Two-Stage Submodular Optimization of Dynamic Thermal Rating for Risk Mitigation Considering Placement and Operation Schedule

Qinfei Long, Student Member, IEEE, Junhong Liu, Student Member, IEEE, Chenhao Ren, Student Member, IEEE, Wengqian Yin, Student Member, IEEE, Feng Liu, Senior Member, IEEE, and Yunhe Hou, Senior Member, IEEE

Abstract—Cascading failure causes a major risk to society currently. And dynamic thermal rating (DTR) technique is a cost-effective approach to mitigate the risk by exploiting potential transmission capability. From the perspectives of service life and Braess paradox, it is important and challenging to jointly optimize the DTR placement and its operation schedule for changing system state, which is a two-stage combinatorial problem with only discrete variables, suffering from no approximation guarantee and dimension curse in traditional solving algorithms. Thus, the present work proposes a novel two-stage submodular optimization (TSSO) of DTR for risk mitigation considering placement and operation schedule. Specifically, it optimizes DTR placement with proper redundancy in first stage, and then determines the corresponding DTR operation for each system state in second stage. Under the condition of the Markov and submodular features in sub-function of risk mitigation, the submodularity of total objective function of TSSO can be proven for the first time. Based on this, a state-of-the-art efficient solving algorithm is developed that can provide a better approximation guarantee than previous studies by coordinating the separate curvature and error form. The performance of the proposed optimization model is verified by case results.

Index Terms—Risk mitigation, dynamic thermal rating, two-stage submodular optimization, cascading failure, Braess paradox, service life, sensor placement, operation schedule, combinatorial optimization, dynamic line rating.

I. INTRODUCTION

WITH rising load demand and increasing extreme weathers, power system faces a growing risk of cascading failure (CF), inflicting significant harm to human life and industrial generation [1]. In 2021, for example, a freezing disaster in Texas left more than 10 million people without electricity and cost more than $130 billion in industrial damages [2]. Thus, mitigating the CF risk becomes a primary concern for each country.

Traditional CF risk mitigation strategies involve unit commitment [3], topology switching [4], line hardening [5] and so on. While these strategies are beneficial to risk mitigation plans, they commonly need large investments or are significantly affected by natural factors such as terrain. Fortunately, the development of smart sensor enables power system to utilize its potential capability to mitigate risk at a lower cost. One promising technology is dynamic thermal rating (DTR), which is often placed in overhead lines to dynamically set the thermal rating depending on real-time environment to maximize the usage of existing transmission assets [6].

As a result of benefits that DTR can provide, several scientists have analyzed DTR through simulation modeling [7], capacity control [8], operation scheduling [9], and risk mitigation [10], [11], in which researchers typically assume that DTRs are already placed on each line. One significant drawback is that, due to the large number of transmission lines, placing DTRs on all lines or most of them still results in high costs. Therefore, it is critical to place the DTR in the precisely selected lines to maximize its influence.

Sensor placement is a classic combinatorial optimization problem with exponential computational complexity. To solve placement optimization, assessment indexes [12], mixed integer programming [13], traditional approximation method [14] and heuristic algorithms [15] have been applied. However, owing to the dimensionality curse, some of them are not appropriate for large-scale cases, and there is a lack of analytical approximation guarantee, defined as the ratio of the worst obtained sub-optimal value to the real optimal value, potentially resulting in severe performance loss.

Another gap is that few studies consider the DTR service life [16] following placement. Generally, sensor power sources are limited and irreplaceable. While some DTR sensors are designed to be self-powered by feeding off the electromagnetic field [17], [18], they are nevertheless difficult to operate throughout the day due to their inefficient charging rate and immature self-powered technique [19]. Also, DTR is prone to failure owing...
to hostile environments, as DTR in operation is susceptible to intense electric and magnetic fields [20]. Moreover, excessive operational DTRs may inversely result in higher system risk, as described by the Braess paradox [21], [22]. Therefore, in addition to planning the DTR placement, its operation schedule also plays a crucial role in achieving a tradeoff between efficient risk mitigation and longer service time.

Furthermore, the problem consisting of placement and operation schedule can be modeled as two-stage optimization, as shown in PMU deployment [23], DG placement [24], fault current limiters stationing [25], charging station planning [26] and edge service configuration [27]. Some [24], [25], [27] contain both discrete and continuous variables, while others [23], [26] only include the discrete. In this study, DTR placement and operation schedule only comprise the discrete variables in two stages, an approach known as two-stage combinatorial optimization [28], which suffers from the similar issues as the sensor placement problem.

To fill the gaps in analytical guarantee and exponential computational complexity, several researchers currently combine two-stage combinatorial optimization with submodular function to develop a new research direction, two-stage submodular optimization (TSSO) [28], [29]. Its solving algorithms can provide an analytical approximation guarantee and give solutions in polynomial time. Specifically, [28] initially proposed the TSSO for the problem of learning sparse combinatorial representations. After that, [29] specializes the TSSO application in the contents of unknown distribution. TSSO is also applied to streaming scenario [30], difference type [31] and so on.

To investigate the benefit of DTR in mitigating CF risk, this study proposes a two-stage submodular optimization model of DTR’s placement and operation schedule. Based on Markov probability and sampling weight, a submodular sub-function for CF risk mitigation has been built considering Braess paradox. Using Markov feature, we analytically show that the total objective function of TSSO is of submodularity. Then, by coordinating the separate curvature and related error, a state-of-the-art solving algorithm is designed, giving a superior guarantee larger than present TSSO related guarantees, and obtaining solution in a polynomial time. The contributions of this article include:

1) We utilize TSSO to establish a DTR model consisting of both placement and operation schedule, which is a two-stage combinatorial optimization with only discrete variable. To the best of the authors’ knowledge, this is the first time the TSSO model is applied to electrical field.
2) We prove that, unlike the existing TSSO analyses, the total objective function of TSSO is submodular when using the Markov feature. Note that this conclusion holds valid for TSSO models that meet the condition: the sub-function has submodular and Markov features.
3) Given TSSO submodularity, we devise an state-of-the-art solving algorithm based on separate curvature, enabling us to analytically derive the performance guarantee, which is better than previous TSSO related guarantees [28], [29], [30], [32], [33] and traditional guarantee $1 - 1/e$ (from single stage optimization) [34], [35]. This algorithm also lowers the computational complexity of combinatorial optimization from exponential to polynomial.
4) Case studies present the impacts of environment, separate curvature, and failure probability parameter on DTR risk mitigation. Additionally, the comparison of TSSO and one-stage optimization is conducted, indicating that two-stage model can offer a superior mitigation effect for each system state, and prolong the DTR service life. A performance comparison of different two-stage strategies is also performed, suggesting that the suggested method outperforms other strategies. And an analysis of calculation efficiency related to case scales is conducted.

The rest of papers is organized as follows. Section II introduces the fundamentals of DTR and submodular optimization. Section III provides the models of DTR-based risk mitigation. Sections IV and V present the TSSO-based risk mitigation model and related solving algorithm analysis. Case studies, conclusion and future work are given in Sections VI and VII.

II. PRELIMINARIES

A. Dynamic Thermal Rating and Sensor Impacted Factors

DTR is an advanced technique that measures the conductor status and ambient environment and then transmits data to system operator to dynamically determine a new thermal rating [6]. DTR is viewed as a cost-effective alternative to other traditional hardening strategies, which exploits potential transmission capability without changing system configuration and defers costly line construction. According to relevant studies from academic [13], [15] and industries [36], [37], DTR system can boost the current line capacity by 10%-30%, with the most increase of 50% in windy areas. In this study, the goal is to mitigate the CF risk through DTR placement and operation schedule, which is a type of long-term planning issue. Hence, the effect of DTR on planning issue is evaluated from historical long-term weather data, especially average weather data over a certain period [11], [12], [13], [15].

In power system protection, the DTR value can be regarded as a new estimated threshold for relay action. Specifically, this threshold value is predefined to guide the relay action based on data for a period, long enough to ignore the non-steady state in heat balance and allow the application of static heat-balance equation, as described below [11], [12], [13], [15]:

$$Q_c + Q_s = I^2 R + Q_s$$  \hspace{1cm} (1)

where $I^2 R$ denotes the conductor heating from power flow, $Q_s$ the solar radiation heating, $Q_c$ the wind cooling, $Q_r$ the radiation cooling. Specifically, the values of $Q_c$, $Q_r$ and $Q_s$ are connected to the ambient weather condition, and their detailed formulas can be found in [6]. For convenience, we adopt $I^2 R$, the power as DTR value in the following risk mitigation model.

On the other hand, DTR sensors consist of two configurations [19], [38]. The first is located within the substation, powered by a battery system and already connected to data network. The second is used at tower sites, where it is powered by the outside environment and only relies on wireless communications. Although tower-based DTR takes the majority in transmission grid, it has some limitations. First, it is hard for sensors to charge themselves directly from high voltage power. Due to immature self-powered technique, it is impractical to
keep DTR operating at all time [17], [18]. To conserve energy, for instance, the DTRs in North Wales system only transmit data at thirty-minute intervals [38]. Second, since the memory capacity of DTR sensor is only a few megabytes, frequent and needless operations will lower the efficiency and accuracy of data communication [16], [20]. Third, the outside environment, particularly some extreme weathers, has an impact on DTR service life. Finally, the DTR tiny units are subjected to intense electric and magnetic fields as well as strong electric transients induced by switching surges and lightning [20]. Thus, excessive operation will accelerate sensor outage, as NERC reported in 2010 that 941 of 3519 tested DTR sensors had discrepancies due to longer running time [19]. In other words, DTR service life can be extended by limiting sensor operating duration. And we only consider the operation time in this work to estimate the DTR service life, since it is easily quantifiable while other factors are not.

B. Two-Stage Submodular Optimization

Submodularity is defined as the diminishing return feature in some special set functions, whose variables are restricted to be discrete. Specifically, it indicates that the incremental “value” of adding components to a set $S_A$ declines as $S_A$ grows larger. If a component is in $S_A$, its value is 1, otherwise it is 0. A typical example is that the cost of combo of hamburger and fried chicken is less than buying them separately. Sensor coverage also exhibits the phenomenon of diminishing return. This property appears in many domains, including economics, computer science, network analysis and so on [34]. It is defined as follows.

**Definition 1:** ([35]) Suppose discrete components set $S_A \subseteq S_B \subseteq S_L$ and set $v \in S_L \setminus S_B$, where $S_L$ is the ground set. A set function $f : 2^{S_L} \rightarrow \mathbb{R}$ is submodular if it satisfies

$$f(S_A \cup v) - f(S_A) \geq f(S_B \cup v) - f(S_B)$$

(2)

In terms of definition, we can obtain the following Lemmas [35] which provide a facilitated tool for determining if a function is submodular.

**Lemma 1:** If $\alpha_i \geq 0$ and $f_i : 2^{S_L} \rightarrow \mathbb{R}$ is submodular function, then so is $\sum \alpha_i f_i$.

**Lemma 2:** Any submodular function $f$ can be represented as a sum of submodular functions $f$ and a modular function $m$, i.e., $f = f + m$.

Based on the submodularity definition, model-free two-stage submodular optimization (TSSO) can be constructed. Specifically, TSSO selects an optimal subset from ground set in first stage, and then decides the appropriate strategy for each submodular sub-function in second stage [28]. Note that each sub-function in TSSO needs to be monotone [28], [29], [30]. And TSSO is a form in which the objective functions in two stages are the same or similar, and only contains discrete variables. The detail is shown in Section IV.

According to early study [28], TSSO is a combinatorial counterpart of representation learning tasks. Following [28], [29] specializes TSSO application for unknown distribution, whereas [30] broadens TSSO to distributed and streaming cases. Moreover, Other references analyze TSSO from different points, such as non-negative nor monotone relaxation [31], curvature index [32], subsampling model [39], meta-learning [40], extended $P$-matroid constraint [33]. In short, the research of TSSO is still in its infancy, and there are numerous improvements and application scenarios to be explored.

C. Curvature

The notion of curvature reflects how much a set function’s marginal values might fall. The definition of curvature is below.

