

Investments in Robustness of Complex Systems: Algorithm Design

Van-Sy Mai¹, Richard J. La², and Abdella Battou¹

¹ NIST, Gaithersburg, MD 20899, USA,

{vansy.mai, richard.la, abdella.battou}@nist.gov

² NIST and University of Maryland, College Park, MD 20742, USA,

WWW home page: <http://www.ece.umd.edu/~hyongla>

Abstract. We study the problem of determining suitable investments in improving the robustness of complex systems comprising many component systems with an aim of minimizing the (time) average costs to system operators. The problem is formulated as an optimization problem that is nonconvex and challenging to solve for large systems. We propose two approaches to finding a good solution to the optimization problem: the first approach is based on a gradient method and finds a local optimizer. The second approach makes use of a convex relaxation of the original problem and provides both a lower bound on the optimal value and a feasible point. The lower bound can be used to bound the optimality gap of the solutions obtained by our methods. We provide numerical results to demonstrate the effectiveness of the proposed approaches.

Keywords: complex systems, optimization, resilience, robustness

1 Introduction

With increasing complexity, modern engineering systems, such as information and communication networks and power systems, consist of many (component) systems that depend on each other to deliver their services. This interdependence among systems makes it possible for a local failure or infection of a system by malware to spread to other systems. From this viewpoint, it is clear that sound investments in robustness of the complex system should consider the interdependence among comprising systems. A similar issue arises also in the problem of managing the spread of an infectious disease via social contacts.

The problem of optimizing the investments in robustness of complex systems or the mitigation of disease spread has been studied extensively. In [4, 5, 8, 9], researchers adopted a game theoretic formulation to study the problem of security investments with distributed agents. In another line of research more closely related to our study, researchers examined optimal strategies using vaccines/immunization (prevention) [3], antidotes or curing rates (recovery) [2, 10, 14] or a combination of both preventive and recovery measures [13, 15]. However, these studies do not take into account dynamics; they focus on either the expected costs stemming from single or cascading failures/infections [7–9] or the

exponential decay rate to the disease-free state as a key performance metric. When systems experience random failures over time, the exponential decay rate adopted in the previous studies is no longer a suitable performance metric.

In a recent study, Mai et al. [11, 12] investigated the problem of minimizing the (time) average costs of a system operator while accounting for dynamics, where the costs include both (security) investments and economic losses incurred following failures or infections. However, the authors considered investments only in resilience, but not in recovery. In this paper, we extend this study and consider investments in both resilience and recovery. It turns out that incorporating two different types of investments complicates the optimization problem for determining optimal investments significantly. This is due to additional *coupling* terms that are introduced in the new model, which were not present in [11, 12]. This leads to a highly nonconvex optimization problem that is difficult to solve in general. However, we show that, under a technical condition, we can formulate a convex relaxation that provides a lower bound on the optimal value and a good feasible solution for the original problem (Theorem 4), whose optimality gap can be bounded. In addition, we show that a gradient-based method also produces a good-quality solution.

Notation and Terminology – Let \mathbb{R} and \mathbb{R}_+ denote the set of real numbers and nonnegative real numbers, respectively. For a matrix $A = [a_{i,j}]$, let $a_{i,j}$ denote its (i, j) element and A^\top its transpose. We use boldface letters to denote (column) vectors, e.g., $\mathbf{x} = [x_1, \dots, x_n]^\top$, $\mathbf{0} = [0, \dots, 0]^\top$, and $\mathbf{1} = [1, \dots, 1]^\top$. For any two vectors \mathbf{x} and \mathbf{y} of the same dimension, $\mathbf{x} \circ \mathbf{y}$, $\frac{\mathbf{x}}{\mathbf{y}}$, and $\mathbf{x}^{\mathbf{y}}$ are their element-wise product, division, and exponentiation, respectively. For $\mathbf{x} \in \mathbb{R}^n$, $\text{diag}(\mathbf{x}) \in \mathbb{R}^{n \times n}$ denotes the diagonal matrix with diagonal elements x_1, \dots, x_n .

A directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of a set of nodes \mathcal{V} and a set of directed edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. A directed path in \mathcal{G} is a sequence of directed edges in the form $((i_1, i_2), (i_2, i_3), \dots, (i_{k-1}, i_k))$. The graph \mathcal{G} is strongly connected if there is a directed path from each node to any other node.

The rest of the paper is organized as follows: Section 2 describes the setup and the problem formulation, including the optimization problem. Our proposed approaches are described in Section 3, followed by numerical studies in Section 4. We conclude in Section 5.

2 Model and Formulation

The complex system under consideration consists of N ($N \gg 1$) systems, and we denote the set of (component) systems by $\mathcal{A} = \{1, \dots, N\}$. Each system in a subset $\mathcal{A}_R \subset \mathcal{A}$ experiences random failures. The frequency with which a system $i \in \mathcal{A}_R$ suffers random failures depends on the amount of investments in its *resilience*, which we denote by s_i^p ; when system i invests s_i^p in improving its *resilience*, it experiences random failures according to a Poisson process with failure rate $\lambda_i(s_i^p)$. Here, we assume that $\lambda_i(s_i^p) = \bar{\lambda}_i \times q_i(s_i^p)$, where $\bar{\lambda}_i \geq 0$ is the failure rate when no investment is made in its resilience, and $q_i : \mathbb{R}_+ \rightarrow [0, 1]$ is a

decreasing function and quantifies how the resilience of system i improves with its investment in resilience. We assume $\bar{\lambda}_j = 0$ for every system $j \in \mathcal{A} \setminus \mathcal{A}_R =: \mathcal{A}_R^c$.

