the outer level.

Actually, the result of the min operators in the inner problem is determined by the outer level variables *x*. To simplify the presentation, we consider the case of $\Delta\theta_i^+ = \Delta\theta_i^-$ and $\Delta\theta_k^+ = \Delta\theta_k^-$ (see Appendix B for other cases). The results of the min operators are given in Propositions 2 and 3.

Proposition 2. For the bilinear terms in Eq. (23a) with positive coefficients, if $x_i = 0$, then $w_{ik}^p = \theta_k^U\theta_i + \theta_i^L\theta_k - \theta_i^L\theta_k^U$; if $x_i = 1$, then $w_{ik}^p = \theta_k^L\theta_i + \theta_i^U\theta_k - \theta_i^U\theta_k^L$, regardless of x_k, $i < k$.

Proof. See appendix C.

Proposition 3. For the bilinear terms in Eq. (23b) with negative coefficients, if $x_i = 0$, then $w_{ik}^n = -\theta_k^L\theta_i - \theta_i^L\theta_k + \theta_i^L\theta_k^L$; if $x_i = 1$, then $w_{ik}^n = -\theta_k^U\theta_i - \theta_i^U\theta_k + \theta_i^U\theta_k^U$, regardless of x_k, $i < k$.

Proof. Similar to the proof of Proposition 2.

Based on Propositions 2 and 3, the min operators in (M1) can be rewritten as the following disjunctions:

$$\begin{bmatrix} \neg x_i \\ w_{ik}^p = \theta_k^U\theta_i + \theta_i^L\theta_k - \theta_i^L\theta_k^U \end{bmatrix} \veebar \begin{bmatrix} x_i \\ w_{ik}^p = \theta_k^L\theta_i + \theta_i^U\theta_k - \theta_i^U\theta_k^L \end{bmatrix}, \qquad \forall (i,k) \in I^p \tag{25}$$

$$\begin{bmatrix} \neg x_i \\ w_{ik}^n = -\theta_k^L\theta_i - \theta_i^L\theta_k + \theta_i^L\theta_k^L \end{bmatrix} \veebar \begin{bmatrix} x_i \\ w_{ik}^n = -\theta_k^U\theta_i - \theta_i^U\theta_k + \theta_i^U\theta_k^U \end{bmatrix}, \qquad \forall (i,k) \in I^n \tag{26}$$

The disjunctions in Eqs. (25) and (26) are reformulated as mixed-integer constraints with the big-M method.[31] Then, (M1) can be reformulated as the following more tractable min-max programming without the min operators in the inner level.

$$
\begin{aligned}
F^L = & \min_{x} \max_{\delta \geq 0, z, w^p, w^n} \delta \\
\text{s.t.} & \sum_i \left[q_{jii} \left(\theta_i^L + \theta_i^U \right) + a_{ji} \right] \theta_i + \sum_{(i,k) \in I^p} \left(q_{jik} + q_{jki} \right) w_{ik}^p \\
& \quad - \sum_{(i,k) \in I^n} \left(q_{jik} + q_{jki} \right) w_{ik}^n + b_j^T z - \sum_i q_{ji} \theta_i^L \theta_i^U + c_j \leq 0, \quad \forall j \in J \\
& - M x_i \leq w_{ik}^s - \left(a_{1ik}^s \theta_i + b_{1ik}^s \theta_k + d_{1ik}^s \right) \leq M x_i, \\
& \qquad \forall s \in \{p, n\}, \forall (i,k) \in I^s \\
& - M \left(1 - x_i \right) \leq w_{ik}^s - \left(a_{2ik}^s \theta_i + b_{2ik}^s \theta_k + d_{2ik}^s \right) \leq M \left(1 - x_i \right), \\
& \qquad \forall s \in \{p, n\}, \forall (i,k) \in I^s \\
& \theta_i = \theta_i^N + \delta \left[x_i \Delta \theta_i^+ - \left(1 - x_i \right) \Delta \theta_i^- \right], \quad \forall i \in I \\
& \delta \leq F^U \\
& x \in \{0,1\}^{n_\theta}
\end{aligned}
\tag{M2}
$$

where $a_{hik}^s, b_{hik}^s, d_{hik}^s, h \in \{1,2\}$, denote the corresponding coefficients and constants of the linear functions in Eqs. (25) and (26) for $s \in \{p,n\}$, while w_{ik}^s represents the aggregation of w_{ik}^p and w_{ik}^n in (M1).

Similar to the dual-based method for linear systems, we can obtain the following MINLP by replacing the inner maximization problem by its dual. The variables λ, μ, υ, and η are the multipliers associated with the inequalities in (M2).

$$
\begin{aligned}
F^L = \min_{\lambda,\mu,\nu,\eta,x}\ & \sum_j \lambda_j \left[-\sum_i \left[q_{jii}\left(\theta_i^L + \theta_i^U\right) + a_{ji} \right] \theta_i^N + \sum_i q_{jii}\theta_i^L\theta_i^U - c_j \right] \\
& + \sum_s \sum_h \sum_{(i,k)\in I^s} \mu_{hik}^s \left(-a_{hik}^s \theta_i^N - b_{hik}^s \theta_k^N - d_{hik}^s \right) \\
& - \sum_s \sum_{(i,k)\in I^s} M x_i \mu_{1ik}^s - \sum_s \sum_{(i,k)\in I^s} M \mu_{2ik}^s + \sum_s \sum_{(i,k)\in I^s} M x_i \mu_{2ik}^s \\
& + \sum_s \sum_h \sum_{(i,k)\in I^s} \nu_{hik}^s \left(a_{hik}^s \theta_i^N + b_{hik}^s \theta_k^N + d_{hik}^s \right) \\
& - \sum_s \sum_{(i,k)\in I^s} M x_i \nu_{1ik}^s - \sum_s \sum_{(i,k)\in I^s} M \nu_{2ik}^s + \sum_s \sum_{(i,k)\in I^s} M x_i \nu_{2ik}^s + F^U \eta \\
\text{s.t.}\ & \sum_j \sum_i \left[q_{jii}\left(\theta_i^L + \theta_i^U\right) + a_{ji} \right] \left[\left(\Delta\theta_i^+ + \Delta\theta_i^-\right) x_i \lambda_j - \Delta\theta_i^- \lambda_j \right] \\
& + \sum_s \sum_h \sum_{(i,k)\in I^s} a_{hik}^s \left(\Delta\theta_i^+ + \Delta\theta_i^-\right) x_i \mu_{hik}^s + b_{hik}^s \left(\Delta\theta_k^+ + \Delta\theta_k^-\right) x_k \mu_{hik}^s \\
& - \sum_s \sum_h \sum_{(i,k)\in I^s} \left(a_{hik}^s \Delta\theta_i^- + b_{hik}^s \Delta\theta_k^- \right) \mu_{hik}^s \\
& - \sum_s \sum_h \sum_{(i,k)\in I^s} a_{hik}^s \left(\Delta\theta_i^+ + \Delta\theta_i^-\right) x_i \nu_{hik}^s + b_{hik}^s \left(\Delta\theta_k^+ + \Delta\theta_k^-\right) x_k \nu_{hik}^s \\
& + \sum_s \sum_h \sum_{(i,k)\in I^s} \left(a_{hik}^s \Delta\theta_i^- + b_{hik}^s \Delta\theta_k^- \right) \nu_{hik}^s + \eta \geq 1 \\
& B^T \lambda = 0 \\
& \sum_j \left(q_{jik} + q_{jki} \right) \lambda_j - \sum_h \mu_{hik}^p + \sum_h \nu_{hik}^p = 0, \qquad \forall (i,k) \in I^p \\
& -\sum_j \left(q_{jik} + q_{jki} \right) \lambda_j - \sum_h \mu_{hik}^n + \sum_h \nu_{hik}^n = 0, \qquad \forall (i,k) \in I^n \\
& \lambda, \mu, \nu, \eta \geq 0 \\
& x \in \{0,1\}^{n_\theta}
\end{aligned}
\tag{M3}
$$

where $B^T = [b_1, b_2, ..., b_{n_j}]$.