**Definition 2:** ([41]) Suppose $S_L$ is ground set, $f$ a monotone submodular function, and $j \in S_L$, a single component. The curvature can be defined as

$$\kappa_f = 1 - \min_{j \in S_L} \frac{f(S_L) - f(S_L \cup j)}{f(j) - f(\emptyset)}$$

(3)

where $S_L \setminus j$ is difference set.

If the reduction of final marginal profit is bigger, there is a larger curvature value, indicating that this function has stronger diminishing return impact. Conversely, a smaller value of curvature expresses that the function is close to linear, implying that its marginal profit is not greatly affected by the selected set size. An example about curvature is that if the long tail effect of product sale is more significant, the curvature value is larger. In [42], the curvature was originally applied to deduce an improved approximation bound, as shown in the lemma below.

**Lemma 3:** ([42]) If $f_i : 2^{S_L} \rightarrow \mathbb{R}$ is a monotone increasing submodular function with curvature $\kappa_f$, then for all $T \subseteq S_L$, it has

$$\sum_{t \in T} [f(S_L) - f(S_L \setminus t)] \geq (1 - \kappa_f) f(T)$$

(4)

III. DYNAMIC THERMAL RATING-BASED RISK MITIGATION MODEL

A. Modeling of Risk Mitigation

Cascading failure is a dynamic process that can be represented by a failure chain, which is described as a sequence of failure generations containing system conditions in each failure propagation stage. Define $M^j = \{M^{(1)}, M^{(2)}, \ldots, M^{(k)}, \ldots\}$ as the failure chain sub-database from the $j$-th initial system state (before failures occur) contained in total database $M$, and $M^{(k)} = \{FG_0^{(k)}, FG_1^{(k)}, \ldots, FG_d^{(k)}\}$ as the $k$-th cascading failure chain in $M^j$, where $FG_i^{(k)} \in 2^{S_L}$ denotes the $i$-th failure generation in $M^{(k)}$, $d$ is the generation number of $M^{(k)}$, and $S_L$ is the line set. In this work, the system states in $M$ have different cascading failure evolution patterns that are determined by different load levels, whose corresponding electric generations are obtained by OPF. And most of states are under heavy load level to find as many failure chains as possible. And there is no renewable generator in our analyzed system, because renewables make it difficult to select representative system states and are negatively related to risk mitigation [43] resulting in a non-monotone TSSO’s objective function.

Owing to the piecewise structure, the single component failure probability $\varphi_c(FG_i^{(k)})$ is approximated smoothly by Sigmoid
form [43]:

$$\varphi_e(FG_i^{(l)}) = Pr_e^{\min} + \frac{Pr_e^{\max} - Pr_e^{\min}}{1 + \exp\left(-\mu \frac{2\delta_e}{\alpha Pr_e^{\max} + \alpha (Pr_e^{\min} + Pr_e^{\max})}\right)} \quad (5)$$

where $Pr_e^{\min}$, $Pr_e^{\max}$, respectively denote current power flow, minimum and maximum transmission capabilities of line $e$, $Pr_e^{\min}$, $Pr_e^{\max}$ the minimum and maximum of failure probability, respectively, and $\mu$ the approximate factor to avoid the vanishing gradient problem. $\alpha$ denotes the DTR improved parameter for power transmission threshold, improving the original to $\alpha Pr_e^{\max}$ related to ambient weather [6]. When $\alpha = 1$ there is no DTR effect.

For cascading failure chain with Markovian property [44], we define $f p_e^{(l)}$ as the probability of a specific cascading failure chain $M_l^{(k)}$ with DTR placed in line $e$:

$$fp_e^{(l)} = \prod_{i=1}^{d} \prod_{j \in S_{fp_i}^{\text{non}}} \varphi_j(FG_i^{(l)} \cdot \prod_{j \in S_{fp_i}^{\text{non}}} (1 - \varphi_j(FG_i^{(l)}))) \quad \text{Other} \quad (7)$$

where $S_{fp_i}^{\text{non}}$ denotes the no-DTR line set that functioned normal in $FG_i^{(l)}$ but failed in $FG_i^{(l)}$, and $S_{fp_i}^{\text{non}}$ denotes the no-DTR line set functioned normal in $FG_i^{(l)}$. Denoting $\varphi'_e$ as a series of line $e$ updated failure probabilities, $H^{(l)}(\varphi'_e)$ is defined as

$$H^{(l)}(\varphi'_e) = \begin{cases} \prod_{i=1}^{d} \left(1 - \varphi'_e(FG_i^{(l)})\right) & d_e > d \\ \varphi'_e(FG_d^{(l)}) \cdot \prod_{i=1}^{d-1} \left(1 - \varphi'_e(FG_i^{(l)})\right) & \text{Other} \end{cases}$$

where $d_e$ is the generation that line $e$ fails. If $d_e > d$, it means there is no failure in line $e$ during $M_l^{(k)}$. For $fp_B^{(l)}$ in which lines set $S_B$ placed with DTR, it is similar to the above, whose details can be found in our previous research [43].

Moreover, define load loss induced by a single failure chain $M_l^{(k)}$ as $Y_{l(k)}$ from the cascading failure simulator [22]. Since cascading failure is treated as an infrequent but extreme event, which has less occurrence probability but causes great damage to the whole power system, we define $\delta_{Y_{l(k)}>Y_{ext}}$ as an indicator that selects the cascading failure chain in which the load loss exceeds $Y_{ext}$, the setting loss threshold. Specifically, $\delta_{Y_{l(k)}>Y_{ext}} = 1$ when $Y_{l(k)} > Y_{ext}$, otherwise $\delta_{Y_{l(k)}>Y_{ext}} = 0$.

Then we propose the original risk of a single failure chain $M_l^{(k)}$ (no DTR) as

$$Risk_{l(k)} = Y_{l(k)} \cdot f p_l^{(k)} \cdot \delta_{Y_{l(k)}>Y_{ext}} \quad (8)$$

Above all, the original risk of whole failure chains in $M_l$ can be modeled as

$$RiskW_{l(k)} = \mathbb{E}(Y \cdot \delta_{Y>Y_{ext}}) = \sum_{M_l^{(k)} \in M_l} Risk_{l(k)}^{(k)}$$

Similarly, if DTR is placed in set $S_A$, there are $Risk_{l(k)}^A$ and $RiskW_{l(k)}^A$ [43].

### B. Sampling Weight Technique

When some parameters change, such as line maximum capacity after DTR placement, the related cascading failure chains may differ, impacting failure chain probability and risk index. However, recreating the failure database based on the new system will incur a considerable computing burden, particularly in optimization issues. To resolve this problem, the sampling weight technique proposed in [44] is applied.

1) Sampling Weight in a Single Failure Chain: In $M_l^{(k)}$, given 2 different line sets $S_A$ and $S_B$ placed with DTR, the underlying relationship between $fp_A^{(l)}$ and $fp_B^{(l)}$ can be expressed as sampling weight

$$W_{A\rightarrow B} = \frac{fp_A^{(l)}}{fp_B^{(l)}} = \prod_{e \in S_A \setminus S_B} \frac{H^{(l)}(\varphi_{Ae})}{H^{(l)}(\varphi_{Be})} \prod_{e \in S_B \setminus S_A} \frac{H^{(l)}(\varphi_{Ce})}{H^{(l)}(\varphi_{Be})} \quad (10)$$

where $\varphi_{Ae}$ and $\varphi_{Be}$ denote the failure probabilities of line $e$ when DTR placed in $S_A$ and $S_B$, respectively, and $\varphi_{Ae}$ and $\varphi_{Be}$ the original probabilities. Note that $S$ can be empty set. From formula (10), if $S_B \subseteq S_A$ the value of $fp_B^{(l)}$ can be calculated by multiplying $W_{A\rightarrow B}$ by $fp_A^{(l)}$, where the calculation of $fp_B^{(l)}$ also can be similarly decomposed, showing Markov feature [43], [44].

2) General Sampling Weight for the Optimal Combination: Define $G_i(S) = \{T \subseteq S : T = \arg\max_i f_i(\cdot)\}$ as the optimal solution for function $f_i(\cdot)$ in feasible region $S$. For model facilitation, the term $W$ is used to represent the general sampling weight, the ratio of optimal objective values in different feasible region. Specifically, for sets $S_A$ and $S_B$, we have

$$\tilde{W}_{A\rightarrow B} = \frac{f_i(G_i(S_A))}{f_i(G_i(S_B))} \quad (11)$$

Similarly, $\tilde{W}_{A\rightarrow 0}$ and $\tilde{W}_{A\cup B\rightarrow 0}$ can be obtained.

### IV. TSSO-BASED RISK MITIGATION OPTIMIZATION

Since system has different states in terms of load, a unique DTR placement (or operation) scheme cannot provide the best risk mitigation effect in each state. On the contrary, an inappropriate DTR operation may result in Braess paradox, which states that improving or adding some components worsens system performance [43], as first discovered in transportation research [21]. Additionally, DTR service life is an essential aspect in planning that has received less attention in relevant research [7], [8], [9], [10], [11]. A better DTR operation schedule, i.e., scheduling the DTR operation wisely to reduce DTR operation time in a specific period while providing enough measurement of weather data in critical lines, not only helps the system mitigate risk more flexibly and pertinently, but prolongs the DTR service life,

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
lowering the failed rate. Compared to one-stage optimization, the TSSO-based method will place DTR with reasonable redundancy, providing additional options for designing the associated operation schedules. In this section, we will establish a DTR risk mitigation optimization model based on TSSO that consists of placement and operation schedule.

A. Submodular Optimization for a Single System State

In order to quantify the DTR placement effect on CF risk mitigation, define \( S_B \subseteq S_L \) as an allocation of potential lines with DTR placement. Then we want to find \( S_B \) to maximize the risk mitigation effect in \( M \) that

\[
f_i(S_B) = \text{Risk}W_B^i - \text{Risk}W_B^j + \eta \cdot \text{BPI} \tag{12}
\]

where \( \eta \in [0, 1] \) denotes an adjustment factor, \( \text{BPI} = \sum_{M(i)\in M} \max\left( w^{(k)}_{\text{com}} - 1, 0 \right) \cdot \text{Risk}_{\text{A}}^{(k)} \) denotes Braess paradox indicator, with a higher value indicating more side effects of Braess paradox. \( w^{(k)}_{\text{com}} \) denotes the compared sampling weight between new updating set \( S_B \) with previous updating set \( S_A \). Note that \( \text{Risk}W_B^i \) is constant. Compared to the non-monotonicity of function \( f \) in [43], function \( f_i(S_B) \) is monotone because it only contains the risk mitigation item, and the related cost item is turned into the cardinality constraint shown below. The model detail can be found in our prior work [43].

Theorem 1: ([43]) Sub-function \( f_i(S_B) \) is submodular.

B. Construction of TSSO

In two-stage submodular optimization (TSSO) application on risk mitigation model, the target is to select a subset of lines to place DTR in first stage, and then schedule the DTR operation mode (on/off) in a series of sub-functions to optimize the target value in second stage. The sub-functions correspond to various system states.

Without loss of generality, let TSSO contains a class of sub-functions \( f_1(\cdot), f_2(\cdot), \ldots \) in different system states, where sub-function \( f_1: 2^S \rightarrow \mathbb{R} \) is defined over \( S_L \). Among these sub-functions, there is a trade-off for the contribution to the selected set based upon diminishing return feature and cost constraint. The aim of TSSO is to find a set \( S \subseteq S_L \) of size less than \( k_{c1} \), whose subsets \( T_1, T_2, \ldots, T_i, \ldots \), with sizes less than constraint \( k_{c2} \), to maximize the mean value \( F(S) \):

\[
\max_{S:|S| \leq k_{c1}} F(S) = \max_{S:|S| \leq k_{c2}} \max_{T_1 \in S, |T_1| \leq k_{c2}} f_i(T_1) \tag{13}
\]

To build TSSO in this study, we can set the sub-function \( f_i \) as the risk mitigation function (12). DTR number constraints \( k_{c1} \) and \( k_{c2} \) are determined by grid potential transmission capacity, budget, risk mitigation requirement, failure rate and so on. Note that even if a class of sub-functions \( f_i \) are submodular in second stage, the total objective function \( F \) ceases to be submodular in general cases [28]. Since TSSO selects variable \( S \) in first stage to serve as a new feasible region for second stage optimizations, the sub-functions optimization may produce a completely different result than that from previous feasible region. On the other hand, Lemmas 1 and 2 cannot be applied directly to the proof of TSSO, since the total objective function of TSSO is composite function related to set \( S \), which is not applicable to the simple function form in Lemmas 1 and 2. Previous studies [29], [30], [31], [32] all adopt this precondition in TSSO, which greatly restrict the solving performance and have an inferior approximation guarantees. One contribution of this research is proving that the total function \( F \) is submodular when its sub-functions are of Markov and submodular features, as detailed in Appendix.