In addition to random failures, systems also experience *secondary* failures brought on by the failures of other systems due to interdependence among systems. The rate at which the failure of system i causes that of another system j depends on system j 's resilience and is equal to $\xi_{i,j} \times q_j(s_j^p)$, where $\xi_{i,j} \in \mathbb{R}_+$. Thus, even systems in \mathcal{A}_R^c can experience secondary failures. Note that this failure transmission rate depends on system j 's investment in resilience. When $\xi_{i,j} > 0$, we say that system i supports system j or system j depends on system i . We adopt the convention $\xi_{i,i} = 0$ for all $i \in \mathcal{A}$.

When system i suffers a failure, the *recovery time* required to repair the system and put it back in service depends on the amount of investment in *recovery*, which we denote by s_i^r ; when system i invests s_i^r in *recovery*, the recovery times are given by independent and identically distributed exponential random variables with parameter $\delta_i(s_i^r)$. We assume that $\delta_i : \mathbb{R}_+ \rightarrow (0, \infty)$ is strictly increasing. Furthermore, the recovery times of different systems are assumed to be mutually independent.

In addition to the investments in the resilience and recovery of systems, the system operator also incurs other costs; when system i fails, it can cause economic losses, e.g., some servers in system i may need to be taken offline for inspection and repair and remain inaccessible during the period to other systems that depend on the servers. We call the economic losses *failure costs*. To model the failure costs, we assume that the failure of system i causes economic losses of c_i per unit time. Define $\mathbf{c} = (c_i : i \in \mathcal{A})$ to be the failure cost vector.

From our discussion, we can define a following dependence graph $\mathcal{G} = (\mathcal{A}, \mathcal{E})$, where $\mathcal{E} := \{(i, j) \mid \xi_{i,j} > 0, i, j \in \mathcal{A}\}$. Let $B = [b_{i,j} : i, j \in \mathcal{A}]$ be an $N \times N$ matrix that describes the failure transmission rates among systems, where the element $b_{i,j}$ is equal to $\xi_{j,i}$. We assume that B is irreducible. Note that B is irreducible if and only if the dependence graph \mathcal{G} is strongly connected.

2.1 Model

We adopt the well-known Susceptible-Infected-Susceptible (SIS) model to capture the evolution of the state of each system: if a system i is up and running at time $t \in \mathbb{R}_+$, we say that the system is at 'susceptible' state. If the system is being repaired following a failure at time t , we say that the system is 'infected'. Let $\mathbf{p}(t)$ be a vector, whose i -th element is the probability that system i is at 'infected' state at time $t \in \mathbb{R}_+$. The dynamics of $\mathbf{p}(t)$ is approximated using the following (Markov) differential equations, starting with $\mathbf{p}(0)$ at $t = 0$:

$$\dot{\mathbf{p}}(t) = (\mathbf{1} - \mathbf{p}(t)) \circ (\bar{\boldsymbol{\lambda}} + B\mathbf{p}(t)) \circ \mathbf{q}(\mathbf{s}) - \boldsymbol{\delta}(\mathbf{s}) \circ \mathbf{p}(t), \quad t \in \mathbb{R}_+, \quad (1)$$

where $\mathbf{s} = (\mathbf{s}_i = (s_i^p, s_i^r) : i \in \mathcal{A})$, $\bar{\boldsymbol{\lambda}} = (\bar{\lambda}_i : i \in \mathcal{A})$, $\mathbf{q}(\mathbf{s}) = (q_i(s_i^p) : i \in \mathcal{A})$ and $\boldsymbol{\delta}(\mathbf{s}) = (\delta_i(s_i^r) : i \in \mathcal{A})$.

Suppose that, for each fixed investment profile $\mathbf{s} = (\mathbf{s}_i : i \in \mathcal{A})$ in \mathbb{R}_+^{2N} , the system state $\mathbf{p}(t)$ converges to a stable equilibrium (the existence and uniqueness

of such an equilibrium will be addressed shortly), which we denote by $\bar{\mathbf{p}}(\mathbf{s})$. From (1), it is clear that $\bar{\mathbf{p}}(\mathbf{s})$ is a solution to the following equation.

$$\mathbf{g}^{\mathbf{s}}(\mathbf{p}) := (\mathbf{1} - \mathbf{p}) \circ (\bar{\boldsymbol{\lambda}} + B\mathbf{p}) \circ \mathbf{q}(\mathbf{s}) - \boldsymbol{\delta}(\mathbf{s}) \circ \mathbf{p} = \mathbf{0} \quad (2)$$

We are interested in solving the following problem:

$$[\mathbf{P0}] \quad \min_{\mathbf{s} \geq \mathbf{0}} F(\mathbf{s}) := w(\mathbf{s}) + \mathbf{c}^T \bar{\mathbf{p}}(\mathbf{s}), \quad (3)$$

where $w : \mathbb{R}_+^{2N} \rightarrow \mathbb{R}_+$ is the cost function that quantifies the investment costs. The second term in (3) is the average failure costs that the system operator suffers due to the failures of systems. Although we do not impose any constraints on \mathbf{s} (other than nonnegativity), our analysis can be extended to handle constraints, e.g., $\mathbf{1}^T \mathbf{s} \leq s_{\text{bgt}}$, where s_{bgt} is the available budget for investments.

