Failure Probability Constrained AC Optimal Power Flow

Anirudh Subramanyam, Jacob Roth, and Mihai Anitescu, Member, IEEE

Abstract—Despite cascading failures being the central cause of blackouts in power transmission systems, existing operational and planning decisions are made largely by ignoring their underlying cascade potential. This paper posits a reliability-aware AC Optimal Power Flow formulation that seeks to design a dispatch point which has a lower operator-specified likelihood of triggering a cascade starting from any single component outage. By exploiting a recently developed analytical model of the probability of component failure, our Failure Probability-constrained AC Optimal Power Flow (FP-ACOPF) utilizes the system’s expected first failure time as a smoothly tunable and interpretable signature of cascade risk. We use techniques from bilevel optimization and numerical linear algebra to efficiently formulate and solve the FP-ACOPF using off-the-shelf solvers. Simulations on the IEEE 118-bus case show that, when compared to the unconstrained and N-1 security-constrained ACOPF, our probability-constrained dispatch points can delay the expected onset of cascades and reduce the probability of long severe cascades, by more than ten and two orders of magnitude, respectively, while incurring less than 1% higher generation costs and load shedding.

Index Terms—cascading failures, AC optimal power flow

I. INTRODUCTION

A cascading failure in a power transmission system refers to a sequence of dependent outages of individual system components that successively disable parts of the grid, leading to a significant loss of served power or large blackout in the worst case. Accounts of major blackouts reveal that cascading failures are often triggered by an initial event that is largely unpredictable (such as extreme weather) but are sustained by subsequent events that are causally linked via Kirchhoff’s laws and automatic control actions of protection devices. For example, the outage of a single component can lead to redistribution of power flows in the remainder of the network in a way that can cause large overcurrents on some transmission lines. This, in turn, may trigger protection relays to disconnect these lines automatically if the current flow exceeds some threshold rating, or it may lead to eventual thermal failure if the overcurrents remain sustained for a long time.

The large direct and indirect costs associated with blackouts, along with mandatory standards set forth by the North American Electric Reliability Corporation (NERC) to address cascading outages, have motivated the development of a plethora of tools for the simulation and analysis of cascading failures; see reviews [1], [2] and references therein. These tools can be broadly classified as (a) complex networks approaches that consider the pure topological properties of power networks while ignoring or simplifying the underlying physics (e.g., see [3], [4]), (b) quasi-steady-state (e.g., see [5], [6]) or dynamic (e.g., see [7]–[9]) methodologies which combine physics-based DC or AC power flow models with Monte Carlo or enumerative sampling methods, and (c) high-level statistical models built either on historical/simulation data or on simplified power system physics (e.g., see [10], [11]).

While these tools have proved immensely useful in deepening our understanding of cascading failures, their use for risk mitigation and decision-making have been largely restricted to limiting the propagation of a cascading failure, e.g., via controlled load shedding or intentional islanding, after severe contingencies have already occurred (e.g., see [12]–[14]). Existing practices for the prevention of cascading failures before they occur, by tuning and modifying the controllable properties of the power network, has largely relied upon N−k security criteria and simulation-based contingency analyses as notional surrogates for reducing cascading likelihood.

Although simulation-based tools can influence long-term cascade mitigation solutions, such as line capacity or generator margin allocations and protection system enhancements (e.g., see [15]–[17]), they are fundamentally limited in preventing cascades in short-term operations such as economic dispatch or optimal power flow. This is because they do not provide any direct functional relationship between control parameters and cascade potential, or they entail expensive Monte Carlo sampling and numerical integration requirements and are thus challenging to incorporate within optimization algorithms.

This work proposes to incorporate in the classical ACOPF model, an analytical – as opposed to simulation-based – model of cascade severity that is an explicit function of the network properties and dispatch point. In contrast to existing methods, we aim to determine a dispatch point subject to the constraint that the probability of individual component failure remains below an operator-prescribed threshold. Our model capitalizes upon results from [18], which in contrast to other approaches for simulating cascading failures, provides an analytic expression for the failure probability as a function of the dispatch point. This is achieved by modeling Gaussian load and generation fluctuations in the AC power flow dynamics, and interpreting the latter as the diffusion of a particle in an energy landscape subject to stochastic forcing. Large deviations theory then provides the means to analytically relate the failure probability to the underlying energy surface.

For a given system state (consisting of voltages and power flows) at equilibrium, the analytical expression for the failure probability of an individual component requires solving a non-linear optimization problem that computes a “most likely” failure state starting from this equilibrium state. However, since
the latter is contingent on the dispatch point, explicitly con- straining the failure probability of an individual component, or of a cascading sequence of multiple components, is tantamount to solving a bilevel optimization problem. We settle for the former and demonstrate that constraining individual failure probabilities, which is equivalent to increasing the system’s expected first failure time—or decreasing its failure rate—has no significant increase in computational cost and operator conservatism, at the cost of increased combinatorial complexity. In contrast, $N - k$ approaches must resort to a non-smooth control of $k$ to achieve the same objective, while invariably incurring a sharp increase in combinatorial complexity. Another subtle, yet practically useful, advantage of our approach is its interpretability. Indeed, the benefit (in terms of reliability) of increasing $k$ in $N - k$ approaches, is difficult to convey outside the domain. Our approach, on the other hand, allows the system operator to decide between a failure rate limit of $10^{-6}$ s$^{-1}$ or $10^{-15}$ s$^{-1}$ (for example), which is equivalent to deciding between observing the first failure once every $10^6$ or $10^{15}$ seconds. This is a statement that is better aligned with the philosophy of regulatory constraints which tend to be in occurrences per unit of time.