V. SEPARATE CURVATURE GREEDY SOLVING ALGORITHM

To obtain the optimal solution, the TSSO needs traversing all candidates repeatedly, resulting in exponential computation complexity. This indicates that TSSO is a NP-hard problem suffering from dimension curse. To get a sub-optimal solution with an acceptable computing time, it is appropriate to design a solving method with a reduced computation complexity, lowering the original exponential complexity to a polynomial-time one [34]. Previously, the solving algorithms were designed as index-based [22], [45], greed-based [28], [42] or surrogate-based strategies [29], [32], [33]. However, if the total objective function of TSSO is not submodular, they yield a suboptimal result with inferior approximation guarantee. In this part, we propose a state-of-the-art solving algorithm to deal with TSSO using separate curvature [41] and the submodularity of TSSO’s total objective function.

A. Solving Algorithm Design

In any system state, \( i \)-th sub-function \( f_i(T_i) \) can be divided into 2 parts, i.e.,

\[
f_i(T_i) = g_i(T_i) + c_i(T_i) \tag{14}
\]

where \( g_i(T_i) \) is a monotone non-negative submodular function and \( c_i(T_i) \) is a modular one. Let \( \Delta_i^j(x, T_i) = f_i(x \cup T_i^j) - f_i(T_i) \) denote the \( i \)-th marginal increasement of element \( x \) adding to set \( T_i \). Unlike the typical modular function settings in [31], [32], [42], \( c_i(T_i) \) in this work is separate as

\[
c_i(T_i) = c_{i1}(T_i) + c_{i2}(T_i) = \sum_{x \in T_i \cap S_{L1}} \Delta_i^j(x, S_{L1} \setminus x) + \sum_{x \in T_i \cap S_{L2}} \Delta_i^j(x, S_{L2} \setminus x) \tag{15}
\]

where \( c_{i1}(T_i) = \sum_{x \in S \cap S_{L1}} \Delta_i^j(x, S_{L1} \setminus x) \) is the \( i \)-th marginal benefit sum only individually counting components intersecting with \( S_{L1} \), \( c_{i2}(T_i) \) is similar form related to \( S_{L2} \). Separate sets \( S_{L1} \) and \( S_{L2} \) have \( S_{L1} \subseteq S_L \) and \( S_{L2} = S_L \setminus S_{L1} \).

Based on the separate form of sub-function, the solving algorithm can be designed. For convenience, the following notations are defined. Let \( \Delta_i^j(x, T_i^j) = g_i(x \cup T_i^j) - g_i(T_i^j) \) denotes the \( i \)-th marginal increasement, \( \nabla_i^j(x, y, T_i^j) = g_i(x \cup T_i^j \setminus y) - g_i(T_i^j) \) the \( i \)-th marginal gain of replacing \( y \) in set \( T_i^j \) with component \( x \notin T_i^j \), and \( i, j \) denote the indexes of sub-function and searching iteration, respectively. In addition,
there are 2 constructed differences
\[ \Delta_i(x, T_j^i) = \left(1 - \frac{p}{k}\right)^{k-j} \Delta_i(x, T_j^i) + c_i(x) \]  \hspace{1cm} (16) 
\[ \nabla_i(x, y, T_j^i) = \left(1 - \frac{p}{k}\right)^{k-j} \nabla_i(x, y, T_j^i) + c_i(x) - c_i(y) \]  \hspace{1cm} (17) 
where \( k, p \) denote the first stage cardinality constraint and constraint’s number of second stage, respectively. Note that \( p \) is not the second stage cardinality constraint \( k_{c2} \). When component \( x \) replaces the component in \( T_j^i \), the new set does not violate the constraints \( |T_j^i \cup x| \leq k_{c2} \) in this work.

Above all, the marginal gain of component \( x \) in each iteration can be expressed as
\[ \nabla_i(x, T_j^i) \]
\[ = \begin{cases} 
\Delta_i(x, T_j^i) & \text{if } |T_j^i \cup x| \leq k_{c2} \\
\max \left\{ 0, \max_{y:(T_j^i \cup x)|y| \leq k_{c2}} \nabla_i(x, y, T_j^i) \right\} & \text{Other} 
\end{cases} \]  \hspace{1cm} (18) 
Similarly, \( \text{Rep}_i(x, T_j^i) \) is defined to represent the element that would be replaced by component \( x \) as follows:
\[ \text{Rep}_i(x, T_j^i) = \begin{cases} 
\emptyset & \text{if } |T_j^i \cup x| \leq k_{c2} \\
\arg \max_{y:(T_j^i \cup x)|y| \leq k_{c2}} \nabla_i(x, y, T_j^i) & \text{Other} 
\end{cases} \]  \hspace{1cm} (19) 
After the notation definition, the solving algorithm is constructed in Algorithm 1. In detail, this algorithm works in \( k = k_{c1} \) rounds, and a carefully designed searching objective function derived from separate curvature is used to select a component \( x \) that maximizes the marginal value in each iteration.

B. Performance Analysis

Based upon the designed solving algorithm, the approximation guarantee can be derived. Given the first stage result \( S_j^i \) in each iteration \( j = 1, \cdots, k \), we define the surrogate function \( \Phi_j(S_j^i) \) as
\[ \Phi_j(S_j^i) = \sum_{i=1}^{m} \left[ (1 - \frac{p}{k})^{k-j} g_i(T_j^i) + c_i(T_j^i) \right] \]  \hspace{1cm} (20) 
where \( k, p \) denote the first stage cardinality constraint and constraint’s number of second stage, respectively, \( m \) is the number of sub-functions, and \( T_j^i, T_i^j \) are the corresponding searching result and practical optimum result for \( i \)-th sub-function, respectively. First, there are Lemmas 4 and 5 about the boundary.

\textbf{Lemma 4:} For \( j = 1, 2, \cdots, k \), it has
\[ \Phi_j(S_j^i) - \Phi_{j-1}(S_j^{i-1}) = \sum_{i=1}^{m} \left[ \nabla_i(x^j, T_j^{i-1}) + \frac{p}{k} \left(1 - \frac{p}{k}\right)^{k-j} g_i(T_j^{i-1}) \right] \]  \hspace{1cm} (21) 
\textbf{Proof:}
\[ \Phi_j(S_j^i) - \Phi_{j-1}(S_j^{i-1}) \]
\[ = \sum_{i=1}^{m} \left[ \nabla_i(x^j, T_j^{i-1}) + \frac{p}{k} \left(1 - \frac{p}{k}\right)^{k-j} g_i(T_j^{i-1}) \right] \]
\[ \hspace{1cm} = \sum_{i=1}^{m} \left[ \left(1 - \frac{p}{k}\right)^{k-j} g_i(T_j^i) + c_i(T_j^i) - \left(1 - \frac{p}{k}\right)^{k-(j-1)} g_i(T_j^{i-1}) \right] \]
\[ \hspace{1cm} = \sum_{i=1}^{m} \left[ \left(1 - \frac{p}{k}\right)^{k-j} (g_i(T_j^i) - g_i(T_j^{i-1})) + c_i(T_j^i) - c_i(T_j^{i-1}) \right] \]
\[ \hspace{1cm} = \sum_{i=1}^{m} \left[ \left(1 - \frac{p}{k}\right)^{k-j} g_i(T_j^{i-1}) + \frac{p}{k} \left(1 - \frac{p}{k}\right)^{k-j} g_i(T_j^{i-1}) \right] \]
\[ \hspace{1cm} = \sum_{i=1}^{m} \left[ \frac{p}{k} \left(1 - \frac{p}{k}\right)^{k-j} g_i(T_j^{i-1}) \right] \]
\[ \hspace{1cm} = \sum_{i=1}^{m} \left[ \nabla_i(x^j, T_j^{i-1}) + \frac{p}{k} \left(1 - \frac{p}{k}\right)^{k-j} g_i(T_j^{i-1}) \right] \]  \hspace{1cm} (22) 
where \( x^j = \arg \max_{x \in S_j^i} \sum_{i=1}^{m} \nabla_i(x, T_j^{i-1}). \]

\textbf{Lemma 5:} In each iteration of Algorithm 1, if \( x^j \) is added to \( S_j^{i-1} \), then
\[ \sum_{i=1}^{m} \nabla_i(x^j, T_j^{i-1}) \geq \frac{1}{k} \sum_{i=1}^{m} \left[ \left(1 - \frac{p}{k}\right)^{k-j} (g_i(T_j^i) - p \cdot g_i(T_j^{i-1})) \right] \]
\[ \hspace{1cm} + (1 - O(\xi)) c_i(T_j^i) \]  \hspace{1cm} (23) 
\textbf{Proof:}
\[ k \sum_{i=1}^{m} \nabla_i(x^j, T_j^{i-1}) \]
\[ \geq k \sum_{i=1}^{m} \max\{0, \max_{y:(T_j^i \cup x)|y| \leq k_{c2}} \nabla_i(x, y, T_j^{i-1})\} \]
\[ \geq \sum_{i=1}^{m} |T_i^j| \max\{0, \max_{y:(T_j^i \cup x)|y| \leq k_{c2}} \nabla_i(x, y, T_j^{i-1})\} \]
\[ \geq \sum_{i=1}^{m} |T_i^j| \max_{x \in T_i^j} \nabla_i(x, y, T_j^{i-1}) \]
\[ = \sum_{i=1}^{m} \left[ \frac{p}{k} \left(1 - \frac{p}{k}\right)^{k-j} \nabla_i(x, T_j^{i-1}) + c_i(x) - c_i(y) \right] \]
\[ \geq \sum_{i=1}^{m} \left[ \left(1 - \frac{p}{k}\right)^{k-j} (g_i(T_j^i) - g_i(T_j^{i-1})) + c_i(T_j^i) - c_i(T_j^{i-1}) \right] \]
\[ \geq \sum_{i=1}^{m} \left[ \left(1 - \frac{p}{k}\right)^{k-j} (g_i(T_j^i) - p \cdot g_i(T_j^{i-1})) + (1 - O(\xi)) c_i(T_j^i) \right] \]  \hspace{1cm} (24) 
In above formulation, the first inequality is due to the restricted maximization of \( \nabla_i(x, T_j^{i-1}) \). The second inequality arises from \( |T_j^i| \leq k \) for all \( i \). The third inequality follows since the element \( x \) is the best result chosen by algorithm in current searching process compared to components in \( T_i^j \). The fifth inequality holds owing to the submodularity of sub-functions. And the final one is correct because \( p \geq 1 \). Define \( O(\xi) = c_i(T_j^{i-1}) / c_i(T_j^i) \) as the ratio, in which the difference between \( c_i(T_j^{i-1}) \) and \( c_i(T_j^i) \) is acceptable. Then the proof is over. 

The above lemmas imply the initial approximation ratio of Algorithm 1.