3 Main Analysis

In order to make progress, we introduce following assumptions.

Assumption 1 *Suppose that \mathcal{G} is strongly connected and the following holds:*

A1-a. *At least one system experiences random failures with a positive rate, i.e., $\mathcal{A}_R \neq \emptyset$ and $\bar{\lambda}_i > 0$ for all $i \in \mathcal{A}_R$.*

A1-b. *For each $i \in \mathcal{A}$, the function $q_i(s_i^p) = (1 + \kappa_i s_i^p)^{-\alpha_i}$, where κ_i and α_i are some positive constants.*

A1-c. *For each $i \in \mathcal{A}$, the recovery rate of system i is given by $\delta_i(s_i^r) = \theta_i(1 + \zeta_i \cdot s_i^r)^{\beta_i} > 0$, where β_i , θ_i and ζ_i are some positive constants.*

A1-d. *The cost of investments is equal to $w(\mathbf{s}) = \mathbf{1}^T \mathbf{s}^p + \mathbf{1}^T \mathbf{s}^r$.*

Assumption A1-b can be viewed as a form of the law of diminishing returns with increasing investments in resilience, where the shape is determined by the parameters α_i and κ_i . Similarly, since the mean recovery time is inversely proportional to the recovery rate, the law of diminishing returns may suggest $\beta_i < 1$ in Assumption A1-c. Larger values of α_i and κ_i (resp. β_i , θ_i , and ζ_i) indicate higher effectiveness of available tools for improving resilience (resp. expediting recovery) and, thus, greater benefits from investments in resilience and recovery.

The following theorem states that, for fixed investments $\mathbf{s} \in \mathbb{R}_+^{2N}$, there is a unique equilibrium of the differential system described by (1); see, e.g., [6].

Theorem 1. *Suppose that Assumption 1 holds. Then, for fixed investments $\mathbf{s} \in \mathbb{R}_+^{2N}$, there is a unique equilibrium $\bar{\mathbf{p}}(\mathbf{s}) \in (0, 1)^N$ that satisfies (2).*

Theorem 1 asserts that the unique equilibrium of (1) satisfying (2) is strictly positive. Hence, under Assumption 1, we can rewrite the constraints in (2) as

$$(\mathbf{p}^{-1} - \mathbf{1}) \circ (\bar{\boldsymbol{\lambda}} + B\mathbf{p}) = \boldsymbol{\theta} \circ (\mathbf{1} + \boldsymbol{\kappa} \circ \mathbf{s}^p)^\alpha \circ (\mathbf{1} + \boldsymbol{\zeta} \circ \mathbf{s}^r)^\beta, \quad (4)$$

where $\mathbf{p}^{-1} = (p_i^{-1} : i \in \mathcal{A})$. Based on this observation, we reformulate our original problem [P0] in (3) as the following equivalent problem:

$$\begin{aligned} \text{[P1]} \quad & \min_{\mathbf{s}, \mathbf{p}} \quad w(\mathbf{s}) + \mathbf{c}^\top \mathbf{p} =: f(\mathbf{s}, \mathbf{p}) \\ & \text{s.t.} \quad (4), \mathbf{p} \in (0, 1]^N, \mathbf{s} \geq \mathbf{0} \end{aligned}$$

The main difficulty in solving this problem lies with the constraint in (4). Specifically, the constraint in (4) involves bilinear terms that are not only nonconvex, but also known to be difficult to handle. For this reason, although the objective function is linear in optimization variables \mathbf{s} and \mathbf{p} , problem [P1] is nonconvex.

In view of the observation that [P1] is nonconvex, finding an optimal point for a large system (with $N \gg 1$) is in general challenging. In this section, we will discuss how we can obtain a feasible point $(\mathbf{s}', \mathbf{p}')$ to [P1] along with a lower bound f_{lb} on the optimal value f^* of [P1] so that we can bound the gap $f(\mathbf{s}', \mathbf{p}') - f^*$ using $f(\mathbf{s}', \mathbf{p}') - f_{lb}$. In order to find a lower bound on f^* , under a technical assumption, we formulate a convex relaxation of [P1], which can be solved efficiently. We find a feasible point to [P1] in two different ways; in the first method, we use a gradient-based method to find a local minimizer of [P1]. In the second method, we use an optimal point to the aforementioned convex relaxation to construct a feasible point and show that it solves [P1] under certain conditions.

In order to cope with the difficulty caused by constraint in (4), we first rewrite the constraint as following two constraints using an auxiliary variable $\phi \geq \mathbf{1}$.

$$(\mathbf{p}^{-1} - \mathbf{1}) \circ (\bar{\boldsymbol{\lambda}} + B\mathbf{p}) = \boldsymbol{\theta} \circ \phi \quad (5a)$$

$$\phi = (\mathbf{1} + \boldsymbol{\kappa} \circ \mathbf{s}^\rho)^\alpha \circ (\mathbf{1} + \boldsymbol{\zeta} \circ \mathbf{s}^r)^\beta \quad (5b)$$