It should be noted that although there are binary variables in the inner problem, it can be rigorously converted to its dual since the binary variables belong to the outer problem. An illustrative example is presented in Appendix D to explain this. With the technique in the basic idea section, by introducing auxiliary variables, (M3) is reformulated as the following MILP, which is the final version we actually solve.

$$
\begin{aligned}
F^L = \min_{\lambda,\mu,\nu,\eta,\bar{\lambda},\bar{\mu},\bar{\nu},x} \; & \sum_j \lambda_j \left[-\sum_i \left[q_{jii}\left(\theta_i^L + \theta_i^U\right) + a_{ji} \right] \theta_i^N + \sum_i q_{jii} \theta_i^L \theta_i^U - c_j \right] \\
& + \sum_s \sum_h \sum_{(i,k)\in I^s} \mu_{hik}^s \left(-a_{hik}^s \theta_i^N - b_{hik}^s \theta_k^N - d_{hik}^s \right) \\
& - \sum_s \sum_{(i,k)\in I^s} M \bar{\mu}_{i1ik}^s - \sum_s \sum_{(i,k)\in I^s} M \mu_{2ik}^s + \sum_s \sum_{(i,k)\in I^s} M \bar{\mu}_{i2ik}^s \\
& + \sum_s \sum_h \sum_{(i,k)\in I^s} \nu_{hik}^s \left(a_{hik}^s \theta_i^N + b_{hik}^s \theta_k^N + d_{hik}^s \right) \\
& - \sum_s \sum_{(i,k)\in I^s} M \bar{\nu}_{i1ik}^s - \sum_s \sum_{(i,k)\in I^s} M \nu_{2ik}^s + \sum_s \sum_{(i,k)\in I^s} M \bar{\nu}_{i2ik}^s + F^U \eta
\end{aligned}
$$

$$
\begin{aligned}
\text{s.t.} \; & \sum_j \sum_i \left[q_{jii}\left(\theta_i^L + \theta_i^U\right) + a_{ji} \right] \left[\left(\Delta\theta_i^+ + \Delta\theta_i^-\right) \bar{\lambda}_{ij} - \Delta\theta_i^- \lambda_j \right] \\
& + \sum_s \sum_h \sum_{(i,k)\in I^s} a_{hik}^s \left(\Delta\theta_i^+ + \Delta\theta_i^-\right) \bar{\mu}_{ihik}^s + b_{hik}^s \left(\Delta\theta_k^+ + \Delta\theta_k^-\right) \bar{\mu}_{khik}^s \\
& - \sum_s \sum_h \sum_{(i,k)\in I^s} \left(a_{hik}^s \Delta\theta_i^- + b_{hik}^s \Delta\theta_k^- \right) \mu_{hik}^s \\
& - \sum_s \sum_h \sum_{(i,k)\in I^s} a_{hik}^s \left(\Delta\theta_i^+ + \Delta\theta_i^-\right) \bar{\nu}_{ihik}^s + b_{hik}^s \left(\Delta\theta_k^+ + \Delta\theta_k^-\right) \bar{\nu}_{khik}^s \\
& + \sum_s \sum_h \sum_{(i,k)\in I^s} \left(a_{hik}^s \Delta\theta_i^- + b_{hik}^s \Delta\theta_k^- \right) \nu_{hik}^s + \eta \geq 1
\end{aligned}
$$

$$
\begin{aligned}
& B^T \lambda = 0 \\
& \sum_j \left(q_{jik} + q_{jki} \right) \lambda_j - \sum_h \mu_{hik}^p + \sum_h \nu_{hik}^p = 0, \qquad \forall (i,k) \in I^p \\
& -\sum_j \left(q_{jik} + q_{jki} \right) \lambda_j - \sum_h \mu_{hik}^n + \sum_h \nu_{hik}^n = 0, \qquad \forall (i,k) \in I^n
\end{aligned}
\tag{M4}
$$

$$
\left.
\begin{aligned}
& \bar{\lambda}_{ij} \geq \lambda_j - M_\lambda + M_\lambda x_i \\
& \bar{\lambda}_{ij} \leq \lambda_j \\
& \bar{\lambda}_{ij} \leq M_\lambda x_i
\end{aligned}
\right\} \forall i \in I, \; j \in J
$$

$$
\left.
\begin{aligned}
& \bar{\mu}_{ihik}^s \geq \mu_{hik}^s - M_\mu + M_\mu x_i \\
& \bar{\mu}_{ihik}^s \leq \mu_{hik}^s \\
& \bar{\mu}_{ihik}^s \leq M_\mu x_i \\
& \bar{\nu}_{ihik}^s \geq \nu_{hik}^s - M_\nu + M_\nu x_i \\
& \bar{\nu}_{ihik}^s \leq \nu_{hik}^s \\
& \bar{\nu}_{ihik}^s \leq M_\nu x_i \\
& \bar{\mu}_{khik}^s \geq \mu_{hik}^s - M_\mu + M_\mu x_k \\
& \bar{\mu}_{khik}^s \leq \mu_{hik}^s \\
& \bar{\mu}_{khik}^s \leq M_\mu x_k \\
& \bar{\nu}_{khik}^s \geq \nu_{hik}^s - M_\nu + M_\nu x_k \\
& \bar{\nu}_{khik}^s \leq \nu_{hik}^s \\
& \bar{\nu}_{khik}^s \leq M_\nu x_k
\end{aligned}
\right\} \forall s \in \{p,n\}, \forall h \in \{1,2\}, \forall (i,k) \in I^s
$$

$$
\lambda, \mu, \nu, \eta, \bar{\lambda}, \bar{\mu}, \bar{\nu} \geq 0
$$

$$
x \in \{0,1\}^{n_\theta}
$$

Properties of the master problem

There are four versions of master problems, from (M1) to (M4). It is interesting to investigate their relations. In most instances, all the four problems are equivalent, but there are some exceptions as listed in Table 1. The reason for this lack of equivalence is two-fold: one lies in introducing big-M parameters for an unbounded problem, and the other is dualizing an infeasible problem.

Table 1. Exceptions when reformulating the master problems

(M1) vs. (M2)	If (M1) is unbounded, i.e., the inner maximization problem of (M1) is unbounded for every x, then the result of (M1) is plus infinity. But (M2) has a finite positive optimum since M is finite.
(M2) vs. (M3)	If for some (or all) vertex directions, the inner maximization problem of (M2) is infeasible, then (M2) has a positive optimum (or infeasible). But (M3) is unbounded and the result is minus infinity since the dual of an infeasible problem is unbounded.
(M3) vs. (M4)	If (M3) is unbounded, then the result of (M3) is minus infinity. But (M4) has a finite negative optimum, since M_λ, M_μ, and M_υ are finite.

Since (M4) is the master problem we actually solve, we only need to consider the exceptions between (M1) and (M4). It can be concluded from Table 1 that,

(1) If (M1) is unbounded, then its optimum is plus infinity. But (M4) has a finite positive optimum since (M2) has a finite positive optimum.

(2) If for some (or all) vertex directions, the inner maximization problem of (M1) is

infeasible, then it has a positive optimum (or infeasible). But (M4) has a negative optimum.

For all the other scenarios except these two exceptions, (M1) and (M4) have the same result. Let x^* denote the optimal solution of (M4), and $\delta^*(x^*)$ the optimal value of the subproblem (S). The relations of F, F^L, F^U, and $\delta^*(x^*)$ are given in Propositions 4 and 5.

Let X^+ denote the set of x in which $\delta^*(x) > F^U$, $\mathrm{X}^=$ the set of x in which $\delta^*(x) = F^U$, and X^- the set of x in which $\delta^*(x) < F^U$. As shown in Appendix E, the feasible region of (M1) for a fixed design d will look like this:

(1) It is inscribed in the intersection of $T(F^U)$ and the feasible region of the original problem.

(2) If $x \in \mathrm{X}^+ \bigcup \mathrm{X}^=$, the vertex $\theta = \theta^N + F^U \left[x\Delta\theta^+ - (1-x)\Delta\theta^- \right]$ of $T(F^U)$ is on the boundary of the feasible region. If $x \in \mathrm{X}^-$, the vertex $\theta = \theta^N + \delta^*(x)\left[x\Delta\theta^+ - (1-x)\Delta\theta^- \right]$ is outside of the feasible region.

Proposition 4. If $F^U = F$, then $F^L = F = F^U$.

Proof. If $F^U = F$, then $\mathrm{X}^- = \varnothing$, and hence $\mathrm{X}^+ \bigcup \mathrm{X}^=$ contains all the vertex directions. According to the aforementioned summary, all the vertices of $T(F^U)$ are on the boundary of the feasible region of (M1). Therefore, for every x, the inner maximization problem of (M1) has the same optimum F^U, and hence the optimum of (M1) is F^U, i.e., F. Since the inner maximization problem of (M1) is bounded and feasible for every vertex, (M1) and (M4) have the same optimum. Therefore, $F^L = F = F^U$ ■

Proposition 5. If $F^U > F$, then $F^L < F \le \delta^*(x^*) < F^U$.

Proof. Since f_j^l is an overestimation of f_j in $T(F^U)$, and $f_j^l = f_j$ only occurs at the vertices/boundaries of $T(F^U)$. Therefore, in a smaller $T(F)$, $f_j^l > f_j$ is always satisfied. We can then obtain the following inequality:

$$\max_{\theta \in T(F)} \min_{z} \max_{j \in J} f_j^l > \max_{\theta \in T(F)} \min_{z} \max_{j \in J} f_j = 0 \tag{27}$$

If the nominal point is feasible in (M1), then its flexibility index must be less than *F*, and (M1), (M4) have the same optimum. On the other hand, if the nominal point is infeasible in (M1), then (M4) has a negative optimum. In either case, $F^L < F$ always holds.

Next, we only need to prove $\delta^*(x^*) < F^U$ since $\delta^*(x^*) \geq F$ always holds. Based on the conclusions in Appendix E, if $x \in \mathrm{X}^+ \bigcup \mathrm{X}^=$, $\varphi(d, x, F^U) = 0$; if $x \in \mathrm{X}^-$, $\varphi\left(d, x, \delta^*(x)\right) > 0$. If the inner maximization problem of (M1) is feasible for every $x \in \mathrm{X}^-$, the solution x^* must be in the set X^- since we need to choose the smallest δ such that $\varphi(d, x, \delta) = 0$. In this case, (M1) and (M4) have the same solution. Therefore, $\delta^*(x^*) < F^U$. On the other hand, if there are some $x \in \mathrm{X}^-$ that are infeasible in the inner maximization problem of (M1), the optimum of (M4) will be negative, and one of these *x* will be the solution of (M4). In either case, $\delta^*(x^*) < F^U$ always holds. ■

Algorithm

The proposed duality-based algorithm is shown in Figure 3, and the procedure is as follows:

Step 1: Choose an initial vertex direction x_0, such as setting all elements to zero (i.e., vertex direction for lower bounds). Choose a stopping criterion ε.