**Theorem 2:** Algorithm 1 returns output $S^k$ and each $T^k_i$, such that

\[
\sum_{i=1}^{m} [g_k(T_i^k) + c_i(T_i^k)] \\
\geq \sum_{i=1}^{m} \left[ \frac{1}{p} (1 - e^{-p}) g_i(T_i^*) + (1 - O(\xi)) c_i(T_i^*) \right] > 0
\]  

(25)

**Proof:** In accordance with the definition of surrogate function $\Phi_j(S^j)$, we have $\Phi_0(S^0) = 0$ and $\Phi_k(S^j) = \sum_{i=1}^{m} [g_k(T_i^k) + c_i(T_i^k)]$. Combining Lemmas 4 and 5, it has $\Phi_j(S^j) - \Phi_{j-1}(S^{j-1}) \geq \frac{1}{k} \sum_{i=1}^{m} [(1 - \frac{p}{k})^{k-j} g_i(T_i^*) + (1 - O(\xi)) c_i(T_i^*)]$. Thus

\[
\sum_{i=1}^{m} [g_k(T_i^k) + c_i(T_i^k)] \\
= \sum_{j=1}^{k} [\Phi_j(S^j) - \Phi_{j-1}(S^{j-1})] \\
\geq \sum_{j=1}^{k} \sum_{i=1}^{m} \left[ (1 - \frac{p}{k})^{k-j} g_i(T_i^*) + (1 - O(\xi)) c_i(T_i^*) \right] \\
= \sum_{i=1}^{m} \left[ \frac{1}{p} (1 - (1 - \frac{p}{k})^k) g_i(T_i^*) + (1 - O(\xi)) c_i(T_i^*) \right]
\]

(26)

In above formulation, the first equality is from definition of $\Phi_j(S^j)$ and $f_i(T_i)$. The third equality arises from geometric series sum. The last one holds due to constant inequality $(1 - \frac{p}{k})^k \leq e^{-p}$. The prove is end.

What is more, the relationship between $c_i(T_i)$ and $f_i(T_i)$ is clarified in Lemma 6 based on separate curvature.

**Lemma 6:** According to the Definition 2, define the separate curvatures for $f(T \cap S_{L1})$ and $f(T \cap S_{L2})$ as $\kappa_{f1} = 1 - \min_{j \in T \cap S_{L1}} f(S_{L1}) / \min_{j \in T \cap S_{L2}} f(S_{L2})$, and $\kappa_{f2} = 1 - \min_{j \in T \cap S_{L2}} f(S_{L2}) / \min_{j \in T \cap S_{L1}} f(S_{L1})$. Then there is

\[
c_i(T_i) = c_i(T_i) + c_{ij}(T_i) \geq (1 - \kappa_{f1} + O(c_2)) f_i(T_i)
\]

(27)

**Proof:** Due to the submodularity of $f_i(T_i)$ and $S_{L2} = S_{L1} \setminus S_{L1}$, we have

\[
f_i(T_i \cap S_{L1}) + f_i(T_i \cap S_{L2}) \geq f_i(T_i)
\]

(28)

Then based upon Lemma 3 and (28), it has

\[
c_{i1}(T_i) + c_{i2}(T_i) \\
= \sum_{x \in T_i \cap S_{L1}} \Delta^f_i(x, S_{L1} \setminus x) + \sum_{x \in T_i \cap S_{L2}} \Delta^f_i(x, S_{L2} \setminus x)
\]

\[
\geq (1 - \kappa_{f1}) f_i(T_i \cap S_{L1}) + (1 - \kappa_{f2}) f_i(T_i \cap S_{L2})
\]

\[
\geq (1 - \kappa_{f1}) f_i(T_i) - f_i(T_i \cap S_{L1}) + (1 - \kappa_{f2}) f_i(T_i \cap S_{L2})
\]

\[
= (1 - \kappa_{f1}) + O(c_2) f_i(T_i)
\]

(29)

where $O(c_2) = (\kappa_{f1} - \kappa_{f2}) f_i(T_i \cap S_{L2}) / f_i(T_i)$.

Finally, the improved approximation ratio of TSSO can be shown in Theorem 3.

**Theorem 3:** For $O(\xi), O(c_2)$ and $\kappa_{f1} \in [0, 1]$, Algorithm 1 returns a set $S^k$ of size $k$ such that

\[
F(S^k) \geq \left[ 1 - \frac{\kappa_{f1}}{p} + \frac{\kappa_{f1}}{p} \right] F(S^*)
\]

(30)

**Proof:** From the Theorem 2 and Lemma 6, we can further obtain that

\[
F(S^k) = \sum_{i=1}^{m} [g_i(T_i^k) + c_i(T_i^k)] \\
\geq \sum_{i=1}^{m} \left[ \frac{1}{p} (1 - e^{-p}) g_i(T_i^*) + (1 - O(\xi)) c_i(T_i^*) \right] \\
= \sum_{i=1}^{m} \left[ \frac{1}{p} (1 - e^{-p}) f_i(T_i^*) + (1 - \frac{1}{p} + \frac{e^{-p}}{p} - O(\xi)) c_i(T_i^*) \right] \\
= \sum_{i=1}^{m} \left[ \frac{1}{p} (1 - e^{-p}) f_i(T_i^*) + (1 - \frac{1}{p} + \frac{e^{-p}}{p} - O(\xi)) c_i(T_i^*) \right] \\
\times (1 - \kappa_{f1} + O(c_2)) f_i(T_i^*)
\]
where $O(\xi, c_2) = (\kappa_{f_1} - 1)O(\xi) + \frac{1}{p} - \frac{e^m}{p} - O(\xi)O(c_2)$ and $S^* = \bigcup_i T_i^\kappa$. The proof is end. \qed

Note that it is acceptable to claim qualitatively that the algorithms with higher approximation guarantee usually have a better objective value than those with lower guarantee [29], [34]. For approximation guarantee in inequality (30), it contains the pure guarantee form $1 - \frac{\kappa_{f_1}e^{-p}}{p} + \frac{\kappa_{f_1}}{p} - \kappa_{f_1}$ similar to the traditional one [35], and an error form $O(\xi, c_2) \leq 0$. From their formulation, as $|S_{1L}|$ increases, the separate curvature $\kappa_{f_1}$ decreases, implying that the pure guarantee will increase. Unfortunately, increasing $|S_{1L}|$ may worsen $O(\xi, c_2)$ determined by both $\kappa_{f_1}$ and $\kappa_{f_2}$. By coordinating the pure guarantee and error, Algorithm 1 can generate a better solution than other TSSO solving methods. Furthermore, under some $\kappa_{f_1}$ and $p$, the value of pure guarantee can be larger than traditional one $1 - e^{-1}$, and its guarantee ratio distribution is illustrated in Fig. 1.

Since there is only one constraint in second stage for TSSO in this research, we can obtain Corollary 1.

Corollary 1: With $p = 1$, the approximation ratio from Theorem 3 can be simplisfied as

$$F(S^k) \geq [1 - \kappa_{f_1}e^{-1} + O'(\xi, c_2)] \cdot F(S^*)$$

Proof: When $p = 1$ is introduced into Theorem 3, it is easy to obtain the result, where $O'(\xi, c_2) = (\kappa_{f_1} - 1)O(\xi) + e^{-1} - O(\xi)O(c_2)$. \qed

Theorem 4: Let $k$ denote the first stage constraint, $k_m = \max_i |M^i|$ the maximum number of unidentical chains in second stage, $D = \max_i |M^i|$ the maximum number of unidentical chains among sub-databases, $d_m = \max_i \max_j |M^i|^j$ the generation maximum, $m$ the number of system state and $n = |S_{1L}|$ the line candidate number. Then the optimization of Algorithm 1 runs in $O(n^2k_mDd_m)$ time.

Proof: The runtime of Algorithm 1 is decided by the searching processes in first and second stages of TSSO and the calculation of sub-function $f(S)$. During the first stage searching process, the algorithm iteratively scans $n$ components at most $k$ times, i.e., $O(\kappa k)$. In the second stage, there are $m$ sub-functions with maximized constraint $k_m$, i.e., $O(mk_m)$. In addition, each sub-function involves at most $O(nDd_m)$ extraction of the failure data. Finally, the total complexity of Algorithm 1 is $O(n^2k_mDd_m)$.

Theorem 4 implies that the dedicated solving algorithm can handle TSSO in polynomial time. In this study, the computation efficiency is compared to the original exponential computation complexity in tackling the identical problem. When dealing with large-scale cases, our method’s scalability is more evident compared to traversal process with at least $O(2^n)$ complexity, particularly in large-scale examples prone to the dimensionality curse [34]. Furthermore, since line serial number is not of the order related to risk mitigation effect, the dichotomy method, which leads to complexity $O(\log(n))$, cannot be applied in this work.

VI. CASE STUDIES

A. Impacts of Environment, Separate Curvature, and $\Pr_t^\text{min}$ on DTR Risk Mitigation

According to the regulation [6], the value of DTR is dependent to ambient environment factors such as wind speed $V$ and surrounding temperature $T_e$. Also, in order to determine the reliable DTR for relay threshold, it is suitable to use the worst environment information among all the critical spans of a transmission line [11]. Thus, the first objective in this part is to investigate the effect of environment factor, represented by the DTR improved factor $\alpha$, on DTR risk mitigation performance.

Experiments are operated in IEEE 39-bus system having $|S_{1L}| = 46$. Specifically, suppose that there are 10 system states, and the constraints in the first and second stages are $k_{c_1} = 8$ and $k_{c_2} = 3, 4, 4, 3, 3, 3, 4, 3, 3$, respectively. Set $\eta = 0.5$ and $Y_{c_{ext}} = 1000$ for submodular sub-function, and each sub-database $M^i$ contains 2000 simulations from failure simulator [22]. And there is $\Pr_t^\text{min} = 0.8$, $\Pr_t^\text{max} = 1.4$, $\Pr_t^\text{min} = 0.1$, $\Pr_t^\text{max} = 1.0$, $\mu = 11.13$ for component failure probability. To assess the fitting rightness of function (5), the error is measured using the squares of error, root mean squared error and R-square error [46], with values of 0.011, 0.026, and 0.996, respectively. These values are deemed acceptable based on the statistical test [46], demonstrating that the applied Sigmoid function can correctly fit the original piecewise linear function in this study. And assume that the default threshold rating is from the condition that $T_e = 40^\circ C$ and $V = 0.61$ m/s [6], implying that $\alpha = 1.0$. Then we select the average worst weather data in different critical line spans from weather databases MRCC [47] and NOAA [48] to calculate $\alpha$.

Table I demonstrates the simulation results. In the weather condition with lower temperature and higher wind speed, $\alpha$ value increases, indicating that more remaining transmission capacity...
can be exploited with the help of DTR. As $\alpha$ rises, the risk mitigation value $F$ becomes larger, raising from $F = 506.181$ ($\alpha = 1.03$) to $F = 1127.598$ ($\alpha = 1.07$), and $F = 1614.099$ ($\alpha = 1.11$). Similarly, the risk value $RiskW$ drops from the original risk $RiskW = 2822.300$ at $\alpha = 1$ to $RiskW = 1543.531$ at $\alpha = 1.05$, and further declines to $RiskW = 836.509$ at $\alpha = 1.11$. These indexes exhibit that the risk mitigation effect will become larger with a feasible weather condition, having a larger $\alpha$ value. What’s more, there is Braess paradox happening in risk mitigation, as shown by $BPI \in (160, 230)$ with DTR placement in system. It reminds us that it is better to apply the DTR to mitigate the failure risk meanwhile retaining the Braess paradox impact at an acceptable level.

On the other hand, the separate sets $S_{L1}$ and $S_{L2}$ can also impact the curvature value, hence affecting the approximation guarantee. Based on in inequalities (32), the roles of $\kappa_{f1}$ and $O'(\xi, c_2)$ are somehow competitive in guarantee improvement, so we need select $S_{L1}$ carefully to achieve a better approximation guarantee considering both cardinality and its components. The settings are the same as above besides fixed $\alpha = 1.05$, and Table II shows the influence of separate curvature on risk mitigation.