Notice that constraint (5a) involves only optimization variables \mathbf{p} , whereas constraint (5b) has optimization variables \mathbf{s} . This observation will be exploited in our algorithm design below. Here, we will briefly illustrate the usefulness of this structure. Consider the following subproblems for fixed $\phi \geq \mathbf{1}$:

$$\begin{aligned} \text{[SP1]} \quad & \min_{\mathbf{p}} \quad \mathbf{c}^\top \mathbf{p} =: g(\mathbf{p}) & \text{[SP2]} \quad & \min_{\mathbf{s}} \quad w(\mathbf{s}) \\ & \text{s.t.} \quad (5a), \mathbf{p} \in (0, 1]^N & & \text{s.t.} \quad (5b), \mathbf{s} \geq \mathbf{0} \end{aligned}$$

Here, in view of Theorem 1, [SP1] is simply the problem of finding the equilibrium failure probability $\bar{\mathbf{p}}$ for fixed $\boldsymbol{\theta} \circ \phi$ corresponding to some investment profile \mathbf{s} . Unfortunately, there appears to be no closed-form solution to this problem. One can, however, resort to a numerical method instead as shown below.

Theorem 2. ([12]) *Suppose $\bar{\boldsymbol{\lambda}} \succeq \mathbf{0}$, $\boldsymbol{\theta} > \mathbf{0}$, and B is irreducible. Then, the iteration*

$$\mathbf{p}_{k+1} = \frac{\bar{\boldsymbol{\lambda}} + B\mathbf{p}_k}{\bar{\boldsymbol{\lambda}} + B\mathbf{p}_k + \boldsymbol{\theta} \circ \phi}, \quad k \in \mathbb{N}, \quad (6)$$

converges to a unique equilibrium $\hat{\mathbf{p}}$ when starting with any \mathbf{p}_0 such that $\hat{\mathbf{p}} \leq \mathbf{p}_0 \leq \mathbf{1}$. Moreover, the convergence is exponential with some rate $\rho_0 < 1 - \min_{i \in \mathcal{A}} \bar{p}_i$.

Next, let us consider problem [SP2], which amounts to finding optimal investments for a given ϕ , or equivalently, for a given failure probability $\hat{\mathbf{p}}$. Unlike [SP1], problem [SP2] can be solved analytically as follows:

Theorem 3. *For each $\phi \geq \mathbf{1}$, the solution to [SP2] is given by*

$$\begin{aligned}\hat{s}_i^r(\phi) &= \zeta_i^{-1} \max \{0, \tau_i \phi_i^{1/(\alpha_i + \beta_i)} - 1\} \quad \text{with } \tau_i := \left(\frac{\beta_i \zeta_i}{\alpha_i \kappa_i} \right)^{\frac{\alpha_i}{\alpha_i + \beta_i}} \\ \hat{s}_i^p(\phi) &= \kappa_i^{-1} \left(\phi_i^{1/\alpha_i} \max \{1, \tau_i \phi_i^{1/(\alpha_i + \beta_i)}\}^{-\beta_i/\alpha_i} - 1 \right).\end{aligned}$$

The proof of this theorem is straightforward and is omitted here.

Denote the optimal values of [SP1] and [SP2] for fixed $\phi \geq \mathbf{1}$ by $g^*(\phi)$ and $w^*(\phi)$, respectively. Then, in principle, we can solve problem [P1] by solving

$$\min_{\phi \geq \mathbf{1}} \quad g^*(\phi) + w^*(\phi). \quad (7)$$

Unfortunately, this is in general not a convex problem due to the implicit function $g^*(\phi)$ and possible nonconvexity of $\hat{s}_i^p(\phi)$ and $\hat{s}_i^r(\phi)$ (which is the case when $\alpha_i + \beta_i > 1$) as described in Theorems 2 and 3, respectively. As a result, we resort to a numerical method that can find a (local) optimizer of the problem. Among different methods for solving nonconvex problems, we consider next in subsection 3.1 a gradient type algorithm due to its simplicity and scalability. Later, in subsection 3.2 we will show that when $\alpha_i + \beta_i \leq 1$, we can obtain practical convex relaxation of the original problem.

3.1 Gradient Method

First, it is tempting to use a gradient method to solve the problem in (7). However, note that w^* is nonsmooth and possibly nonconvex since $\hat{s}_i^r(\phi)$ is nonsmooth and nonconvex. Thus, directly solving this problem using gradient methods is known to be difficult. As a result, we will use [P0], which is of higher dimension than (7) but smooth with simple constraints and hence is easier to apply gradient methods to.

To this end, we show how to compute gradient $\nabla F(\mathbf{s})$ efficiently. Note that

$$\nabla F(\mathbf{s}) = \nabla w(\mathbf{s}) + \nabla g(\mathbf{p}(\phi(\mathbf{s}))) = \mathbf{1} + J_{\mathbf{p}}(\mathbf{s})^T \mathbf{c}, \quad (8)$$

where $J_{\mathbf{p}}(\mathbf{s}) = [\partial p_i(\mathbf{s})/\partial s_j]$ is the Jacobian matrix. Thus, the bulk of computation lies in evaluating $J_{\mathbf{p}}(\mathbf{s})$. By applying the chain rule, we obtain

$$J_{\mathbf{p}}(\mathbf{s}) = J_{\mathbf{p}}(\phi) J_{\phi}(\mathbf{s}). \quad (9)$$

Here, $J_{\phi}(\mathbf{s})$ can be evaluated easily from (5b), i.e.,

$$J_{\phi}(\mathbf{s}) = \left[\text{diag} \left(\frac{\phi \circ \alpha \circ \kappa}{\mathbf{1} + \kappa \circ \mathbf{s}^p} \right), \text{diag} \left(\frac{\phi \circ \beta \circ \zeta}{\mathbf{1} + \zeta \circ \mathbf{s}^r} \right) \right]. \quad (10)$$