Step 2: Use x_0 to solve the subproblem (S) with the constraint in Eq. (19), and the result is $\delta^*(x_0)$. Set $F^U = \delta^*(x_0)$.

Step 3: Set $\theta^L = \theta^N - F^U \Delta\theta^-$, and $\theta^U = \theta^N + F^U \Delta\theta^+$. Use these parameters to construct and solve the master problem (M4), the optimal results are F^L and x^*.

Step 4: If $F^U - F^L \le \varepsilon$, stop; otherwise, set $x_0 = x^*$, and return to step 2.

The convergence of the proposed algorithm is proved in the following proposition.

Proposition 6. The proposed algorithm will converge to the optimal solution in a finite number of iterations regardless of the choice of x_0.

Proof. Let *F* denote the flexibility index, and x^F denote the direction that the flexibility index is obtained, i.e., $\delta^*(x^F) = F$. If $x_0 = x^F$, then in the first iteration $F^U = F$. According to Proposition 4, $F^L = F = F^U$, therefore, the algorithm converges to *F* in one iteration.

If $x_0 \neq x^F$, and let the superscript in the parentheses denote the number of iteration. In the first iteration, $F^{U(1)} = \delta^*(x_0) > F$. According to Proposition 5, $F^{L(1)} < F \le \delta^*\left(x^{*(1)}\right) < F^{U(1)}$. In the second iteration, $F^{U(2)} = \delta^*\left(x^{*(1)}\right) < F^{U(1)}$, meaning F^U is strictly decreasing. Since the number of vertices is finite, F^U will finally be equal to *F*; and the corresponding F^L will also be equal to *F*. Therefore, the algorithm will converge to *F* in a finite number of iterations regardless of x_0. ■

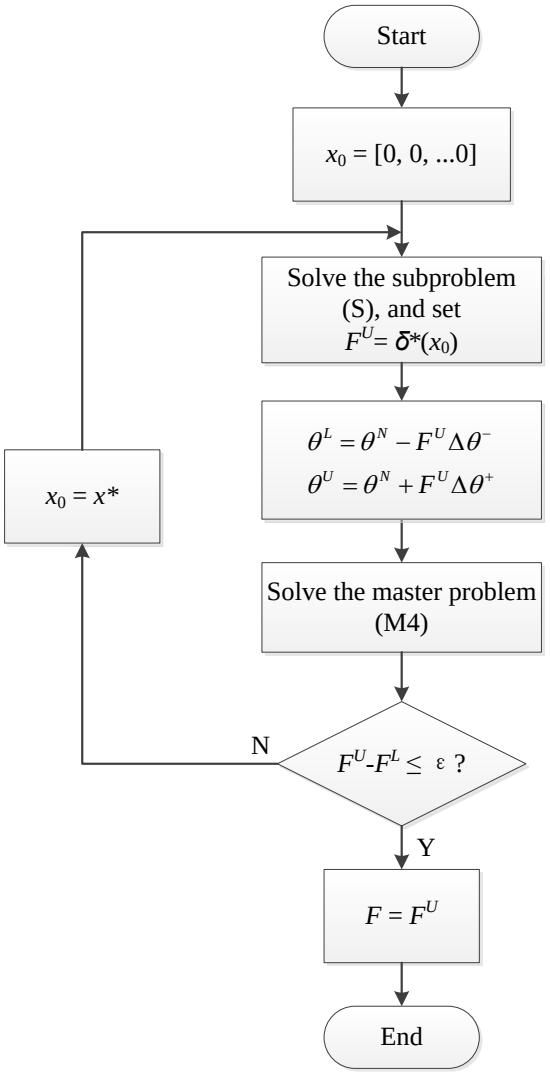

Figure 3. Flowchart of the proposed duality-based flexibility index algorithm

Numerical examples

In the following, four examples are presented to compare the traditional active-set method with the proposed duality-based method. The first one is a low dimensional problem to show the details of the proposed method, the second one is a small HEN, the last two are a process network problem, and a security constrained unit commitment problem to compare the computational performance of different methods. All models are solved in GAMS 24.8.2, the MINLP in the active-set method is solved by BARON 16.12.7,[32] and the NLP and MILP in the proposed duality-based method are solved by CONOPT 3 and CPLEX 12.7, respectively. The

computation is carried out on an Intel Core™ i5-6200U laptop at 2.30 GHz with 8GB RAM. Note that the sets and indices used in the last two models are independent, and therefore, should not be confused with each other and the ones used in the previous sections.

Low dimensional problem

This problem is designed to show some diagrams to illustrate the details of the algorithm. Assume that there are two uncertain parameters θ_1 and θ_2, no control variables. The nominal values and expected deviations are $\theta_1^N = 7$, $\theta_2^N = 8$, $\Delta\theta_1^{\pm} = 4$, $\Delta\theta_2^{\pm} = 6$, respectively. The following two constraints must be satisfied.

$$\begin{aligned} f_1 &= \theta_1\theta_2 \le 100 \\ f_2 &= -\theta_1\theta_2 \le -10 \end{aligned} \tag{28}$$

Problem (S) is solved with four different *x*, and the solutions of four vertex directions are shown in Figure 4. Note that, when solving problem (S) with x_0 = (0, 0), an extra constraint like the one in Eq. (19) should be added to guarantee the smallest optimum is obtained.

$$\delta \le \frac{\Delta\theta_1^-\theta_2^N + \Delta\theta_2^-\theta_1^N}{2\Delta\theta_1^-\Delta\theta_2^-} \tag{29}$$

The two blue curves are the boundary of the feasible region, the black dot in the center is the nominal point, and the four arrows represent the maximum distance from the nominal point to the boundary. The flexibility index $F = 0.51$ corresponds to the smallest vertex solution, and its corresponding flexible region is marked by the red solid rectangle. The dotted rectangle is the reference rectangle, i.e., the expected flexible region.

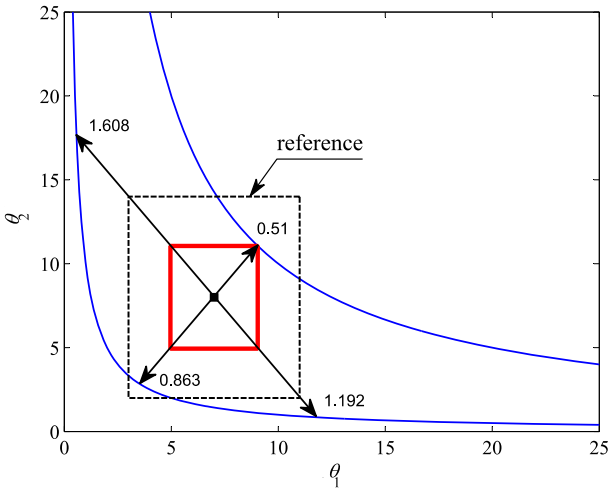

Figure 4. Vertex solutions and the feasible region

The proposed algorithm is carried out step by step to show the details. The big-M parameters used in this problem are $M = 100$, $M_\lambda = M_\mu = M_\nu = 1$, and these big-M parameters are also used in the subsequent three examples. The value of M should satisfy the following inequality:

$$M \geq \max\left\{\begin{array}{l}\left|w_{ik}^{p}-\theta_k^U\theta_i-\theta_i^L\theta_k+\theta_i^L\theta_k^U\right| \\ \left|w_{ik}^{p}-\theta_k^L\theta_i-\theta_i^U\theta_k+\theta_i^U\theta_k^L\right| \\ \left|w_{ik}^{n}+\theta_k^L\theta_i+\theta_i^L\theta_k-\theta_i^L\theta_k^L\right| \\ \left|w_{ik}^{n}+\theta_k^U\theta_i+\theta_i^U\theta_k-\theta_i^U\theta_k^U\right|\end{array}\right\} \tag{30}$$

It is tedious and unnecessary to derive M with Eq. (30), we can just set a sufficiently large number for M, say 100 for these case studies, which is sufficiently large with regard to the range of θ. The typical values for M_λ, M_μ, M_ν are 1. After solving problem (M4), we need to check if there exist some λ_j, μ_{hik}^s, ν_{hik}^s such that $\lambda_j = M_\lambda$, $\mu_{hik}^s = M_\mu$, or $\nu_{hik}^s = M_\nu$. If the answer is no, the values of M_λ, M_μ, M_ν are valid; otherwise, we need to increase M_λ, M_μ, M_ν and repeat this procedure.

It takes two iterations to obtain the optimal solution, and the results of the subproblem (S), master problem (M4) and some intermediate results are given in Table 2.

Table 2. Results of the low dimensional problem with x_0 = (0, 0)

Iter	subproblem (S)			master problem (M4)	
	x_0	δ^*	F^U	F^L	x^*
1	(0, 0)	0.863	0.863	0.353	(1, 1)
2	(1, 1)	0.51	0.51	0.51	(1, 1)[1]

Note: [1] This optimal value can be any combination of x, since every direction yields the same result in the last iteration.