We can see that when separate number $|S_{L1}|$ decreases, the pure guarantee grows, while there is a worser $O'(\xi, c_2)$ resulting in a poor guarantee. For instance, even if $1 - \kappa_{f1} e^{-1} = 0.803$ when $|S_{L1}| = 15$, the worser error $O'(\xi, c_2) = -0.384$ yields a lower guarantee value 0.419. As $|S_{L1}|$ rises, both pure guarantee and $O'(\xi, c_2)$ have downward trends, but $O'(\xi, c_2)$ decrease more. Then, we can find a key balance between pure guarantee and error form to obtain a larger guarantee. In this experiment, when $|S_{L1}| = 36$, the guarantee is 0.674, which is more than the traditional optimal guarantee $1 - e^{-1}$ [35], and its objective value $F = 940.643$ is nearly the best compared to other $|S_{L1}|$. Note that the objective value $F = 941.036$ at $|S_{L1}| = 38$ is close to $F$ value at $|S_{L1}| = 36$, but its guarantee is 0.589 lower than guarantee at $|S_{L1}| = 36$, suggesting qualitatively that under some extreme conditions, like poor or missing state data, the worst value at $|S_{L1}| = 38$ will be less than the worst value at $|S_{L1}| = 36$.

Furthermore, we investigate the impact of parameter $Pr_{e}^{\max}$ on model performance. In related analyses [1, 43], $Pr_{e}^{\max}$ is not universally constant and is frequently set within the range of 0 to 0.2. Therefore, we set $Pr_{e}^{\min}$ as 0.05, 0.10 and 0.15, respectively. For concise explanation, only one system state is selected to conduct the model, and the result is given in Table IV. It indicates that as $Pr_{e}^{\min}$ increases within a feasible region, our method can target the critical line, i.e., line 15 in this case, while the remaining lines adjust as auxiliary parts to varying $Pr_{e}^{\min}$ values. Meanwhile, sub-objective function rises from 84.328 to 116.637, since a higher $Pr_{e}^{\min}$ implies that the power lines are more susceptible to being cut, thereby elevating cascading failure risk required to be mitigated. Conversely, $Pr_{e}^{\max}$ is typically established proximal or equal to 1 [1, 43]. The related experiment indicates that within the allowable change of $Pr_{e}^{\max}$, the proposed model can also identify the critical line effectively, which are not listed here to save space.

In summary, the DTR can indeed assist in mitigating the risk. Regarding the weather factor, DTR can release more remaining transmission capacity when weather condition has a higher $\alpha$ value. The separate curvature, on the other hand, can influence the DTR placement and operation scheme as well. Careful selection of separate sets can reach a suitable balance between pure guarantee and error form, leading to a superior performance guarantee. When $Pr_{e}^{\min}$ is within a proper region, our method can also focus on critical lines.

### B. Performance Comparison With One-Stage Optimization

The goal of this section is to compare the performances of one-stage and two-stage DTR models. The comparison experiment is conducted in IEEE 39-bus system. Except for a fixed $\alpha = 1.05$, the settings of one-stage and two-stage model are the same as above. But the difference is that one-stage model only contains constraint $k = 5$, meaning that all the placed DTR in one-stage model will operate constantly regardless of the system state.

The result is shown in Table III. In one-stage model, the selected lines are 3, 6, 9, 16, 27, and that in two-stage model are 3, 9, 11, 16, 19, 23, 27, 45. As can be seen, two-stage model enables DTR operating flexibly with redundant placement, resulting in a better effect. For instance, even though two-stage model has fewer DTR operation in state 3, $f_{\text{two}}^{\max} = 1645.408$ is greater than $f_{\text{one}}^{\max} = 771.740$ in one-stage model. Similar results also occur in states 2, 4, 5, 6, 8, 9, 10. Also from $BPI$ value, the fixed DTR operation in one-stage model can cause more Braess paradox. In state 2, for example, $BPI_{\text{one}}^{\max} = 379.013$ is larger than $BPI_{\text{two}}^{\max} = 29.772$, meaning that one-stage model inversely brings more extra risk into system. For the whole performance, mean value $F_{\text{one}}^{\max} = 776.223$ is lower than $F_{\text{two}}^{\max} = 836.509$.

### Table I

| Average worst weather | $\alpha$ | $F$ | $RiskW$ | $BPI$ |
|----------------------|---------|-----|---------|-------|
| $T_e = 40$, $V = 0.61$ | $1$     | $0.000$ | $2822.300$ | $0.000$ |
| $T_e = 38.5$, $V = 1.29$ | $1.03$ | $506.181$ | $1944.031$ | $161.548$ |
| $T_e = 37.2$, $V = 1.76$ | $1.05$ | $940.643$ | $1543.531$ | $229.475$ |
| $T_e = 36.8$, $V = 1.79$ | $1.07$ | $1127.598$ | $1339.127$ | $194.576$ |
| $T_e = 36.3$, $V = 1.85$ | $1.09$ | $1395.570$ | $1083.234$ | $218.735$ |
| $T_e = 35.7$, $V = 1.94$ | $1.11$ | $1614.099$ | $836.569$ | $162.462$ |

### Table II

| $|S_{L1}|$ | $F$ | $\kappa_{f1}$ | Pure Guarantee | $O'(\xi, c_2)$ | Guarantee |
|----------|-----|---------------|----------------|----------------|-----------|
| 11       | 750.14 | 0.510 | 0.812 | -0.314 | 0.498 |
| 15       | 827.283 | 0.556 | 0.803 | -0.384 | 0.419 |
| 18       | 771.792 | 0.665 | 0.755 | -0.213 | 0.542 |
| 21       | 792.483 | 0.674 | 0.742 | -0.489 | 0.262 |
| 24       | 772.430 | 0.687 | 0.747 | -0.588 | 0.159 |
| 27       | 810.735 | 0.783 | 0.712 | -0.111 | 0.601 |
| 30       | 846.226 | 0.697 | 0.744 | -0.199 | 0.545 |
| 33       | 880.912 | 0.837 | 0.692 | -0.139 | 0.554 |
| 36       | 940.643 | 0.765 | 0.719 | -0.045 | 0.674 |
| 39       | 941.036 | 0.675 | 0.792 | -0.163 | 0.589 |
| 41       | 901.453 | 0.735 | 0.730 | -0.166 | 0.564 |
| 45       | 889.956 | 0.662 | 0.755 | -0.203 | 0.552|

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
TABLE III
Performance Comparison Between One-Stage and Two-Stage Models

| System state | One-stage model | Two-stage model |
|--------------|----------------|-----------------|
|              | f       | BPI   | f       | BPI   | T     |
| 1            | 151.809 | 71.268 | 105.301 | 28.972 | 11.1623 |
| 2            | 340.600 | 379.013 | 454.516 | 29.772 | 9.161923 |
| 3            | 771.740 | 219.562 | 1347.633 | 106.127 | 3.9111623 |
| 4            | 792.133 | 28.022 | 922.339 | 28.577 | 3.923 |
| 5            | 1088.631 | 0.000 | 1378.213 | 0.000 | 9.1619 |
| 6            | 868.428 | 524.382 | 981.752 | 410.267 | 9.1927 |
| 7            | 1320.952 | 257.718 | 1263.288 | 264.595 | 3.927 |
| 8            | 258.908 | 561.247 | 341.233 | 504.690 | 3.111623 |
| 9            | 1313.603 | 488.344 | 1448.525 | 396.493 | 11.1927 |
| 10           | 855.424 | 290.773 | 1161.833 | 525.254 | 3.1645 |
| Mean         | 776.223 | 281.983 | 940.643 | 229.475 | - |

TABLE IV
P2min Effect on DTR Risk Mitigation

| P2min | f | T   |
|-------|---|-----|
| 0.05  | 84.328 | 6, 15, 38 |
| 0.10  | 106.402 | 6, 12, 15 |
| 0.15  | 118.637 | 12, 15, 30 |

940.643, while mean value \( BPI_{\text{one}} = 281.983 \) is worse than \( BPI_{\text{two}} = 229.475 \). These results demonstrate that two-stage model can properly schedule fewer DTR to achieve a higher risk mitigation effect. Since there is only one stage optimization, one-stage model needs to balance the risk mitigation effect in each system state, thus the final DTR operation is not the best scheme for all states.

Moreover, we contrast the residual service life of DTR in two models. Assuming that each DTR can work around the clock for 6 years, the residual service life ratio for each DTR can be calculated, shown in Table V. From the result, the residual service life ratio for each DTR in one-stage model is only 67% after 2 years and 33% after 4 years. However, after 2 years and 4 years, the residual service life ratios for each DTR in two-stage model are all larger than 80% and 60%, respectively, indicating that the flexible DTR operation can extend the DTR service life.

Furthermore, we also investigate the two-stage model’s flexibility to potential load increase. In addition to the one-stage and two-stage model defined before, we include a flexible two-stage model that adds an extra DTR in each system state. Setting load ratios \( \text{LoadR} \) are 1.02 and 1.06 times of the original, respectively. The result in Table VI reveals that, as system load grows the two-stage model performs better in risk mitigation than one-stage model, implied by the fact that \( F^{\text{two}} = 934.957 \) is larger than \( F^{\text{one}} = 805.863 \) at \( \text{LoadR} = 1.02 \), and by the similar findings at \( \text{LoadR} = 1.06 \). In comparison to one-stage and original two-stage models, flexibility of redundant DTR allows the flexible two-stage model to achieve the highest risk mitigation effect, demonstrated by \( F^{\text{flex}} = 982.375 \) at \( \text{LoadR} = 1.02 \) and \( F^{\text{flex}} = 991.318 \) at \( \text{LoadR} = 1.06 \). It indicates that the two-stage model has scalability potential to handle the load growth and other emergencies, achieving a more flexible and effective performance of risk mitigation.

To sum up, two-stage method considers the suitable redundant DTR placement, not only allowing operators to build up the certain DTR operation schedule for a specific system state, but also prolonging DTR service life. It can also function better when faced with load growth and other situations, showing advantages compared to one-stage model.

C. Performance Comparison With Different Strategies

This section compares the performance of our proposed strategy, SCG, with that of other two-stage strategies. Through the literature review, we take 3 categories as the comparable forms. The first category is known as the index-based strategy, which selects DTR lines based upon some traditional assessment indexes [22], [45], including:

- **Random line strategy (RL):** Select lines randomly to place DTR.
- **Failure rate strategy (FR):** Sort lines in decreasing order from failure number in database, and place DTR in the highest ranked lines.
- **Largest power flow strategy (LPF):** Sort lines in decreasing order from initial power flow, and place DTR in the highest ranked lines.
- **Largest hidden failure strategy (LHF):** Using \( N - 1 \) security test, sort lines in decreasing order from hidden failure probabilities, and place DTR in the highest ranked lines.

The second category, known as greed-based strategy, includes the core greedy algorithm but does so in a variant way [28], [42]. It includes:

- **Greedy sum strategy (GS):** Select \( k \) lines with the highest objective function \( F \) iteratively, then choose the matching constrained lines for each system state from these \( k \) lines.
- **Modular approximation strategy (MA):** Approximate the sub-function \( f \) is modular, then select \( k \) lines with the highest objective function \( F \) iteratively.
- **Local search strategy (LS):** Select a line with the highest objective function \( F \) and arbitrary \( k - 1 \) lines, then choose an unselected line to replace the selective line constantly. If the new objective value is larger than the previous, adopt this replacement set. Note that it can achieve a \( \frac{1}{p+1} \) guarantee where \( p \) is second stage constraint’s number.