The matrix $J_{\mathbf{p}}(\phi)$ can be computed by totally differentiating (5a) with respect to ϕ ; this is similar to the approach of [12]. In fact,

$$M(\phi)J_{\mathbf{p}}(\phi) = -\text{diag}(\boldsymbol{\theta} \circ \bar{\mathbf{p}}(\mathbf{s})) \quad (11)$$

where $M(\phi) = \text{diag}(\boldsymbol{\theta} \circ \phi + \bar{\boldsymbol{\lambda}} + B\bar{\mathbf{p}}(\mathbf{s})) - \text{diag}(\mathbf{1} - \bar{\mathbf{p}}(\mathbf{s}))B$, and $\bar{\mathbf{p}}(\mathbf{s}) = \hat{\mathbf{p}}(\phi(\mathbf{s}))$, which can be computed efficiently using the fixed point iteration in Theorem 2. We can show that $M(\phi)$ is a nonsingular M-matrix. Thus, from (11) we get

$$J_{\mathbf{p}}(\phi) = -M(\phi)^{-1}\text{diag}(\boldsymbol{\theta} \circ \bar{\mathbf{p}}(\mathbf{s})). \quad (12)$$

Substituting (10) and (12) in (9) and using it in (8), we obtain

$$\nabla F(\mathbf{s}) = \mathbf{1} - J_{\phi}(\mathbf{s})^{\top}\text{diag}(\boldsymbol{\theta} \circ \bar{\mathbf{p}}(\mathbf{s}))\mathbf{z}, \quad \text{where } \mathbf{z} := M(\phi)^{-\top}\mathbf{c}.$$

As a result, we can now use a projected gradient method for solving [P0]; see [12] for the detailed algorithm as well as an efficient and scalable approach for computing \mathbf{z} without matrix inversions for large systems (by employing the structure of $M(\phi)$ and using the power method).

3.2 Convex Relaxation

We will introduce relaxations to the constraints in (5). First, to relax (5a), we use the same approach used in [12]. Specifically, define

$$\mathbf{y} := -\ln \mathbf{p}, \quad \mathbf{t} := \bar{\boldsymbol{\lambda}} \circ e^{\mathbf{y}}, \quad U := \text{diag}(e^{\mathbf{y}})B\text{diag}(e^{-\mathbf{y}}). \quad (13)$$

Using these new variables, (5a) can be rewritten as

$$\mathbf{t} + U\mathbf{1} = \bar{\boldsymbol{\lambda}} + B\mathbf{p} + \boldsymbol{\theta} \circ \phi, \quad (14)$$

which is linear in the variables $\mathbf{t}, \mathbf{p}, \phi$ and U . Next, we relax the nonconvex equality constraints in (13) with the following convex inequality constraints:

$$e^{-\mathbf{y}} \leq \mathbf{p} \leq \mathbf{1}, \quad \bar{\boldsymbol{\lambda}} \circ e^{\mathbf{y}} \leq \mathbf{t}, \quad \text{diag}(e^{\mathbf{y}})B\text{diag}(e^{-\mathbf{y}}) \leq U \quad (15)$$

We can express these inequality constraints as a following set of at most $2N + |\mathcal{E}|$ exponential cone constraints:

$$(p_i, 1, -y_i) \in \mathcal{K}_{\text{exp}} \quad \text{for all } i \in \mathcal{A} \quad (16a)$$

$$(t_i, 1, y_i + \log \bar{\lambda}_i) \in \mathcal{K}_{\text{exp}} \quad \text{for all } i \in \mathcal{A}_R \quad (16b)$$

$$(u_{ij}, 1, y_i - y_j + \log b_{ij}) \in \mathcal{K}_{\text{exp}} \quad \text{for all } (i, j) \in \mathcal{E}, \quad (16c)$$

where $\mathcal{K}_{\text{exp}} := \text{cl}(\{(x_1, x_2, x_3) \mid x_1 \geq x_2 e^{x_3/x_2}, x_2 > 0\})$. These constraints in (16) (as well as those in (19) below) can be handled efficiently by conic optimization solvers, e.g., MOSEK [1].³

³ Any mention of commercial products is for information only; it does not imply a recommendation or endorsement by NIST.

We now consider the constraint (5b). Note that since we aim to minimize the investment costs from \mathbf{s}^p and \mathbf{s}^r and the right-hand side of (5b) is strictly increasing in each element, at an optimal point the constraint will be active, allowing us to replace the equality with the inequality, i.e.,

$$\boldsymbol{\phi} \leq (\mathbf{1} + \boldsymbol{\kappa} \circ \mathbf{s}^p)^\alpha \circ (\mathbf{1} + \boldsymbol{\zeta} \circ \mathbf{s}^r)^\beta. \quad (17)$$

In general this is not a convex constraint because of the product term on the right-hand side; however, it can be recast as convex constraints when $\boldsymbol{\alpha} + \boldsymbol{\beta} \leq \mathbf{1}$. To see this, let us first use a change of variable to rewrite the right-hand side of (17): for each $i \in \mathcal{A}$, define

$$\eta_i = 1 + \kappa_i s_i^p \quad \text{and} \quad \vartheta_i = 1 + \zeta_i s_i^r. \quad (18)$$

From their relations, we have $s_i^p(\eta_i) := (\eta_i - 1)/\kappa_i$ and $s_i^r(\vartheta_i) := (\vartheta_i - 1)/\zeta_i$. Note that η_i and ϑ_i are linear in s_i^p and s_i^r , respectively, and vice versa. With a little abuse of notation, we denote $(s_i^p(\eta_i) : i \in \mathcal{A})$ and $(s_i^r(\vartheta_i) : i \in \mathcal{A})$ by $\mathbf{s}^p(\boldsymbol{\eta})$ and $\mathbf{s}^r(\boldsymbol{\vartheta})$, respectively. In order to rewrite constraint (17) as conic constraints, we need the following assumption.