In the first iteration, we start from x_0 = (0, 0) to solve the subproblem, and the upper bound of the flexibility index is 0.863. The range in which f_j is overestimated is $\theta_1 \in \left[\theta_1^L, \theta_1^U\right] = [3.546, 10.454]$, and $\theta_2 \in \left[\theta_2^L, \theta_2^U\right] = [2.819, 13.181]$. We can then construct the master problem in the form of (M1), as shown below:

$$
\begin{aligned}
& F^L = \min_{x} \max_{\delta, w_{12}^p, w_{12}^n} \delta \\
& \text{s.t. } w_{12}^p - 100 \le 0 \\
& \quad w_{12}^n + 10 \le 0 \\
& \quad w_{12}^p = \min\left(\theta_2^U\theta_1 + \theta_1^L\theta_2 - \theta_1^L\theta_2^U,\ \theta_2^L\theta_1 + \theta_1^U\theta_2 - \theta_1^U\theta_2^L\right) \\
& \quad w_{12}^n = \min\left(-\theta_2^L\theta_1 - \theta_1^L\theta_2 + \theta_1^L\theta_2^L,\ -\theta_2^U\theta_1 - \theta_1^U\theta_2 + \theta_1^U\theta_2^U\right) \\
& \quad \theta_1 = 7 + \delta\left[4x_1 - 4(1 - x_1)\right] \\
& \quad \theta_2 = 8 + \delta\left[6x_2 - 6(1 - x_2)\right] \\
& \quad \delta \le F^U \\
& \quad x_1, x_2 \in \{0,1\}
\end{aligned}
\tag{31}
$$

Its corresponding feasibility function is stated in the following:

$$
\varphi(d, \theta) = \max\left\{
\begin{array}{l}
f_1^l = \min\left(\theta_2^U\theta_1 + \theta_1^L\theta_2 - \theta_1^L\theta_2^U,\ \theta_2^L\theta_1 + \theta_1^U\theta_2 - \theta_1^U\theta_2^L\right) - 100, \\
f_2^l = \min\left(-\theta_2^L\theta_1 - \theta_1^L\theta_2 + \theta_1^L\theta_2^L,\ -\theta_2^U\theta_1 - \theta_1^U\theta_2 + \theta_1^U\theta_2^U\right) + 10 \\
\theta^L - \theta \\
\theta - \theta^U
\end{array}
\right\}
\tag{32}
$$

The feasible region is depicted in Figure 5, in which the black dot in the center is the nominal

point. The dotted rectangle is the region in which the constraint is overestimated, i.e., the hyperrectangle $T(F^U)$. The shaded area and the solid rectangle with $F^L = 0.353$ are the feasible region and flexible region for the master problem, respectively. The four arrows represent the maximum distance from the nominal point to the boundary of the feasible region of the linearized system. We also depict the feasible region for the original problem in Figure 5, corresponding to the region between the two blue solid curves.

In the first iteration, X^+={(0, 1), (1, 0)}, $X^=$={(0, 0)}, X^-={(1, 1)}. The graph is consistent with the aforementioned conclusions about the feasible region in the properties of the master problem section. The three vertices of $T(F^U)$ along the directions in $X^+ \bigcup X^=$ are on the boundary, and the vertex corresponding to the maximum deviation along the direction in X^- in the original problem is outside of the feasible region. There is only one element in X^-, which is exactly the critical direction for the original problem. Therefore, the solution will converge in the next iteration.

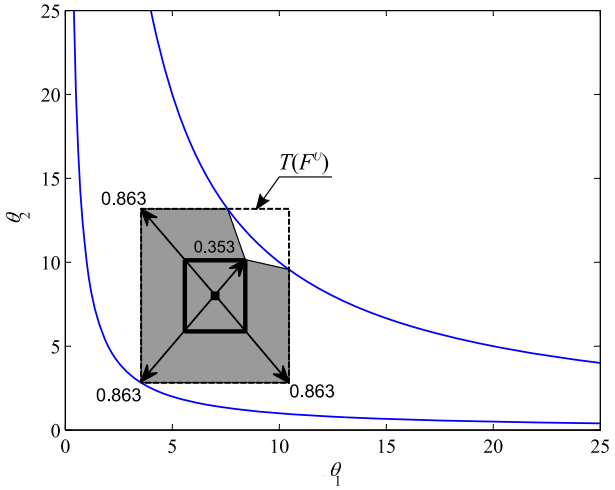

Figure 5. Feasible region in the first iteration with x_0 = (0, 0)

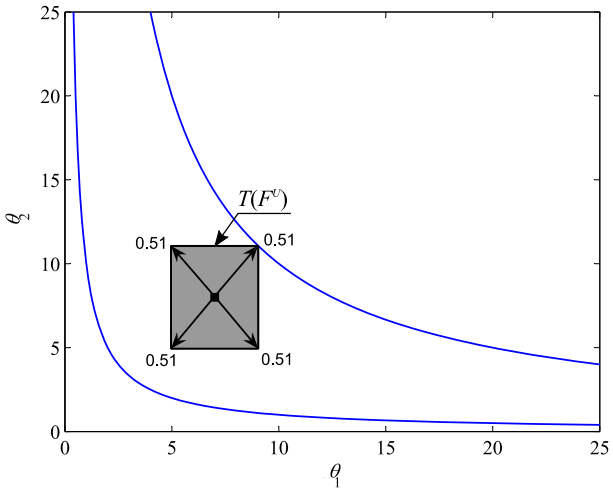

Figure 6. Feasible region in the second iteration with x_0 = (0, 0)

In the second iteration, we use the optimal solution x^* of the first iteration as the vertex direction to solve the subproblem, and the new upper bound of the flexibility index is 0.51, which is exactly the flexibility index. Therefore, X^- is empty, and all the vertices of $T(F^U)$ are on the boundary of the feasible region, which means $T(F^U)$ is exactly the same as the feasible region of the master problem as shown in Figure 6. Hence, the optimum of the master problem, 0.51, is equal to F^U, and the algorithm converges in two iterations.

If we start from x_0 = (0, 1), the results are given in Table 3. In the first iteration, the master problem (M4) has a negative optimum which means that the constraints are greatly overestimated, such that even the nominal point is infeasible in the linearized system as shown in Figure 7. Strictly speaking, this is not a flexibility index problem since the nominal point is infeasible. If we solve problem (31) directly, the only solution is x^*= (0, 1), and the algorithm cannot converge. But in practice, we solve (M4) instead, which dualizes the inner maximization problem of (M1). As discussed in the properties of the master problem section, if there are infeasible directions in (M1), the optimum of (M4) will be negative and the

optimal value of x will be one of these infeasible directions.

Table 3. Results of the low dimensional problem with x_0 = (0, 1)

Iter	subproblem (S)			master problem (M4)	
	x_0	δ^*	F^U	F^L	x^*
1	(0, 1)	1.608	1.608	-25.849	(1, 1)
2	(1, 1)	0.51	0.51	0.51	(1, 1)

The conclusions about the feasible region of the master problem are still valid. In this case, X^+ is empty, $X^=$={(0, 1), and X^-={(0, 0), (1, 0), (1, 1)}. We can only exclude the single direction in $X^=$. But fortunately, the optimal solution x^* is exactly the critical direction for the original problem. Therefore, the algorithm converges in the next iteration, just like starting from x_0 = (0, 0).

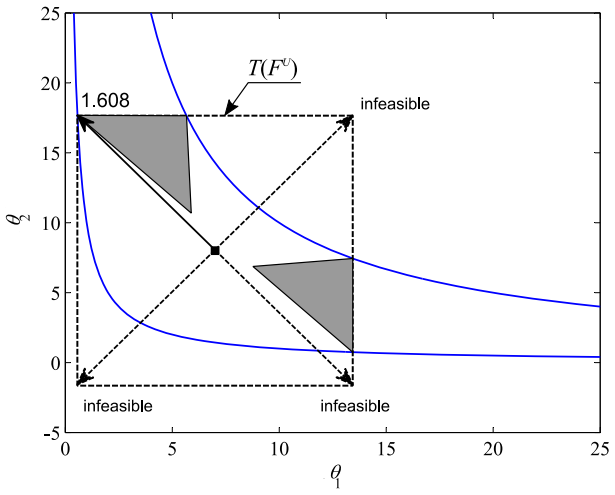

Figure 7. Feasible region in the first iteration with x_0 = (0, 1)

Heat exchanger network

This example of a small HEN is a modification of the example introduced by Grossmann and Floudas.[8] There are two hot streams, two cold streams, three heat exchangers and one

cooler in the HEN, as shown in Figure 8. The uncertain parameters are the inlet temperatures of the four streams and the heat capacity flow rates of H1 and C2, namely, T_1, T_3, T_5, T_8, F_{H1}, and F_{C2}. All the nominal values are shown in Figure 8, the expected deviations for the inlet temperatures are ±5 K, and the expected deviations for the heat capacity flows of H1 and C2 are assumed to be ±0.3 kW/K and ±0.5 kW/K, respectively. The only control variable lies in the heat load of the cooler (Q_c).