The last category is called surrogate-based strategy, which designs several surrogate functions to judge DTR line selection, including:

- **Replacement greedy strategy (RG):** Select \( k \) lines with the highest surrogate function \( f(S_A \cup s) - f(S_A) \) iteratively [29]. And it can achieve a \( \frac{1}{p+1} \) guarantee.
- **General P-matroid greedy strategy (GPG):** Select \( k \) lines with the highest surrogate function \( (1 - \frac{1}{k})^{k-i} g(S_A) + c(S_A) \) iteratively, where \( k \) and \( i \) denote the first stage constraint and searching step [33]. And it can achieve the same guarantee as RG.
- **General curvature-based greedy strategy (GCC):** Select \( k \) lines with the highest surrogate function \( (1 - \frac{1}{k})^{k-i} g(S_A) + (1 - \frac{1}{k})^{k-i} c(S_A) \) iteratively [32]. And it can achieve a \( \frac{1 - \kappa}{\kappa} L(1 - e^{-p}) + \frac{\kappa^{p+1}}{p+1} (1 - e^{-(p+1)}) \) guarantee where \( \kappa \) is curvature.
The experiment settings are the same as in Section VI.B and the result is given in Table VII. Even though LHF has a higher objective value $F^{LHF} = 743.385$, other index-based strategies perform worse than other categories since the indexes in these strategies are not strongly correlated to the risk and capture inadequate features in the cascading failure process. For greed-based strategies, they all exhibit relatively better performance, as evidenced by $F^{GS} = 872.411$, $F^{MA} = 856.520$ and $F^{LS} = 886.701$. For surrogate-based strategy, RG outperforms greed-based strategies as $F^{RG} = 890.748$, while the other two perform badly. The reason is that, while our analyzed problem only contains a single cardinality constraint, GPG and GCG is designed for general use and perform better in complex constraints [32], [33], as opposed to the index-based strategy and greed-based strategy which are only intended for use with single constraint type. When compared to others, our strategy SCG has the best performance, as evidenced by $F^{SCG} = 940.643$ with adaptive regulation of separate curvature to theoretically boost the approximation guarantee. What’s more, the $BPI$ value does not equal zero in all strategies, highlighting that the Braess paradox must be considered when designing a risk mitigation scheme. In Table VII, we also compare their computed time and complexity. For RL, FR, LPF, LHF and MA, their complexities are first-order of $n$, which means they need less calculation time but produce inferior output. For other strategies, they have complexities with multi-order of $n$, responsible for longer calculated time, while our method can obtain a superior result in a similar computed time.

Moreover, Fig. 1 compares the guarantee distributions of LS, RG, GPG, GCG and SCG, neglecting their error forms. When curvature or constraint’s number lowers, the guarantees of all strategies improve. With the help of separate curvature, the proposed strategy SCG can reach a higher guarantee than others under the same conditions. It is also notable that SCG’s guarantee may be raised close to 1.0, indicating its greater improvement potential.

In conclusion, the proposed strategy SCG outperforms other strategies in TSSO-based risk mitigation. With the aid of separate curvature, SCG can improve the objective function value adaptively, seen from the highest approximation guarantee. SCG is also suitable to multiple constraint types owing to the consideration of constraint’s number.

### TABLE VII

| Strategy | $F$ | Time (min) | Complexity |
|----------|-----|------------|-------------|
| RL       | 488.998 | 140.033 | $O(nk_mD_m)$ |
| FR       | 595.917 | 271.307 | $O(nk_mD_m)$ |
| LPF      | 31.496  | 3.942   | $O(nk_mD_m)$ |
| LHF      | 743.385 | 243.807 | $O(nk_mD_m)$ |
| GS       | 872.411 | 293.569 | $O((n^2 + n k_m) k_m D_m)$ |
| MA       | 856.520 | 263.058 | $O(n k_m k_m D_m)$ |
| LS       | 886.701 | 189.308 | $O(n k_m k_m D_m)$ |
| RG       | 890.748 | 201.953 | $O(n k_m k_m D_m)$ |
| GPG      | 497.852 | 136.508 | $O(n k_m k_m D_m)$ |
| GCG      | 465.902 | 175.473 | $O(n k_m k_m D_m)$ |
| SCG (ours) | 940.643 | 229.475 | $O(n k_m k_m D_m)$ |
| Nu. DTR  | 0.000  | 0.000   | -            |

### TABLE VIII

| Case  | $n$ | $k$ | $k_m$ | $m$ | $D$ | $d_m$ | Time (h) |
|-------|-----|-----|-------|-----|-----|-------|----------|
| 39-bus| 46  | 8   | 4     | 10  | 1364| 5     | 3.297    |
| 118-bus| 186 | 14  | 8     | 12  | 1974| 9     | 459.623  |

D. Computational Efficiency Analysis Related to Case Scales

To investigate the computational efficiency related to case scales in proposed algorithm, we conduct the experiment in IEEE 118-bus system with $|S_u| = 186$. Set $k_{c_1} = 14$ and $k_{c_2} = [6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]$, respectively, and the rest of parameters are the same as in 39-bus system. Based on simulation, the selected lines are 8, 9, 19, 31, 32, 38, 94, 96, 97, 98, 117, 126, 183, 185 and 186, yielding $F = 1758.819$. Also, the relevant complexity parameters and computation times of IEEE 39-bus and 118-bus systems are included in Table VIII. We can see that their computation time ratio is nearly $4.299 \approx 139$ and the complexity ratio is $186^2 + 14 + 8 + 12 + 1974 + 9 \approx 179$, which are consistent within the tolerance of multiple 2. When the ratio tolerance is within one order of magnitude, it is acceptable to utilize the complexity to approximate the computation time. Thanks to the coding acceleration approach we used, the computation time ratio is less than the complexity ratio in this comparison experiment. However, as case scale grows, the computation burden is increasingly affected by hardware and software conditions, such as memory size and floating point limit [43]. Thus, as case scale grows, the computation time ratio will exceed the complexity ratio, necessitating the usage of more advanced computation environment.

### VII. Conclusion

This work investigated a DTR optimization problem for risk mitigation considering placement and operation schedule in...
terms of service life and Braess paradox. The proposed model is based on two-stage submodular optimization, with the first stage optimizing DTR placement with proper redundancy and the second stage designing a flexible operation schedule for each system state. First, we established a sub-function of CF risk mitigation. Then using the Markov and submodular properties in sub-function, the submodularity of total objective function of TSSO is proven for the first time. Consequently, a state-of-the-art solving algorithm based on separate curvature is devised, which can provide a better provable approximation guarantee than the current researches and obtain the solution in polynomial time. Case results demonstrate the impacts of environment, separate curvature, and $Pr_{min}$ on risk mitigation. Also, it validates that the suggested model outperforms the one-stage model both in risk mitigation and service life extension, and excels other two-stage strategies in performance.

For future work, we regard these directions as having exploration value. 1) Examine the influence of unexplored factors described in Section II.A on DTR placement plan, like self-powered technique and outside environment. 2) Explore the DTR sensor maintenance planning in light of DTR failure. 3) Look at how to choose $S_{L1}$ and $S_{L2}$ wisely for separate curvature. 4) Investigate the reducible effect on DTR-based risk mitigation. Furthermore, we believe that these possible applications of TSSO merit further investigation. 1) Traditional electrical applications: PMU setting for system observation, charging station placement and traffic schedule, meter coverage for load estimation, and so on. 2) Current trends in smart grid: electric privacy, data summarization, recommendation system, representative scenario selection, and others. 3) Further application: electrical problems that use the stochastic combinatorial optimization, meta-learning, or dictionary learning as mathematical model.

APPENDIX

Proof of the Submodularity of TSSO with Markov property: To simplify the notation, we apply set function $Y_i(T_i) = Cons_i - y_i(T_i)$ to represent the sub-function $f_i(T_i)$, where $Cons_i$ is a constant, $y_i(T_i)$ is a Markov-based decreasing function having $y_i(a ∪ b) = y_i(a) · y_i(b)$ where $a ∩ b = 0$, and $i$ is the sub-function index. Note that $Y_i(T_i) ≥ 0$, and $y_i(T_i)$ can be treated as a general form of risk mitigation $RiskW_i(T_i)$ (the smaller the better). And $η · BPI$ does not affect the submodularity due to Lemma 2. Define $G_i(S) = \{ T \subseteq S : T = \arg\max_{S \subseteq T} f_i(T) \}$ as the optimal solution for $y_i(T_i)$ in the feasible region $S$, and $S_L$ the ground set. Then TSSO can be formulated by the weight sum of $Y_i(G_i(S))$, which chooses the feasible region $S$ for sub-functions $Y_i(T_i)$ to get the optimal results. Therefore, the goal of checking the submodularity of TSSO under Markov property is to show whether the $Y_i(G_i(S))$ is submodular. For simplicity, we can analyze the submodularity of $y_i(G_i(S))$, which allows us to easily infer the $Y_i(G_i(S))$ properties.

In the beginning, Lemma A1 is introduced to clarify the inequality properties of $y_i(G_i(S))$.

Lemma A1: For Markov-based function $y_i(T_i)$, if there are subsets $S_A, S_B, v$ satisfying $S_A \subseteq S_B \subseteq S_L, v \in S_L \setminus S_B$, $y_i(T_i)$ has these inequalities in combinatorial optimization: $y_i(G_i(S_A)) ≥ y_i(G_i(S_B ∪ v)), y_i(G_i(S_B)) ≥ y_i(G_i(S_B ∪ v))$, $y_i(G_i(S_A)) ≥ y_i(G_i(S_B))$ and $y_i(G_i(S_A ∪ v)) ≥ y_i(G_i(S_B ∪ v))$. Note that for $y_i(T_i)$, the smaller the better, and $Y_i(T_i) ≥ v$ has invasive inequalities from $y_i(T_i)$.

Proof: For $y_i(G_i(S_A)) ≥ y_i(G_i(S_A ∪ v))$, based on the definition of $G_i(S)$, the adding of $v$ to $S_A$ constructs a new feasible region, including 2 scenarios.

1) $G_i(S_A ∪ v)$ cannot diminish the value $y_i(G_i(S_A))$ by adding $v$ to $G_i(S_A)$ or replacing certain components in $G_i(S_A)$ with $v$, then it has $G_i(S_A ∪ v) = G_i(S_A)$ such that $y_i(G_i(S_A ∪ v)) = y_i(G_i(S_A)).$

2) $G_i(S_A ∪ v)$ reduces the value $y_i(G_i(S_A))$, meaning that $y_i(v) < y_i(G_i(S_A ∪ v)), so that it has $y_i(G_i(S_A ∪ v)) < y_i(G_i(S_A)).$

Thus, there is $y_i(G_i(S_A)) ≥ y_i(G_i(S_A ∪ v)).$ Other inequalities can be obtained by the similar process as described above.

Then some operations are defined for $G_i(S)_i$.

When $G_i(S_A ∪ v) = G_i(S_A)$ and there is a set $S_B$ where $S_B \neq S_A$, we define Absorb operation (Abs-ope): $G_i(S_B ∪ S_A ∪ v) = G_i(S_B ∪ S_A).$

When $G_i(S_A ∪ v) = G_i(S_A)$, we define three operations in set $G_i(S_A ∪ v)$:

- Addition operation (Add-ope): $G_i(S_A ∪ v) = G_i(S_A) ∪ v$.
- Exchange operation (Exc-ope): $G_i(S_A ∪ v) = G_i(S_A \setminus a) ∪ v$ where $a$ is the element in $G_i(S_A)$ replaced by $v$ and $|v| = 1$.
- Reduction operation (Red-ope): $G_i(S_A ∪ v) = G_i(S_A \setminus a) ∧ v$ where $a$ are the elements in $G_i(S_A)$ replaced by $v$ and $|v| > 1$.

After that, there are some important set relationships of $G_i(S)$ shown in Lemmas A2, A3 and A4.

Lemma A2: For subsets $S_A, S_B, v$ satisfying $S_A \subseteq S_B \subseteq S_L, v \in S_L \setminus S_B, G_i(S_A ∪ v) = G_i(S_A)$ leads to $G_i(S_B ∪ v) = G_i(S_B).$

Proof: Since $G_i(S_A ∪ v) = G_i(S_A)$, it signifies that the adding of component $v$ to current feasible region $S_A$ cannot decrease $y_i(G_i(S_A))$. Due to $y_i(G_i(S_A)) ≥ y_i(G_i(S_B))$, we know that the improvement $y_i(G_i(S_B ∪ v)) − y_i(G_i(S_A))$ is from components in set $S_B \setminus S_A$, not in $v$. Based upon Abs-ope, there is

\begin{align*}
G_i(S_B ∪ v) &= G_i(S_B ∪ v ∪ S_A) \\
&= G_i(S_B ∪ S_A ∪ v) \\
&= G_i(S_B ∪ S_A) \\
&= G_i(S_B) \quad (33)
\end{align*}

Lemma A3: For subsets $S_A, S_B, v$ satisfying $S_A \subseteq S_B \subseteq S_L, v \in S_L \setminus S_B, G_i(S_B ∪ v) \neq G_i(S_B)$ leads to $G_i(S_A ∪ v) \neq G_i(S_A)$.