Assumption 2 *We assume $\boldsymbol{\alpha} + \boldsymbol{\beta} \leq \mathbf{1}$.*

This assumption implies that the (marginal) rates of both the increase in recovery rates and the decrease in failure rates slow down relatively quickly with increasing investments. In other words, the available tools are not very effective and, as a result, the failure rates do not diminish quickly and the recovery rates do not improve rapidly with increasing investments in resilience and recovery, respectively. Under Assumption 2, we can express the constraint in (17) as the following conic constraints:

$$\begin{cases} (\eta_i, \vartheta_i, \phi_i) \in \mathcal{P}_3^{\alpha_i, 1-\alpha_i} & \text{if } \alpha_i + \beta_i = 1, \\ (\eta_i, \vartheta_i, 1, \phi_i) \in \mathcal{P}_4^{\alpha_i, \beta_i, 1-\alpha_i-\beta_i} & \text{if } \alpha_i + \beta_i < 1, \end{cases} \quad i \in \mathcal{A}, \quad (19)$$

where $\mathcal{P}_n^{\alpha_1, \dots, \alpha_m} = \left\{ \mathbf{x} \in \mathbb{R}^n \mid \prod_{i=1}^m x_i^{\alpha_i} \geq \sqrt{\sum_{j=m+1}^n x_j^2}, x_1, \dots, x_m \geq 0 \right\}$, $m < n$, is an n -dimensional power cone ($n \geq 3$), which is convex. Clearly, the above power cone constraints require Assumption 2. When this assumption does not hold, one must resort to other techniques to obtain a convex relaxation of (17).

Based on these new constraints (14)–(19), we obtain the following convex relaxation of [P1]: define $\tilde{w}(\boldsymbol{\eta}, \boldsymbol{\vartheta}) := w(\mathbf{s}^p(\boldsymbol{\eta}), \mathbf{s}^r(\boldsymbol{\vartheta})) = \sum_{i=1}^N (s_i^p(\eta_i) + s_i^r(\vartheta_i))$.

$$\begin{aligned} \text{[CR]} \quad & \min_{\mathbf{p}, \mathbf{t}, \mathbf{y}, \boldsymbol{\eta}, \boldsymbol{\vartheta}, \boldsymbol{\phi}, U} \quad \tilde{w}(\boldsymbol{\eta}, \boldsymbol{\vartheta}) + \mathbf{c}^\top \mathbf{p} \\ & \text{s.t.} \quad (14), (16), (19), \\ & \quad \mathbf{p} \in (0, 1]^N, \boldsymbol{\eta} \geq \mathbf{1}, \boldsymbol{\vartheta} \geq \mathbf{1}, \mathbf{y} \geq \mathbf{0} \end{aligned}$$

Suppose that $\mathbf{x}_R^+ := (\mathbf{p}^+, \mathbf{t}^+, \mathbf{y}^+, \boldsymbol{\eta}^+, \boldsymbol{\vartheta}^+, \boldsymbol{\phi}^+, U^+)$ is an optimal point of [CR], and let $\mathbf{s}^+ = (\mathbf{s}^p(\boldsymbol{\eta}^+), \mathbf{s}^r(\boldsymbol{\vartheta}^+))$. Define

$$\mathbf{p}' = e^{-\mathbf{y}^+}, \boldsymbol{\phi}' = \boldsymbol{\phi}^+ + \text{diag}(\boldsymbol{\theta}^{-1})B(\mathbf{p}^+ - \mathbf{p}'), \text{ and } \mathbf{s}' = \hat{\mathbf{s}}(\boldsymbol{\phi}'), \quad (20)$$

where $\hat{\mathbf{s}}(\phi')$ is an optimal point of [SP2] for ϕ' , which was given in Theorem 3. The above relaxation provides both an upper bound and a lower bound on the optimal cost as shown below.

Theorem 4. *Suppose that f^* is the optimal value of [P1]. Then, $(\mathbf{s}', \mathbf{p}')$ is a feasible point of [P1], and we have*

$$f(\mathbf{s}^+, \mathbf{p}^+) \leq f^* \leq f(\mathbf{s}', \mathbf{p}'). \quad (21)$$

Moreover, the last two constraints of (15) are active at \mathbf{x}^+ , i.e.,

$$\mathbf{t}^+ = \bar{\lambda} \circ e^{\mathbf{y}^+} \quad \text{and} \quad U^+ = \text{diag}(e^{\mathbf{y}^+}) B \text{diag}(e^{-\mathbf{y}^+}). \quad (22)$$