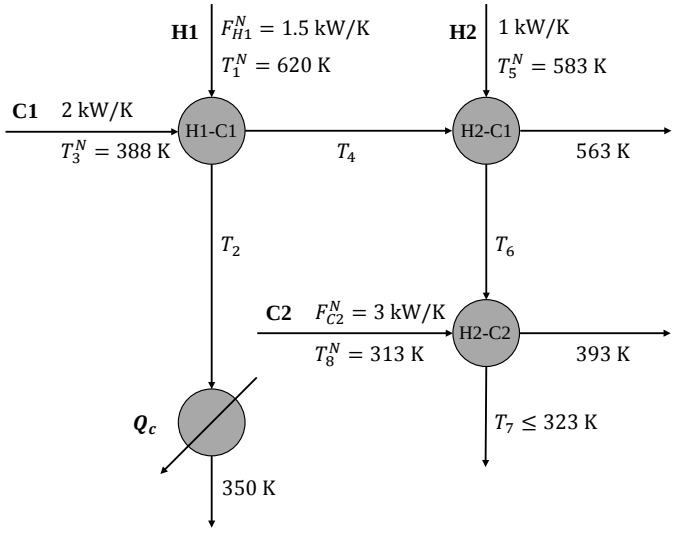

Figure 8. HEN with six uncertain parameters and one control variable

The inequality constraints are given as follows:

$$
\begin{aligned}
&f_1 = -350F_{H1} + F_{H1}T_3 - Q_c \leq 0 \\
&f_2 = -T_3 - T_5 + 175F_{H1} - 0.5F_{H1}T_1 + 0.5Q_c + 1126 \leq 0 \\
&f_3 = -2T_3 - T_5 + 350F_{H1} - F_{H1}T_1 + Q_c + 1519 \leq 0 \\
&f_4 = -2T_3 - T_5 + T_8 + 350F_{H1} + 393F_{C2} - F_{H1}T_1 - F_{C2}T_8 + Q_c + 1126 \leq 0 \\
&f_5 = 2T_3 + T_5 - 350F_{H1} - 393F_{C2} + F_{H1}T_1 + F_{C2}T_8 - Q_c - 1449 \leq 0 \\
&T_1, T_3, T_5, T_8, F_{H1}, F_{C2}, Q_c \geq 0
\end{aligned}
\tag{33}
$$

The traditional active-set method and the proposed duality-based method are applied to analyze the flexibility index of this system. The results, which yield F = 0.174, are given in Table 4. Although the MILP problem in the proposed duality-based method has more variables

than the MINLP problem in the active-set method, the duality-based method requires less time to solve this small example (0.094 s vs. 0.32 s).

Table 4. Flexibility index results of the HEN

Method	# of bin. variables	# of cont. variables	# of constraints	# of iterations[2]	F	Solution time [s]
Active-set	12	23	52	-	0.174	0.320
Duality	0/6[1]	2/145[1]	12/344[1]	2	0.174	0.094[3]

Notes: [1] The first number denotes the subproblem, the second denotes the master problem

[2] The number of iterations in the duality-based method

[3] Summation of all time used in the solvers

It is clear from Table 4 that the active-set method has much fewer constraints and continuous variables than the MILP problem in the duality-based method, but it takes longer to solve. The reasons are two-fold. On one hand the active-set method has to solve a complex MINLP problem with nonconvex bilinear terms, while the duality-based method only needs to solve some small NLPs with $n_z + 1$ variables (where n_z is the number of control variables) and easier MILPs; on the other hand, the MINLP in the active-set method usually has more binary variables than the MILP in the duality-based method, because the former has n_j (number of constraints) binary variables, while the latter has n_θ (number of uncertain parameters) binary variables, and there are usually more constraints than uncertain parameters.

The performance of the duality-based method depends on the initial vertex direction used to solve the subproblem. If the optimal vertex direction is used as the initial vertex direction, only

one iteration is needed. Here, the duality-based method is tested with all possible initial vertex directions. All the runs yield the correct result, and on average, they take 2.4 iterations and 0.16 seconds to converge.

Planning of process network

Consider a process network consisting of 6 processes and 10 chemicals as shown in Figure 9.[33] The following planning model of this process network is the one used by Zhang et al.[3]

$$Q_j^0 + \sum_{t'=1}^{t}\left(\sum_{i\in\hat{I}_j}\mu_{ij}P_{it'} - \sum_{i\in\bar{I}_j}\mu_{ij}P_{it'} + W_{jt'} - D_{jt'}\right) \geq Q_j^{\min}, \qquad \forall j \in J, t \in T \tag{34a}$$

$$Q_j^0 + \sum_{t'=1}^{t}\left(\sum_{i\in\hat{I}_j}\mu_{ij}P_{it'} - \sum_{i\in\bar{I}_j}\mu_{ij}P_{it'} + W_{jt'} - D_{jt'}\right) \leq Q_j^{\max}, \qquad \forall j \in J, t \in T \tag{34b}$$

$$P_{it} \leq P_i^{\max}, \qquad \forall i \in I, t \in T \tag{34c}$$

$$W_{jt} \leq W_{jt}^{\max} - \sum_{j'\in\bar{J}_j}\xi_{jj'}D_{j't}, \quad \forall j \in \bar{J}, t \in T \tag{34d}$$

$$P_{it} \geq 0, \quad \forall i \in I, t \in T \tag{34e}$$

$$W_{jt} \geq 0, \quad \forall j \in \bar{J}, t \in T \tag{34f}$$

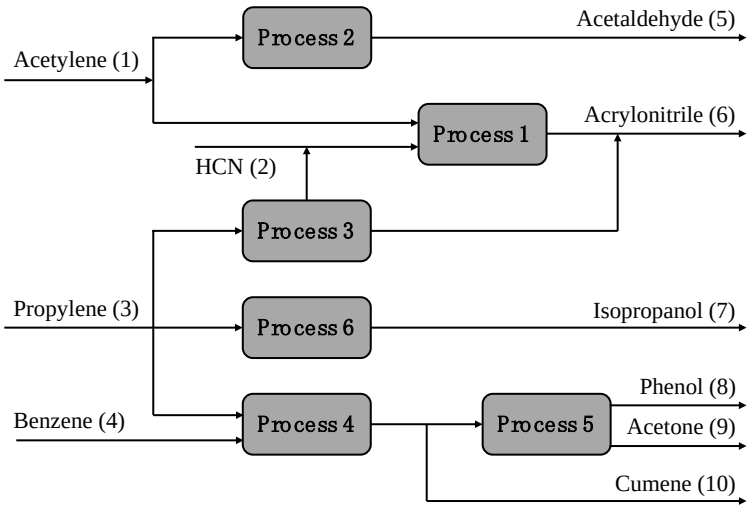

Figure 9. Process network superstructure

In this model, *I*, *J*, and *T* are sets of processes, chemicals, and time periods, respectively. Q_j^0,

$Q_j^{\min}$, and $Q_j^{\max}$ are the initial, minimum and maximum inventories, respectively. P_{it} is the total amount of chemicals produced by process *i* in time period *t*. μ_{ij} denotes the conversion fraction of chemical *j* by process *i* with respect to the total amount. Therefore, the amount of chemical *j* consumed or produced by process *i* in time period *t* is given by $\mu_{ij}P_{it}$. $\hat{I}_j$ and $\bar{I}_j$ are sets of processes producing and consuming chemical *j*, respectively. W_{jt} denotes the purchased amount of raw material *j* in time period *t*, and D_{jt} denotes the demand for product *j* in time period *t*. $\bar{J}$ is the set of raw materials (chemicals 1 to 4). The data is available in the supplementary material.

Equations (34a) and (34b) ensure that the inventory level lies within the bounds at every point in time. Equation (34c) corresponds to the capacity constraint. Equation (34d) is the purchase limit for raw materials. As stated by Zhang et al.,[3] it is assumed that the purchase limit of chemical *j* depends on the demand of products consuming *j*, denoted by the set $\bar{J}_j$. The qualitative explanation is that as the demand increases, the availability of its corresponding raw materials decreases. For example, the purchase limit of acetylene is affected by the demand of acetaldehyde and acrylonitrile. The coefficient $\xi_{jj'}$ describes how much the availability is affected by the demand. Equations (34e) and (34f) are nonnegativity constraints.

In this model, the coefficient $\xi_{jj'}$ and the demand D_{jt} for products (chemicals 5 to 10) are considered to be uncertain parameters, and P_{it} and W_{jt} are the control variables. For different number of time periods, N_T, the number of uncertain parameters, control variables, and constraints are $6N_T + 7$, $10N_T$, and $40N_T$, respectively.

The flexibility index problem for this process network with N_T uncertain parameters

varying from 1 to 4 is solved by the active-set method and the proposed duality-based method, as shown in Table 5. The time limit for each case is one hour.

Table 5. Flexibility index results of the process network problem

N_T	Method	# of bin. variables	# of cont. variables	# of constraints	# of Iterations	F	Gap [%]	Solution time [s]
1	Active-set	40	105	159	-	2.131	0	0.51
	Duality	0/13	24/646	53/1747	2	2.131	0	0.58
2	Active-set	80	201	301	-	0.829	0	198.5
	Duality	0/19	40/1770	99/4932	2	0.829	0	5.0
3	Active-set	120	297	443	-	0.556	100	3600
	Duality	0/25	56/3374	145/9557	2	0.556	0	18.8
4	Active-set	160	393	585	-	0.050	100	3600
	Duality	0/31	72/5458	191/15,622	2	0.050	0	51.7

In the smallest instance ($N_T = 1$), both methods can solve the flexibility index problem to optimality quite efficiently. The active-set method is slightly faster than the duality-based method since the latter needs to solve 2 NLPs and 2 MILPs, which are not significantly easier than the small MINLP in the former. As the problem size increases, the number of binary variables and bilinear terms in the active-set method increases quickly. Therefore, the duality-based method becomes superior to the active-set method. In the larger instances ($N_T = 3$, $N_T = 4$), the active-set method can obtain the correct result, but it cannot prove optimality. The lower bound does not improve (remains zero) during the computation. However, the proposed

duality-based method can solve all the instances within one minute.