Proof: Since $G_i(S_B ∪ v) \neq G_i(S_B)$, we can obtain that $y_i(G_i(S_B ∪ v)) < y_i(G_i(S_B))$, showing that the reduced value is contributed from component $v$. Due to $S_A \subseteq S_B$, the adding
of \(v\) can also reduce the value of \(y_i(G_i(S_A))\), i.e., \(y_i(G_i(S_A)) > y_i(G_i(S_A \cup v))\). Then, we can obtain \(G_i(S_A \cup v) \neq G_i(S_A)\). □

Lemma A4: For subsets \(S_A, S_B, v\) satisfying \(S_A \subseteq S_B \subseteq S_L, v \in S_L \setminus S_B\), it exists that
1) If \(G_i(S_B \cup v)\) executes Add-ope from \(G_i(S_B)\), \(G_i(S_A \cup v)\) will perform Add-ope from \(G_i(S_A)\).
2) If \(G_i(S_B \cup v)\) executes Exc-ope from \(G_i(S_B)\), \(G_i(S_A \cup v)\) will perform Add-ope or Exc-ope from \(G_i(S_A)\).
3) If \(G_i(S_B \cup v)\) executes Red-ope from \(G_i(S_B)\), \(G_i(S_A \cup v)\) will perform Add-ope, Exc-ope or Red-ope from \(G_i(S_A)\).

Proof: These statements are proven in the perspectives of cost performance and residual budget. From the definition, \(G_i(S)\) selects components from \(S\) that are close to the budget threshold or have greater cost performance, to optimize the objective function. Specifically, the cost nearing to the budget threshold indicates that there is still enough residual budget to apply more components, and the limitation is the insufficient feasible region. On the other hand, in the situation of greater cost performance, the improvement limit is component efficiency, and the budget does not allow for the addition of another component.

The above analysis suggests that execution of Add-ope means that there is still enough residual budget to add \(v\) to \(G_i(S)\) to optimize the objective value. And the executions of Exc-ope and Red-ope demonstrate that the residual budget is insufficient to add more component to the selected set, and the cost performance of \(v\) is larger than some components in \(G_i(S)\).

Based upon Lemmas A1 and A3, \(G_i(S_B)\) has either less residual budget or higher cost efficiency compared with \(G_i(S_A)\). Thus, when \(G_i(S_B \cup v)\) performs Add-ope from \(G_i(S_B)\), \(G_i(S_A \cup v)\) must have a lower budget to execute Add-ope in statement 1). Inversely, if \(G_i(S_A \cup v)\) takes Exc-ope and Red-ope in this situation, it means that \(G_i(S_A)\) has a smaller budget than \(G_i(S_B)\), which violates the formulation of \(G_i(S)\).

Similarly, when \(G_i(S_B \cup v)\) executes Exc-ope from \(G_i(S_B)\), i.e., \(G_i(S_B \cup v)\) has used the cost of a single component in \(G_i(S_B)\) to replace with \(v\), \(G_i(S_A \cup v)\) with larger budget will not replace more components in \(G_i(S_A)\) with \(v\), such that \(G_i(S_A \cup v)\) will execute Add-ope or Exc-ope in statement 2). And the situation of the statement 3) is similar to the analysis of statement 2).

After obtaining Lemmas A1~ A4, the submodularity of TSSO in Markov setting can be shown.

Theorem A1: Under the Markov and submodular features of sub-function, the total objective function of TSSO is submodular.

Proof: Suppose there are subsets \(S_A, S_B, v\) satisfying \(S_A \subseteq S_B \subseteq S_L, v \in S_L \setminus S_B\). \(y_i(T_i)\) is a Markov-based decreasing function having \(y_i(a \cup b) = y_i(a) \cdot y_i(b)\), and \(y_i(G_i(S))\) can be represented as the optimal value of \(y_i(T_i)\) under feasible region \(S\). Due to the relationship of \(S_A, S_B\) and \(v\), there are \(y_i(G_i(S_A)) \geq y_i(G_i(S_B))\) and \(y_i(G_i(S)) \geq y_i(G_i(S_U \cup v))\). Also define the general sampling weight as \(W\) [44], e.g., \(\hat{W}_{B-A} = y_i(G_i(S_B))/y_i(G_i(S_A))\). And it is easy to deduce that \(\hat{W}_{B-A} \leq 1, \hat{W}_{A-U-A} \leq 1\) and so on. Note that the Markov property of sub-function allows for the application of sampling weight technique on \(y_i(G_i(S))\) [44].

Then, there are 3 situations for the searching of the optimal value in TSSO.

1) When \(v\) is added to the current feasible region, it occurs that \(G_i(S_A \cup v) = G_i(S_A)\) and \(G_i(S_B \cup v) = G_i(S_B)\), as stated in Lemma A2. This suggests that the addition of \(v\) cannot further optimize the objective function in the current feasible regions, both in \(S_A\) and \(S_B\). Based on Lemma A1, it can be shown that \(\hat{W}_{B-A} = 1\). Then there is
\[
y_i(G_i(S_A)) - y_i(G_i(S_A \cup v)) \geq \hat{W}_{B-A} \cdot y_i(G_i(S_A)) - \hat{W}_{B-A \cup v} \cdot y_i(G_i(S_A \cup v)) = y_i(G_i(S_B)) - y_i(G_i(S_B \cup v)) \tag{34}
\]

2) When \(v\) is added to the current feasible region, it occurs that \(G_i(S_A \cup v) \neq G_i(S_A)\) and \(G_i(S_B \cup v) = G_i(S_B)\). It implies that \(G_i(S_A)\) has a larger budget to obtain \(v\) than \(G_i(S_B)\), or \(v\) has a higher cost performance for \(G_i(S_A)\) than \(G_i(S_B)\). In this situation, \(y_i(G_i(S_A \cup v))\) will be less than \(y_i(G_i(S_A))\) such that \(\hat{W}_{B-A \cup v} > \hat{W}_{B-A}\). Then there is
\[
y_i(G_i(S_A)) - y_i(G_i(S_A \cup v)) \geq \hat{W}_{B-A} \cdot y_i(G_i(S_A)) - \hat{W}_{B-A \cup v} \cdot y_i(G_i(S_A \cup v)) \geq y_i(G_i(S_B)) - y_i(G_i(S_B \cup v)) \tag{35}
\]

3) When \(v\) is added to the current feasible region, it occurs that \(G_i(S_A \cup v) \neq G_i(S_A)\) and \(G_i(S_B \cup v) \neq G_i(S_B)\), as stated in Lemma A3. This situation is complicated, which can be categorized into 3 cases indicated in Lemma A4.

Case (1): When \(G_i(S_B \cup v)\) executes Add-ope from \(G_i(S_B)\), \(G_i(S_A \cup v)\) will perform Add-ope from \(G_i(S_A)\). Due to the Markov property of \(y_i(T_i)\), we can deduce that
\[
y_i(G_i(S_A \cup v)) = y_i(G_i(S_A)) \cdot y_i(v) \quad \text{and} \quad y_i(G_i(S_B \cup v)) = y_i(G_i(S_B)) \cdot y_i(v).
\]
Then there is
\[
y_i(G_i(S_B \cup v)) = y_i(G_i(S_B)) \cdot y_i(v) \geq y_i(G_i(S_A \cup v)) = y_i(G_i(S_A)) \cdot y_i(v)
\]

Case (2): When \(G_i(S_B \cup v)\) executes Exc-ope from \(G_i(S_B)\), \(G_i(S_A \cup v)\) will perform Add-ope or Exc-ope from \(G_i(S_A)\). In this case, \(y_i(G_i(S_B \cup v)) = \frac{y_i(G_i(S_B))}{y_i(b)}\) where component \(b\) in \(G_i(S_B)\) is replaced by \(v\), implying that \(y_i(v) < y_i(b)\). On the other hand, there is \(y_i(G_i(S_A \cup v)) = y_i(G_i(S_A)) \cdot y_i(v)\) for Add-ope, or \(y_i(G_i(S_A \cup v)) = y_i(G_i(S_A \cup v))\) for Exc-ope, where component \(a\) in \(G_i(S_A)\) is replaced by \(v\). Due to the submodularity, if \(G_i(S_A \cup v)\) executes Add-ope, the reduction effect from component \(a\) is equal to or less than that from \(b\), i.e., \(y_i(a) \leq y_i(b)\).
The reason is that the worse cost-effective component in $G_i(S_B)$ can have equal or greater effect on reduction compared to the worse cost-effective component in $G_i(S_A)$ due to $S_A \subseteq S_B$. Then it has

$$\frac{y_i(G_i(S_B)) \cdot y_i(v)}{y_i(G_i(S_B)) \cdot y_i(b)} \geq \frac{y_i(G_i(S_A)) \cdot y_i(v)}{y_i(G_i(S_A)) \cdot y_i(a)}$$

$$\Rightarrow \frac{y_i(G_i(S_B \cup v))}{y_i(G_i(S_B))} \geq \frac{y_i(G_i(S_A \cup v))}{y_i(G_i(S_A))}$$

$$\Rightarrow \hat{W}_{B \cup v} \geq \hat{W}_B$$

(37)

**Case (3):** When $G_i(S_B \cup v)$ executes Red-ope from $G_i(S_B)$, $G_i(S_B \cup v)$ will perform Add-ope, Exc-ope or Red-ope from $G_i(S_A)$. In this case, $y_i(G_i(S_B \cup v)) = \frac{S_i(G_i(S_B)) \cdot y_i(v)}{\prod_j y_i(b_j)}$ where a class of components $b_j$ in $G_i(S_B)$ is replaced by $v$. Since the cost-performance of $G_i(S_A)$ must equal to or less than that of $G_i(S_B)$, we can deduce similarly that $\prod_j y_i(b_j)$ will equal to or less than the value resulting from addition or replacements in $G_i(S_A)$. Based on the analysis in Case (2), there is $W_{B \cup v} \geq W_B$.

As a result of these cases, it can be concluded that when $G_i(S_A \cup v) \neq G_i(S_A)$ and $G_i(S_B \cup v) \neq G_i(S_B)$, there is $W_{B \cup v} \geq W_B$, resulting similarly in $y_i(G_i(S_A)) - y_i(G_i(S_B \cup v)) \geq y_i(G_i(S_B)) - y_i(G_i(S_B \cup v))$ in three situations of TSSO with Markov property. Transforming $y_i(G_i(S))$ to $Y_i(G_i(S))$, it has $Y_i(G_i(S_A \cup v)) - Y_i(G_i(S_B \cup v)) \geq Y_i(G_i(S_B \cup v)) - Y_i(G_i(S_B))$, satisfying Definition 1, i.e., $Y_i(G_i(S))$ is submodular. According to TSSO formulation $\mathbb{E}[Y_i(G_i(S))]$ and Lemma 1, it shows that the total objective function of TSSO is submodular when its sub-functions contain Markov and submodular features.

According to the Theorem A1, there are some observations.

**Observation A1:** In general, the non-submodularity of total objective function in TSSO derives from the mechanism that the non-additivity of marginal benefit is related with components that have been selected, rather than the scale of selective set.

**Proof:** Suppose that there are subsets $S_A, S_B, v$ satisfying $S_A \subseteq S_B \subseteq S_L$, $v \in S_L \backslash S_B$, and $Y_i(G_i(S))$ is nondecreasing total function of TSSO. Imagining this situation that the benefit of component $v$ cannot be additive in $Y_i(G_i(S_A))$, i.e., $Y_i(G_i(S_A \cup v)) = Y_i(G_i(S_A))$, due to the constraints. And there is $Y_i(G_i(S_B \cup v)) > Y_i(G_i(S_B))$ since the previous set in $S_B$ can recombine with $v$ to optimize the total objective value. Then there is $Y_i(G_i(S_A \cup v)) - Y_i(G_i(S_B \cup v)) \leq Y_i(G_i(S_B \cup v)) - Y_i(G_i(S_B))$ such that the total function of TSSO is not submodular. The reason for this situation is that the non-additivity in TSSO results in that increasing the feasible region in first stage cannot guarantee the expansion of effective feasible region in second stage. In other words, the expanding of feasible region in second stage is only connected to certain components rather than all unselected components. This situation is easily observed in some applications, like bipartite graph coverage [28].