This result shows that the tightness of our relaxation can be judged via the gap $f(\mathbf{s}', \mathbf{p}') - f(\mathbf{s}^+, \mathbf{p}^+)$. Note that in view of Theorem 4, $w(\mathbf{s})$ can be expressed as $w(\hat{\mathbf{s}}(\phi)) =: \hat{w}(\phi)$, which is a convex function of ϕ under Assumption 2. Thus,

$$\begin{aligned} f(\mathbf{s}', \mathbf{p}') - f(\mathbf{s}^+, \mathbf{p}^+) &= \hat{w}(\phi') - \hat{w}(\phi^+) + \mathbf{c}^\top (\mathbf{p}' - \mathbf{p}^+) \\ &\leq \nabla \hat{w}(\phi')^\top (\phi' - \phi^+) + \mathbf{c}^\top (\mathbf{p}' - \mathbf{p}^+) \quad (\text{convexity of } \hat{w}) \\ &= (\mathbf{c} - B^\top \nabla \hat{w}(\phi'))^\top (\mathbf{p}' - \mathbf{p}^+), \quad (\text{from (20)}) \end{aligned}$$

where $\nabla \hat{w}(\phi')$ is the gradient of \hat{w} at ϕ' (or any subgradient if non-differentiable). Since $\mathbf{p}' \leq \mathbf{p}^+$, the above bound suggests that the gap $f(\mathbf{s}', \mathbf{p}') - f(\mathbf{s}^+, \mathbf{p}^+)$ is likely to be small when the failure cost vector \mathbf{c} is sufficiently large (which is often the case in practice). Clearly, this gap is 0 when $\mathbf{c} \geq B^\top \nabla \hat{w}(\phi')$, i.e., $(\mathbf{s}', \mathbf{p}')$ is a global optimal solution to the original problem; this is the case, for example, when $\boldsymbol{\alpha} + \boldsymbol{\beta} = \mathbf{1}$ and \mathbf{c} is sufficiently large (independent of ϕ'). Moreover, when the convex relaxation is not tight, we can obtain an improved solution using the gradient method in Section 3.1 with $(\mathbf{s}', \mathbf{p}')$ as a starting point.

Finally, we note that when either $(\alpha_i, \beta_i) = (0, 1)$ or $(\alpha_i, \beta_i) = (1, 0)$ for all $i \in \mathcal{A}$, both (5b) and (17) are simple affine constraints and can be used directly in our relaxed problem without the need to transform them into conic constraints provided in (19); a similar approach for the special case with $(\boldsymbol{\alpha}, \boldsymbol{\beta}) = (\mathbf{0}, \mathbf{1})$ can be found in [12].

4 Numerical Results

In this section, we provide some numerical results to demonstrate the usefulness and efficacy of our proposed approaches. Our studies are carried out in MATLAB (version 9.5) on a laptop with 8GB RAM and a 2.4GHz Intel Core i5 processor. For our numerical studies, we generate a set of strongly connected scale-free networks with the power law parameter for node degrees set to 1.5, and the minimum and maximum node degrees equal to 2 and $\lceil 3 \log N \rceil$, respectively, in order to ensure network connectivity with high probability.

For all considered networks, we fix $\theta_i = 1$ and $\alpha_i = \beta_i = 0.5$ for all $i \in \mathcal{A}$. The failure transmission rates $b_{j,i}$, $(j, i) \in \mathcal{E}$, and the parameters κ_i and ζ_i are

selected uniformly at random in $[0.01, 1]$, $[1, 1.5]$, and $[0.5, 1]$, respectively. We choose failure cost vector $\mathbf{c} = \nu B^T \mathbf{1}$ with a varying parameter $\nu > 0$ to reflect an observation that nodes that support more neighbors should, on the average, cause larger economic losses. In each considered network, we assign positive failure rates of $\bar{\lambda}_i = 0.1$ to 20% of the nodes (sampled without replacement).

We solve the relaxed problem [CR] using MOSEK package and define the relative optimality gap at $(\mathbf{s}', \mathbf{p}')$ as $\text{opt_gap} = \left| 1 - \frac{f(\mathbf{s}', \mathbf{p}')}{f(\mathbf{s}^+, \mathbf{p}^+)} \right|$. We use the Reduced Gradient Method (RGM) in [12] to find a local optimizer $\tilde{\mathbf{s}}$ of [P0] with initial point $\mathbf{s}^{(0)} = \mathbf{0}$. This gives us $F(\tilde{\mathbf{s}})$, which is an upper bound on the optimal cost, and the relative optimality gap for RGM's solution is given by $\text{opt_gap} = \left| 1 - \frac{f(\mathbf{s}^+, \mathbf{p}^+)}{F(\tilde{\mathbf{s}})} \right|$. The results (averaged over 5 runs) are shown in Fig. 1.