Another observation is that, it takes only a few iterations to converge to the optimum in the duality-based method, regardless of the number of uncertain parameters and the choice of the initial vertex direction. There is no theoretical guarantee of this property, but the direction corresponding to the smallest deviation in the original problem tends to also have the smallest deviation in the master problem.

Security constrained unit commitment problem

In this example, we consider the unit commitment model, which arises in electric power systems.[34] It deals with the scheduling of I power generators over T time periods in a power generation plant. The MINLP model, which involves uncertainties in the power generated, consists of the following constraints:

$$\sum_{i=1}^{I}\sum_{t=1}^{T}\left[\left(a_i + b_i p_{it} + c_i p_{it}^2\right) + y_{it} SD_i + z_{it} SU_i\right] \le C^U \tag{35a}$$

$$\sum_{i=1}^{I} p_{it} + R_t \le \sum_{i=1}^{I} p_i^U \qquad \forall t \tag{35b}$$

$$p_{it}^L \le p_{it} \le p_{it}^U \qquad \forall i, t \tag{35c}$$

$$\sum_{i=1}^{I}\sum_{t=1}^{T} EC_i p_{it} \le E^U \tag{35d}$$

$$R_t^L \le R_t \le R_t^U \qquad \forall t \tag{35e}$$

$$p_{it} \le p_{i,t-1} + RU_i \qquad \forall i, t \tag{35f}$$

$$p_{it} \ge p_{i,t-1} - RD_i \qquad \forall i, t \tag{35g}$$

where constraint (35a) is the convex objective function in the original scheduling problem solved by Niknam et al.,[35] but we introduce an upper limit of cost, C^U, in the flexibility index

problem. Equation (35b) is the spinning reserve constraint. Equation (35c) sets the power limit for every generator at every time period. Equation (35d) ensures that the carbon dioxide emissions satisfy the emission limit. Equation (35e) sets the spinning reserve limit for every time period. Equations (35f) and (35g) are the ramp rate limits.

In this model, the spinning reserve R_t are the control variables, the uncertain parameters correspond to the generated power p_{it}, and all the others are constants. In particular, y_{it} and z_{it} are binary variables in the original scheduling problem, but in the flexibility index problem they are fixed to the optima of the scheduling problem. All the constraints except Eq. (35a) are linear, and there are only quadratic terms with positive coefficients in the nonlinear constraint. Therefore, the proposed duality-based method is applicable.

We consider a case consisting of 10 units and 24 time periods. The data is available from the website www.minlp.org/library/problem/index.php?i=41&lib=MINLP. The corresponding flexibility index problem is a very large problem with 240 uncertain parameters, 24 control variables, and 1034 constraints. The nominal operating point is the optimal solution of the scheduling problem, in which all the generators are forced to operate during all time periods. This can simplify the formulation of the flexibility index problem, since we do not need to deal with the special case where some generators are shut down. The expected deviation is set to be 5 percent of the nominal value, and the spinning reserve can only be adjusted within 10 percent of the nominal value. To avoid the flexibility index being zero, the upper limit of the total cost C^U is 700,000 \$/h, which is greater than the optimal cost. The emission limit E^U is 200,000 g higher than the value (15,500,000 g) used in the original scheduling problem. The power

limit p_i^U of every generator is 10 MW higher than the original value. The ramp rate limits RU_i and RD_i are 10 MW/h higher than the original values.

The flexibility index problem is solved by the active-set method and the proposed duality-based method, and the results are shown in Table 6.

Table 6. Flexibility index results of the unit commitment problem

Method	# of bin. variables	# of cont. variables	# of constraints	# of Iterations	F	Solution time [s]
Active-set	1034	2334	3609	-	0.258	206.9
Duality	0/240	25/249,196	1034/744,506	2	0.258	51.5

Both methods yield the same result (F = 0.258), but the duality-based method is significantly faster than the active-set method (51.5 s vs. 206.9 s). The reason is the same as the one explained in the second example for HEN. The master problem of the duality-based method can be very large, since we have to introduce new variables and new constraints for every combination of binary variables and constraints to eliminate the nonlinearity. However, it only takes less than 30 seconds to solve the large MILP since the relaxed MILP (RMIP) has the same optimum as the MILP, although the binary variables are not integral in the RMIP. In contrast, the optimum of the relaxed MINLP of the active-set method is zero.

Conclusions

In this work, we have presented a new flexibility index algorithm for systems in which all the inequalities are quadratic or linear in θ, and linear in z. This class of problems is proved to have a vertex solution if all the diagonal elements of Q_j, $j \in J$, are non-negative. Based on

this property, the subproblem, which provides an upper bound of the flexibility index, is easily obtained. Similar to the idea of OA in convex MINLP, the master problem is constructed by overestimating the nonlinear constraints, providing a lower bound of the flexibility index. After eliminating the min operators in the inner maximization problem, dualizing the inner maximization problem, and introducing new variables and constraints, the master problem is reformulated as a tractable MILP. By iteratively solving the subproblem and the master problem, the algorithm can be guaranteed to converge to the flexibility index.

Four computational studies, which include a small example HEN, a process network and unit commitment, are presented to illustrate its applicability in solving flexibility index problem. The results show that the proposed algorithm is more efficient than the active-set method, especially for large-scale problems with more constraints than uncertain parameters.

Finally, this algorithm is not restricted to quadratic systems. It can be directly extended to inequalities with univariate convex nonlinear terms, since they can be overestimated by their secants just like the quadratic terms. For general nonlinear systems with vertex solutions, this algorithm also provides a promising approach. If one can develop the overestimation satisfying these conditions: (1) the overestimation is linear or piecewise line, (2) the overestimation is equal to the original function at all the vertices of $T(F^U)$, and greater than the original function inside of $T(F^U)$, (3) the maximum of the overestimation in $T(F^U)$ lies at a vertex, the algorithm in this work can also be applied to the general systems, but the master problem must be reformulated in a similar way.

Acknowledgment

The authors gratefully acknowledge the financial support from the China Scholarship Council and the Center for Advanced Process Decision-making at Carnegie Mellon University.

Notation

Variables: Algorithm

θ = uncertain parameters

θ^L, θ^U = lower, upper bounds of θ

θ^N = nominal values of uncertain parameters

$\Delta\theta^+$, $\Delta\theta^-$ = expected positive, negative deviations

λ, μ, v, η = Lagrangian multipliers

A = coefficient matrix of uncertain parameters

a_j = coefficient vector of uncertain parameters in f_j

a_{hik}^s, b_{hik}^s, d_{hik}^s = coefficients and constants for the linearized function

B = coefficient matrix of control variables

b_j = coefficient vector of control variables in f_j

C = constant vector of the inequalities

c_j = constant in f_j

d = design variables

f = process inequalities

f^l = linearized process inequalities

F = flexibility index

F^U = upper bound of flexibility index

F^L = lower bound of flexibility index

I = set of uncertain parameters

$I^p = \{(i,k) \mid i,k \in I, i < k, \exists j \in J, q_{jik} + q_{jki} > 0\}$

$I^n = \{(i,k) \mid i,k \in I, i < k, \exists j \in J, q_{jik} + q_{jki} < 0\}$

J = set of inequalities

M, M_λ, M_μ, M_ν = big-M values

Q_j = coefficient matrix of quadratic terms

q_{jii} = diagonal elements of Q_j

w_{ik}^p, w_{ik}^n = overestimations of $\theta_i\theta_k$, $-\theta_i\theta_k$

X^+ = set of x in which $\delta^*(x) > F^U$

$\mathrm{X}^=$ = set of x in which $\delta^*(x) = F^U$

X^- = set of x in which $\delta^*(x) < F^U$

x = vertex directions

x_0 = initial vertex direction

x^* = optimal solution of (M4)

z = control variables

Subscripts: Algorithm

i, k = index of uncertain parameters

j = index of inequalities

h = index of positive and negative bilinear terms

Superscripts: Algorithm

s = label of p and n

Literature Cited

1. Grossmann IE, Sargent RWH. Optimum design of chemical plants with uncertain parameters. *AIChE J.* 1978;24(6):1021-1028.

2. Grossmann IE, Calfa BA, Garcia-Herreros P. Evolution of concepts and models for quantifying resiliency and flexibility of chemical processes. *Comput Chem Eng.* 2014;70:22-34.

3. Zhang Q, Lima RM, Grossmann IE. On the Relation Between Flexibility Analysis and Robust Optimization for Linear Systems. *AIChE J.* 2016;62(9):3109-3123.

4. Swaney RE, Grossmann IE. An index for operational flexibility in chemical process design. part I. *AIChE J.* 1985;31(4):621-630.

5. Halemane KP, Grossmann IE. Optimal process design under uncertainty. *AIChE J.* 1983;29(3):425-433.

6. Grossmann IE, Halemane KP, Swaney RE. Optimization strategies for flexible chemical process. *Comput Chem Eng.* 1983;7(4):439-462.

7. Swaney RE, Grossmann IE. An index for operational flexibility in chemical process design. Part II. *AIChE J.* 1985;31(4):631-641.

8. Grossmann IE, Floudas CA. Active constraint strategy for flexibility analysis in chemical process. *Comput Chem Eng.* 1987;11(6):675-693.

9. Ostrovsky GM, Volin YM, Barit EI, Senyavin MM. Flexibility analysis and

optimization of chemical plants with uncertain parameters. *Comput Chem Eng.* 1994;18(8):755-767.