**Observation A2:** The sufficient conditions for total objective function in TSSO to be submodular are

1) the sub-function is submodular, and
2) there is no non-additivity in TSSO’s sub-function, or its non-additivity is related to the selected set scale.

**REFERENCES**

[1] K. Sun, Y. Hou, W. Sun, and J. Qi, *Power System Control Under Cascading Failures: Understanding, Mitigation, and System Restoration*. Hoboken, NJ, USA: Wiley, 2019.

[2] J. W. Busby et al., “Cascading risks: Understanding the 2021 winter blackout in Texas,” *Energy Res. Soc. Sci.*, vol. 77, 2021, Art. no. 102106.

[3] C. Peng, Y. Hou, N. Yu, and W. Wang, “Risk-limiting unit commitment in smart grid with intelligent periphery,” *IEEE Trans. Power Syst.*, vol. 32, no. 6, pp. 4696–4707, Nov. 2017.

[4] C. Wang, P. Ju, S. Lei, Z. Wang, F. Wu, and Y. Hou, “Markov decision process-based resilience enhancement for distribution systems: An approximate dynamic programming approach,” *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2498–2510, May 2020.

[5] F. Liu, J. Guo, X. Zhang, Y. Hou, and S. Mei, “Mitigating the risk of cascading blackouts: A data inference based maintenance method,” *IEEE Access*, vol. 6, pp. 39197–39207, 2018.

[6] J. Iglesias et al., *Guide for Thermal Rating Calculations of Overhead Lines*. Paris, France: Cigre, 2014.

[7] J. Zhan, C. Chung, and E. Demeter, “Time series modeling for dynamic thermal rating of overhead lines,” *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 2172–2182, May 2017.

[8] M. A. Bucher and G. Andersson, “Robust corrective control measures in power systems with dynamic line rating,” *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 2034–2043, May 2016.

[9] F. Qiu and J. Wang, “Distributionally robust congestion management with dynamic line ratings,” *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 2198–2199, Jul. 2015.

[10] M. Wang, M. Yang, J. Wang, M. Wang, and X. Han, “Contingency analysis considering the transient thermal behavior of overhead transmission lines,” *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 4982–4993, Sep. 2018.

[11] R. Xiao, Y. Xiang, L. Wang, and K. Xie, “Power system reliability evaluation incorporating dynamic thermal rating and network topology optimization,” *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6000–6012, Nov. 2018.

[12] R. Chu, “On selecting transmission lines for dynamic thermal line rating system implementation,” *IEEE Trans. Power Syst.*, vol. 7, no. 2, pp. 612–619, May 1992.

[13] J. Zhan, W. Liu, and C. Chung, “Stochastic transmission expansion planning considering uncertain dynamic thermal rating of overhead lines,” *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 432–443, Jan. 2019.

[14] S. Noel and S. Jajodia, “Optimal IDS sensor placement and alert prioritization using attack graphs,” *J. Netw. Syst. Manage.*, vol. 16, no. 3, pp. 259–275, 2008.

[15] J.-A. Jiang et al., “A novel weather information-based optimization algorithm for thermal sensor placement in smart grid,” *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 911–922, Mar. 2018.

[16] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “Wireless sensor networks: A survey,” *Comput. Netw.*, vol. 38, no. 4, pp. 393–422, 2002.

[17] “Power-donut 2 wireless intelligent electrical device for high voltage overhead conductors,” 2014. [Online]. Available: http://www.usi-power.com/.

[18] “Overhead powerline - strain and temperature monitoring,” 2016. [Online]. Available: https://www.sensorsnet.co.uk/overhead-powerline-strain-temperature-monitoring/.

[19] R. Black and W. A. Chisholm, “Key considerations for the selection of dynamic thermal line rating systems,” *IEEE Trans. Power Del.*, vol. 30, no. 5, pp. 2154–2162, Oct. 2015.

[20] Y. Yang, F. Lambert, and D. Divan, “A survey on technologies for implementing sensor networks for power delivery systems,” in *IEEE Power Eng. Soc. Gen. Meeting*, 2007 pp. 1–8.

[21] M. Frank, “The Braess paradox,” *Math. Program.*, vol. 20, no. 1, pp. 283–302, 1981.

[22] Q. Long, Z. Ma, F. Liu, S. Mei, and Y. Hou, “Analyzing patterns transference and mitigation of cascading failures with interaction graphs,” in *Proc. IEEE PES Innov. Smart Grid Technol. Euro.*, 2021, pp. 01–05.
[23] R. Sodhi, S. Srivastava, and S. Singh, “Optimal PMU placement method for complete topological and numerical observability of power system,” Electric Power Syst. Res., vol. 80, no. 9, pp. 1154–1159, 2010.

[24] Z. Wang, B. Chen, J. Wang, J. Kim, and M. M. Begovic, “Robust optimization based optimal DG placement in microgrids,” IEEE Trans. Smart Grid, vol. 5, no. 5, pp. 2173–2182, Sep. 2014.

[25] H.-T. Yang, W.-J. Tang, and P. R. Lubicki, “Placement of fault current limiters in a power system through a two-stage optimization approach,” IEEE Trans. Power Syst., vol. 33, no. 1, pp. 131–140, Jan. 2018.

[26] S. Deb, K. Tammi, X.-Z. Gao, K. Kalita, P. Mahanta, and S. Cross, “A robust two-stage planning model for the charging station placement problem considering road traffic uncertainty,” IEEE Trans. Intell. Transp. Syst., vol. 23, no. 7, pp. 6571–6585, Jul. 2022.

[27] D. T. Nguyen, H. T. Nguyen, N. Trieu, and V. K. Bhargava, “Two-stage robust edge service placement and sizing under demand uncertainty,” IEEE Internet Things J., vol. 9, no. 2, pp. 1560–1574, Jan. 2022.

[28] E. Balkanski, B. Mirzasoleiman, A. Krause, and Y. Singer, “Learning sparse combinatorial representations via two-stage submodular maximization,” in Proc. Int. Conf. Mach. Learn., 2016, pp. 2207–2216.

[29] S. Pan, M. Zadimoghaddam, A. Krause, and A. Karbasi, “Probabilistic submodular maximization in sub-linear time,” in Proc. Int. Conf. Mach. Learn., 2017, pp. 3241–3250.

[30] M. Mitrovic, E. Kazemi, M. Zadimoghaddam, and A. Karbasi, “Data summarization at scale: A two-stage submodular approach,” in Proc. Int. Conf. Mach. Learn., 2018, pp. 3596–3605.

[31] Z. Liu, H. Chang, R. Ma, D. Du, and X. Zhang, “Two-stage submodular maximization problem beyond non-negative and monotone,” in Proc. Int. Conf. Theory Appl. Models Computation, 2020, pp. 144–155.

[32] Y. Li, Z. Liu, C. Xu, F. Li, H. Chang, and X. Zhang, “Two-stage submodular maximization under curvature,” in Proc. Int. Conf. Combinatorial Optim. Appl., 2021, pp. 25–34.

[33] R. Yang, S. Gu, C. Gao, W. Wu, H. Wang, and D. Xu, “A constrained two-stage submodular maximization,” Theor. Comput. Sci., vol. 853, pp. 57–64, 2021.

[34] A. Krause and D. Golovin, “Submodular function maximization,” Tractability, vol. 3, pp. 71–104, 2014.

[35] S. Fujishige, Submodular Functions and Optimization. Amsterdam, The Netherlands: Elsevier, 2005.

[36] R. Billinton, H. Chen, and R. Ghajar, “Time-series models for reliability evaluation of power systems including wind energy,” Microelectron. Rel., vol. 36, no. 9, pp. 1253–1261, 1996.

[37] R. Puffer, M. Schmale, B. Rusek, C. Neumann, and M. Scheufen, “Area-wide dynamic line ratings based on weather measurements. Report B2-106,” Proc. Conf. Cigre Session, vol. 44, 2012.

[38] G. Murphy, “Implementation of real-time thermal ratings,” Scottish Power Energy Network, Tech. Rep., 2013.

[39] C. Harshaw, E. Kazemi, M. Feldman, and A. Karbasi, “The power of sub-sampling in submodular maximization,” Math. Operations Res., vol. 47, no. 2, pp. 1365–1393, 2022.

[40] A. Adibi, A. Mokhtari, and H. Hassani, “Submodular Meta-learning,” in Proc. 34th Int. Conf. Neural Inf. Process. Syst., vol. 33, pp. 3821–3828, 2020.

[41] M. Conforti and G. Cornuéjols, “Submodular set functions, matroids and combinatorial optimization,” Progr. Math., vol. 133, 1998.

[42] A. Krause and D. Golovin, “Submodular function maximization,” Discrete Math., vol. 8, no. 3, pp. 251–274, 1984.

[43] M. Vondrak and J. Ward, “Optimal approximation for submodular and supermodular optimization with bounded curvature,” Math. Program., vol. 126, no. 1, pp. 219–242, 2011.

[44] Q. Long, J. Liu, F. Liu, and Y. Hou, “Submodular optimization of dynamic thermal rating for cascading failure risk mitigation considering Braess paradox,” IEEE Trans. Power Syst., vol. 33, no. 4, pp. 3605–3620, Jul. 2023.

[45] J. Guo, F. Liu, J. Wang, J. Lin, and S. Mei, “Toward efficient cascading outage simulation and probability analysis in power systems,” IEEE Trans. Power Syst., vol. 33, no. 3, pp. 2370–2382, May 2018.

[46] X. Wu, R. Yao, and F. Qiu, “Mitigating cascading outages in severe weather using simulation-based optimization,” IEEE Trans. Power Syst., vol. 36, no. 1, pp. 204–213, Jan. 2021.

[47] P. Vidyullatha and D. R. Rao, “Machine learning techniques on multidimensional curve fitting data based on R-square and chi-square methods,” Int. J. Electr. Comput. Eng., vol. 6, no. 3, 2016, Art. no. 974.

[48] A. Arguez et al., “NOAA’s 1981–2010 US climate normals: An overview,” Bull. Amer. Meteorological Soc., vol. 93, no. 11, pp. 1687–1697, 2012. [Online]. Available: https://journals.ametsoc.org/view/journals/bams/93/11/bams-d-11-00197.1.xml
Feng Liu (Senior Member, IEEE) received the B.Sc. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1999 and 2004, respectively. He is currently an Associate Professor with Tsinghua University. From 2015 to 2016, he was a Visiting Associate with the California Institute of Technology, CA, USA. He is the author/coauthor of more than 200 peer-reviewed technical papers and two books, and holds more than 20 issued/pending patents. His research interests include stability analysis, optimal control, robust dispatch and game theory based decision making in energy and power systems. He is also an Associate Editor of several international journals including IEEE TRANSACTIONS ON SMART GRID and IEEE PES LETTER. He was the Guest Editor of IEEE TRANSACTIONS ON ENERGY CONVERSION.

Yunhe Hou (Senior Member, IEEE) received the B.E. and Ph.D. degrees in electrical engineering from the Huazhong University of Science and Technology, Wuhan, China, in 1999 and 2005, respectively. From 2005 to 2007, he was a Postdoctoral Research Fellow with Tsinghua University, Beijing, China, and a Postdoctoral Researcher with Iowa State University, Ames, IA, USA, and the University College Dublin, Dublin, Ireland, from 2008 to 2009. In 2010, he was also a Visiting Scientist with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA. Since 2017, he has been a Guest Professor with the Huazhong University of Science and Technology, China and has been an Academic Adviser of China Electric Power Research Institute since 2019. In 2009, he joined the Faculty of the University of Hong Kong, Hong Kong, where he currently is an Associate Professor with the Department of Electrical and Electronic Engineering. From 2016 to 2021, he was an Associate Editor for the IEEE TRANSACTIONS ON SMART GRID. Dr. Hou is also an Associate Editor for the IEEE TRANSACTIONS ON POWER SYSTEMS and Journal of Modern Power Systems and Clean Energy.