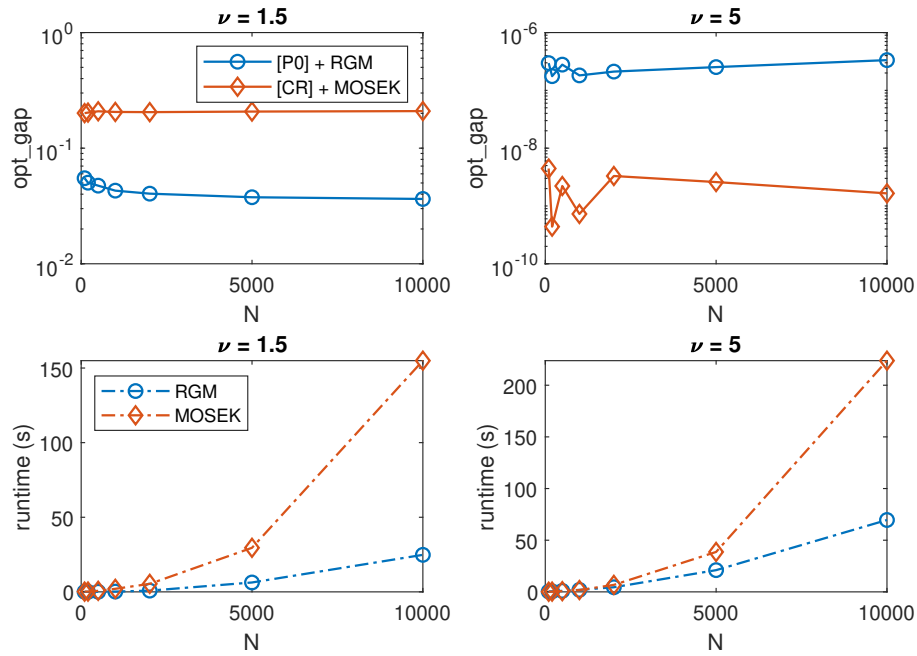


Fig. 1. Comparison between RGM and MOSEK package in two cases $\nu = 1.5$ and $\nu = 5$; here ν is the parameter associated with the cost vector $\mathbf{c} = \nu B^T \mathbf{1}$.

As we can see, when ν is small with smaller failure costs, the relaxation is not exact, and the RGM yields a solution with a smaller optimality gap (less than 5.5% for $\nu = 1.5$). When ν is large, the relaxation becomes tight, and both approaches provide (nearly) optimal solutions. In this case, MOSEK yields a slightly better solution since it uses an interior-point method. In terms of runtimes, RGM outperforms MOSEK, especially for large networks. The above

results suggest that RGM not only is scalable, but also can find a good solution to the original problem, if not (nearly) optimal.

5 Conclusion

We studied the problem of determining suitable investments in improving the robustness of complex systems. Unlike in previous studies, we considered investments in both resilience and recovery, while taking into account dynamics. The problem of minimizing the average costs of a system operator is formulated as an optimization problem, which is shown to be nonconvex. We then proposed two approaches to determining (nearly optimal) investments based on a gradient-based method and a convex relaxation. The effectiveness of the proposed approaches are demonstrated using numerical studies.

References

1. MOSEK ApS. *The MOSEK optim. toolbox for MATLAB manual. Version 9.0.*, 2019. <http://docs.mosek.com/9.0/toolbox/index.html>.
2. Christian Borgs, Jennifer Chayes, Ayalvadi Ganesh, and Amin Saberi. How to distributed antidote to control epidemics. *Random Structures & Algorithms*, 37(2):204–222, September 2010.
3. Reuven Cohen, Shlomo Havlin, and Daniel ben Avraham. Efficient immunization strategies for computer networks and populations. *Phys. Rev. Lett.*, 91(247901), December 2003.
4. Ashish R. Hota and Shreyas Sundaram. Interdependent security games on networks under behavioral probability weighting. *IEEE Control Netw. Syst.*, 5(1):262–273, March 2018.
5. Mohammad M. Khalili, Xueru Zhang, and Mingyan Liu. Incentivizing effort in interdependent security games using resource pooling. In *Proc. of NetEcon*, 2019.
6. Ali Khanafer, Tamer Başar, and Bahman Ghahsifard. Stability of epidemic models over directed graphs: A positive systems approach. *Automatica*, 74:126–134, 2016.
7. Howard Kunreuther and Geoffrey Heal. Interdependent security. *J. Risk and Uncertainty*, 26(2/3):231–249, 2003.
8. Richard J. La. Interdependent security with strategic agents and global cascades. *IEEE/ACM Trans. Netw.*, 24(3):1378–1391, June 2016.
9. Marc Lelarge and Jean Bolot. A local mean field analysis of security investments in networks. In *Proc. of International Workshop on Economics of Networked Systems*, pages 25–30, 2008.
10. Van Sy Mai, Abdella Battou, and Kevin Mills. Distributed algorithm for suppressing epidemic spread in networks. *IEEE Contr. Syst. Lett.*, 2(3):555–560, 2018.
11. Van-Sy Mai, Richard J. La, and Abdella Battou. Optimal cybersecurity investments for SIS model. In *Proc. of IEEE Globecom*, 2020.
12. Van-Sy Mai, Richard J. La, and Abdella Battou. Optimal cybersecurity investments in large networks using SIS model: algorithm design. *IEEE/ACM Trans. Netw.*, 29(6):2453–2466, December 2021.
13. Cameron Nowzari, Victor M Preciado, and George J. Pappas. Optimal resource allocation for control of networked epidemic models. *IEEE Control Netw. Syst.*, 4(2):159–169, June 2017.

14. Stefania Ottaviano, Francesco De Pellegrini, Stefano Bonaccorsi, and Piet Van Mieghem. Optimal curing policy for epidemic spreading over a community network with heterogeneous population. *J. Complex Networks*, 6(6), October 2018.
15. Victor M. Preciado, Michael Zargham, Chinwendu Enyioha, Ali Jadbabaie, and George J. Pappas. Optimal resource allocation for network protection against spreading processes. *IEEE Control Netw. Syst.*, 1(1):99–108, March 2014.