10. Ostrovsky GM, Achenie LEK, Wang YP, Volin YM. A new algorithm for computing process flexibility. *Ind Eng Chem Res.* 2000;39(7):2368-2377.

11. Bansal V, Perkins JD, Pistikopoulos EN. Flexibility analysis and design of linear systems by parametric programming. *AIChE J.* 2000;46(2):335-354.

12. Bansal V, Perkins JD, Pistikopoulos EN. Flexibility analysis and design using a parametric programming framework. *AIChE J.* 2002;48(12):2851-2868.

13. Floudas CA, Gumus ZH, Ierapetritou MG. Global optimization in design under uncertainty: Feasibility test and flexibility index problems. *Ind Eng Chem Res.* 2001;40(20):4267-4282.

14. Ierapetritou MG. New approach for quantifying process feasibility: Convex and 1-D quasi-convex regions. *AIChE J.* 2001;47(6):1407-1417.

15. Goyal V, Ierapetritou MG. Determination of operability limits using simplicial approximation. *AIChE J.* 2002;48(12):2902-2909.

16. Grossmann IE, Halemane KP. Decomposition strategy for designing flexible chemical plants. *AIChE J.* 1982;28(4):686-694.

17. Pistikopoulos EN, Grossmann IE. Optimal retrofit design for improving process flexibility in linear systems. *Comput Chem Eng.* 1988;12(7):719-731.

18. Pistikopoulos EN, Grossmann IE. Optimal retrofit design for improving process flexibility in nonlinear systems-I. Fixed degree of flexibility. *Comput Chem Eng.*

1989;13(9):1003-1016.

19. Pistikopoulos EN, Grossmann IE. Optimal retrofit design for improving process flexibility in nonlinear systems-II. Optimal level of flexibility. *Comput Chem Eng.* 1989;13(10):1087-1096.

20. Pistikopoulos E, Ierapetritou M. Novel approach for optimal process design under uncertainty. *Comput Chem Eng.* 1995;19(10):1089-1110.

21. Varvarezos DK, Grossmann IE, Biegler LT. A sensitivity based approach for flexibility analysis and design of linear process systems. *Comput Chem Eng.* 1995;19(12):1301-1316.

22. Floudas CA, Grossmann IE. Synthesis of flexible heat exchanger networks with uncertain flowrates and temperatures. *Comput Chem Eng.* 1987;11(4):319-336.

23. Papalexandri KP, Pistikopoulos EN. Synthesis and retrofit design of operable heat exchanger networks. 1. Flexibility and structural controllability aspects. *Ind Eng Chem Res.* 1994;33(7):1718-1737.

24. Pistikopoulos E, Thomaidis T, Melin A, Ierapetritou M. Flexibility, reliability and maintenance considerations in batch plant design under uncertainty. *Comput Chem Eng.* 1996;20:S1209-S1214.

25. Sirdeshpande AR, Ierapetritou MG, Andrecovich MJ, Naumovitz JP. Process synthesis optimization and flexibility evaluation of air separation cycles. *AIChE J.* 2005;51(4):1190-1200.

26. Kabatek U, Swaney RE. Worst-case identification in structured process systems.

Comput Chem Eng. 1992;16(12):1063-1071.

27. Glover F. Improved linear integer programming formulations of nonlinear integer problems. *Manage Sci.* 1975;22(4):455-460.

28. Duran MA, Grossmann IE. An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Math Program.* 1986;36(3):307-339.

29. McCormick GP. Computability of global solutions to factorable nonconvex programs: Part I-Convex underestimating problems. *Math Program.* 1976;10(1):147-175.

30. Al-Khayyal FA, Falk JE. Jointly constrained biconvex programming. *Math Oper Res.* 1983;8(2):273-286.

31. Trespalacios F, Grossmann IE. Review of Mixed-Integer Nonlinear and Generalized Disjunctive Programming Methods. *Chem Ing Tech.* 2014;86(7):991-1012.

32. Tawarmalani M, Sahinidis NV. A polyhedral branch-and-cut approach to global optimization. *Math Program.* 2005;103(2):225-249.

33. Iyer RR, Grossmann IE. A bilevel decomposition algorithm for long-range planning of process networks. *Ind Eng Chem Res.* 1998;37(2):474-481.

34. Yamin HY. Review on methods of generation scheduling in electric power systems. *Electr Pow Syst Res.* 2004;69(2):227-248.

35. Niknam T, Khodaei A, Fallahi F. A new decomposition approach for the thermal unit commitment problem. *Appl Energ.* 2009;86(9):1667-1674.

Appendix A: Proof of proposition 1

Based on Theorems 2 and 3 proved by Swaney and Grossmann,[4] if all the constraints f_j,

$j \in J$, are jointly quasi-convex in *z* and one-dimensional quasi-convex (1-DQC) in θ, then the solution of the flexibility index problem will lie at a vertex of the hyperrectangle $T(\delta)$. f_j is obviously quasi-convex in *z* since it is linear in *z*. Therefore, we only need to prove that f_j is 1-DQC in θ.

Define $\theta^1 = \left(\theta_1,...,\theta_i,...,\theta_{n_\theta}\right)^T$, $\theta^2 = \left(\theta_1,...,\theta_i+\beta,...,\theta_{n_\theta}\right)^T = \theta^1 + \beta e_i$, where β is a scalar, and e_i is the *i*th unit vector, i.e., the *i*th column of the identity matrix. f_j is 1-DQC in θ if and only if

$$\max\left\{ f_j\left(d,z,\theta^1\right),\ f_j\left(d,z,\theta^2\right)\right\} \geq f_j\left(d,z,\alpha\theta^1+\left(1-\alpha\right)\theta^2\right),\ \forall \alpha \in [0,1] \tag{A1}$$

After cancelling the $b_j^T z + c_j$ on both sides, the right-hand side (RHS) of Eq. (A1) can be written as follows:

$$\begin{aligned}
&\text{RHS}\\
&= \left[\alpha\theta^1+\left(1-\alpha\right)\left(\theta^1+\beta e_i\right)\right]^T Q_j \left[\alpha\theta^1+\left(1-\alpha\right)\left(\theta^1+\beta e_i\right)\right] + a_j^T\left[\alpha\theta^1+\left(1-\alpha\right)\theta^2\right]\\
&= \left[\theta^1+\left(1-\alpha\right)\beta e_i\right]^T Q_j \left[\theta^1+\left(1-\alpha\right)\beta e_i\right] + a_j^T\left[\alpha\theta^1+\left(1-\alpha\right)\left(\theta^1+\beta e_i\right)\right]\\
&= \theta^{1T}Q_j\theta^1 + \left(1-\alpha\right)\beta\theta^{1T}Q_j e_i + \left(1-\alpha\right)\beta e_i^T Q_j\theta^1 + \left(1-\alpha\right)^2\beta^2 e_i^T Q_j e_i + a_j^T\theta^1 + \left(1-\alpha\right)\beta a_j^T e_i\\
&= \theta^{1T}Q_j\theta^1 + \left(1-\alpha\right)\beta\theta^{1T}Q_j e_i + \left(1-\alpha\right)\beta e_i^T Q_j\theta^1 + \left(1-\alpha\right)^2\beta^2 q_{jii} + a_j^T\theta^1 + \left(1-\alpha\right)\beta a_j^T e_i
\end{aligned} \tag{A2}$$

If $\theta^{1T}Q_j\theta^1 + a_j^T\theta^1 \geq \theta^{2T}Q_j\theta^2 + a_j^T\theta^2$, the left-hand side (LHS) of Eq. (A1) after cancelling $b_j^T z + c_j$ will be $\theta^{1T}Q_j\theta^1 + a_j^T\theta^1$. By substituting $\theta^2 = \theta^1 + \beta e_i$ into $\theta^{1T}Q_j\theta^1 + a_j^T\theta^1 \geq \theta^{2T}Q_j\theta^2 + a_j^T\theta^2$, we get $\beta\theta^{1T}Q_j e_i + \beta e_i^T Q_j\theta^1 + \beta^2 Q_{jii} + \beta a_j^T e_i \leq 0$. Since $0 \leq 1-\alpha \leq 1$ and $q_{jii} \geq 0$, the following inequality holds:

$$\begin{aligned}
&\left(1-\alpha\right)\beta\theta^{1T}Q_j e_i + \left(1-\alpha\right)\beta e_i^T Q_j\theta^1 + \left(1-\alpha\right)^2\beta^2 q_{jii} + \left(1-\alpha\right)\beta a_j^T e_i\\
&\leq \left(1-\alpha\right)\beta\theta^{1T}Q_j e_i + \left(1-\alpha\right)\beta e_i^T Q_j\theta^1 + \left(1-\alpha\right)\beta^2 q_{jii} + \left(1-\alpha\right)\beta a_j^T e_i \leq 0
\end{aligned} \tag{A3}$$

Therefore, the following inequality is obtained:

$$\text{LHS} = \theta^{1T}Q_j\theta^1 + a_j^T\theta^1 \geq \text{RHS} \tag{A4}$$

which means Eq. (A1) holds. It is similar when $\theta^{1T}Q_j\theta^1 + a_j^T\theta^1 < \theta^{2T}Q_j\theta^2 + a_j^T\theta^2$.

Therefore, f_j is jointly quasi-convex in *z* and 1-DQC in θ, and hence the solution lies at a vertex. ■

Appendix B: Other cases for reformulating the min operators in (M1)

(1) $\Delta\theta_i^+ / \Delta\theta_i^- > \Delta\theta_k^+ / \Delta\theta_k^-$:

If $x_i = 1$ and $x_k = 0$, $w_{ik}^p = \theta_k^L\theta_i + \theta_i^U\theta_k - \theta_i^U\theta_k^L$; otherwise, $w_{ik}^p = \theta_k^U\theta_i + \theta_i^L\theta_k - \theta_i^L\theta_k^U$.

(2) $\Delta\theta_i^+ / \Delta\theta_i^- < \Delta\theta_k^+ / \Delta\theta_k^-$:

If $x_i = 0$ and $x_k = 1$, $w_{ik}^p = \theta_k^U\theta_i + \theta_i^L\theta_k - \theta_i^L\theta_k^U$; otherwise, $w_{ik}^p = \theta_k^L\theta_i + \theta_i^U\theta_k - \theta_i^U\theta_k^L$.

(3) $\Delta\theta_i^+ / \Delta\theta_i^- > \Delta\theta_k^- / \Delta\theta_k^+$:

If $x_i = 1$ and $x_k = 1$, $w_{ik}^n = -\theta_k^U\theta_i - \theta_i^U\theta_k + \theta_i^U\theta_k^U$; otherwise, $w_{ik}^n = -\theta_k^L\theta_i - \theta_i^L\theta_k + \theta_i^L\theta_k^L$.

(4) $\Delta\theta_i^+ / \Delta\theta_i^- < \Delta\theta_k^- / \Delta\theta_k^+$:

If $x_i = 0$ and $x_k = 0$, $w_{ik}^n = -\theta_k^L\theta_i - \theta_i^L\theta_k + \theta_i^L\theta_k^L$; otherwise, $w_{ik}^n = -\theta_k^U\theta_i - \theta_i^U\theta_k + \theta_i^U\theta_k^U$.

Appendix C: Proof of proposition 2

Let us consider $x_i = 0$ first. Since x_i is fixed to zero, but x_k is not fixed, θ_i and θ_k can be expressed as functions of δ:

$$\begin{aligned} \theta_i &= \theta_i^N - \delta\Delta\theta_i^- \\ \theta_k &= \theta_k^N \pm \delta\Delta\theta_k^\pm \end{aligned} \tag{C1}$$

Together with Eq. (20), the first term in the min operator of w_{ik}^p in (M1) minus the second one can be written as

$$
\begin{aligned}
\Delta &= \theta_k^U\theta_i + \theta_i^L\theta_k - \theta_i^L\theta_k^U - \left(\theta_k^L\theta_i + \theta_i^U\theta_k - \theta_i^U\theta_k^L\right) \\
&= \left(\theta_k^U - \theta_k^L\right)\theta_i + \left(\theta_i^L - \theta_i^U\right)\theta_k - \theta_i^L\theta_k^U + \theta_i^U\theta_k^L \\
&= F^U\left(\Delta\theta_k^+ + \Delta\theta_k^-\right)\left(\theta_i^N - \delta\Delta\theta_i^-\right) - F^U\left(\Delta\theta_i^+ + \Delta\theta_i^-\right)\left(\theta_k^N \pm \delta\Delta\theta_i^\pm\right) \\
&\quad - \left(\theta_i^N - F^U\Delta\theta_i^-\right)\left(\theta_k^N + F^U\Delta\theta_k^+\right) + \left(\theta_i^N + F^U\Delta\theta_i^+\right)\left(\theta_k^N - F^U\Delta\theta_k^-\right) \\
&= -\delta F^U\Delta\theta_i^-\left(\Delta\theta_k^+ + \Delta\theta_k^-\right) \mp \delta F^U\Delta\theta_k^\pm\left(\Delta\theta_i^+ + \Delta\theta_i^-\right) + F^{U2}\left(\Delta\theta_i^-\Delta\theta_k^+ - \Delta\theta_i^+\Delta\theta_k^-\right)
\end{aligned}
\tag{C2}
$$

Recall that we assume $\Delta\theta_i^+ = \Delta\theta_i^-$ and $\Delta\theta_k^+ = \Delta\theta_k^-$, then Eq. (C2) is equal to 0 if $x_k = 0$, i.e., $\theta_k = \theta_k^N - \delta\Delta\theta_k^-$; and can be simplified to the following equation if $x_k = 1$, i.e., $\theta_k = \theta_k^N + \delta\Delta\theta_k^+$.

$$
\Delta = -\delta F^U\Delta\theta_i^-\left(\Delta\theta_k^+ + \Delta\theta_k^-\right) - \delta F^U\Delta\theta_k^+\left(\Delta\theta_i^+ + \Delta\theta_i^-\right) \tag{C3}
$$

It is clear that $\Delta \le 0$ always holds if $x_i = 0$. Therefore, w_{ik}^p is equal to the first term of the min operator if $x_i = 0$, i.e., $w_{ik}^p = \theta_k^U\theta_i + \theta_i^L\theta_k - \theta_i^L\theta_k^U$.

The proof for $x_i = 1$ is similar, and is omitted for simplicity. ■

Appendix D: Example for converting mixed integer linear min-max problem to its dual

Let us consider the following min-max problem with a disjunction in the constraint similar to (M1):

$$
\begin{aligned}
&\min_{y\in\{0,1\}} \max_x c^T x \\
&\text{s.t.} \begin{bmatrix} \neg\, y \\ Ax \le b \end{bmatrix} \veebar \begin{bmatrix} y \\ Dx \le e \end{bmatrix}
\end{aligned}
\tag{D1}
$$

It can be decomposed to two maximization problems, and subsequently convert them to their dual respectively.

$$
\begin{array}{cc}
\text{If } y = 0 & \text{If } y = 1 \\
\min\limits_{\lambda} \lambda^T b & \min\limits_{\lambda} \mu^T e \\
\text{s.t. } A^T\lambda = c & \text{s.t. } D^T\mu = c \\
\lambda \ge 0 & \mu \ge 0
\end{array}
\tag{D2}
$$

where λ and μ are the dual variables for the two linear inequalities, respectively.

The result of problem (D1) is the minimum of two problems in (D2). Another method to solve problem (D1) is to apply big-M reformulation as shown below.

$$\begin{aligned} &\min_{y\in\{0,1\}} \max_{x} c^T x \\ &\text{s.t. } Ax \le b + My \\ &\quad Dx \le e + M(1-y) \end{aligned} \tag{D3}$$

By using the dual of the inner problem, it can be rewritten as

$$\begin{aligned} &\min_{y\in\{0,1\}} \min_{\lambda,\mu} \lambda^T (b+My) + \mu^T \left[e + M(1-y) \right] \\ &\text{s.t. } A^T\lambda + D^T\mu = c \\ &\quad \lambda, \mu \ge 0 \end{aligned} \tag{D4}$$

If y is fixed to be 0, μ will be 0 to minimize the objective function as long as $M > b - e$, which is exactly the same as the first scenario of problem (D2). On the other hand, λ will be 0 if y is fixed to be 1 and $M > e - b$, corresponding to the second scenario of problem (D2). This shows that problem (D1) is equivalent to problem (D4) if $M > |b - e|$. Therefore, the reformulation from (M2) to (M3) is correct if M is sufficiently large.

Appendix E: Properties of the feasible region of (M1)

We consider the feasibility function $\varphi(d, \theta)$ of (M1):

$$\varphi(d, \theta) = \min_{z} \max_{j\in J} \{ f_j^l, \theta^L - \theta, \theta - \theta^U \} \tag{E1}$$

Since θ is a function of x and δ, $\varphi(d, \theta)$ can also be denoted by $\varphi(d, x, \delta)$. We can draw some conclusions about the feasible region of (M1) defined by $\varphi(d, \theta) \le 0$ in the following:

(1) In the hyperrectangle $T(F^U)$, $\theta^L - \theta \le 0$ and $\theta - \theta^U \le 0$ always hold. But outside of $T(F^U)$, $\varphi(d, \theta) > 0$. Therefore, the feasible region can be regarded as the intersection of $T(F^U)$ and $\psi^l(d, \theta) \le 0$, where $\psi^l(d, \theta) = \min_{z} \max_{j\in J} f_j^l$.

(2) Since f_j^l is an overestimation of f_j in $T(F^U)$, $f_j^l \geq f_j$, $\forall j \in J$, always hold, therefore, $\psi^l(d,\theta) \geq \psi(d,\theta)$ must hold, which means the feasible region of (M1) must be contained in the one of the original problem.

(3) Since f_j^l is equal to f_j at all the vertices of $T(F^U)$, $\psi^l(d,\theta)$ is equal to $\psi(d,\theta)$ at all the vertices of $T(F^U)$. If $x \in \mathrm{X}^+ \bigcup \mathrm{X}^=$, the vertex $\theta = \theta^N + F^U\left[x\Delta\theta^+ - (1-x)\Delta\theta^-\right]$ is feasible in the original problem. Therefore, $\psi^l(d,x,F^U) = \psi(d,x,F^U) \leq 0$, but $\varphi(d,x,F^U) = 0$ since one of the last two terms in Eq. (E1) is equal to zero. If $x \in \mathrm{X}^-$, the vertex $\theta = \theta^N + \delta^*(x)\left[x\Delta\theta^+ - (1-x)\Delta\theta^-\right]$ is on the boundary of the feasible region of the original problem, therefore, $\psi^l\left(d,x,\delta^*(x)\right) > \psi\left(d,x,\delta^*(x)\right) = 0$, and hence $\varphi\left(d,x,\delta^*(x)\right) > 0$.

List of Figure Captions