This paper is organized as follows. Section II presents assumptions, and reviews the probabilistic failure model of [18]. Section III presents the failure probability-constrained ACOPF model along with its reformulation. Section IV demonstrates the empirical performance of our method and Section V offers concluding remarks and directions for future work.

II. Failure Probability Model

A. Notation

We use $\mathcal{N} = \{1, \ldots, n_b\}$ to denote the set of buses, and $\mathcal{L} \subseteq \mathcal{N} \times \mathcal{N}$ to denote the set of transmission lines, where $l = (i, j) \in \mathcal{L}$ is a line from bus $i$ to bus $j$. We denote the set of generators connected to bus $i \in \mathcal{N}$; note that $\mathcal{G} = \cup_{i \in \mathcal{N}} \mathcal{G}_i$. For ease of notation, we define $\mathcal{N}^c := \{i \in \mathcal{N} : \mathcal{G}_i = \emptyset\}$ to denote the set of non-generator buses, i.e., those that are not connected to any generator and similarly, $\mathcal{L}^c := \{(i, j) \in \mathcal{L} : \mathcal{G}_i = \emptyset \text{ or } \mathcal{G}_j = \emptyset\}$ to denote the set of lines that are connected to at least one non-generator bus. The nodal admittance matrix $Y = G + \sqrt{-1} B \in \mathbb{C}^{n_b \times n_b}$ with conductance and susceptance matrices $G \in \mathbb{R}^{n_b \times n_b}$ and $B \in \mathbb{R}^{n_b \times n_b}$, respectively.

We denote active and reactive power generations by $p_g, q_g \in \mathbb{R}^{n_b}$, and active and reactive power demands by $d_p, d_q \in \mathbb{R}^{n_b}$, and net active and reactive powers by $p_{net}, q_{net} \in \mathbb{R}^{n_b}$, where we define $p_{net,i} := p_{di} - \sum_{k \in \mathcal{G}} p_{g,k}$ and $q_{net,i} := q_{di} - \sum_{k \in \mathcal{G}} q_{g,k}$ for $i \in \mathcal{N}$. The voltage magnitudes and angles are denoted by $V \in \mathbb{R}^{n_b}$ and $\theta \in \mathbb{R}^{n_b}$.

For a vector $z \in \mathbb{C}^n$, we use $\|z\|$ and $z^\ast$ to denote its Euclidean norm and Hermitian transpose, respectively, and we use $e^z$ and $\log(z)$ to denote the vectors $(e^{z_1}, \ldots, e^{z_n}) \in \mathbb{C}^n$ and $(\log(z_1), \ldots, \log(z_n)) \in \mathbb{C}^n$. Given vectors $z, \tilde{z} \in \mathbb{C}^n$, we use $z \circ \tilde{z}$ to denote the Hadamard product $(z_1 \tilde{z}_1, \ldots, z_n \tilde{z}_n) \in \mathbb{C}^n$. For a matrix $A \in \mathbb{C}^{n \times n}$, we use $\rho(A)$, det$(A)$, and ad$(A)$ to denote its spectral radius, determinant and adjugate, respectively, and $A \succeq 0 \geq 0$ to indicate that it is positive definite (semi-definite). For a given matrix $A \geq 0$ and vector $z \in \mathbb{C}^n$, we use $\|z\|_A$ to denote $\sqrt{z^\ast A z}$.

In formulating the probability model, it will be convenient to divide the power generations and voltages $(p_g, q_g, V, \theta) \in \mathbb{R}^{n_b}$ ($m := 2n_g + 2n_b$) into two distinct sets. After selecting an arbitrary generator bus $\sigma \in \mathcal{N} \setminus \mathcal{N}'$ as the slack bus in the steady-state context and reference bus in the dynamics context, we collect all voltage magnitudes at non-generator buses and voltage angles at non-slab buses in the state vector $x = (\{V_i\}_{i \in \mathcal{N}}, \{\theta_i\}_{i \in \mathcal{N} \setminus \{\sigma\}}) \in \mathbb{R}^d$, $d = |\mathcal{N}'| + |\mathcal{N}'| - 1$. The remaining voltage components and power generations are aggregated in the vector $y = (\{V_i\}_{i \in \mathcal{N} \setminus \mathcal{N}'}, \theta_\sigma, p_g, q_g) \in \mathbb{R}^{m-d}$.

B. Assumptions

Motivated by the automatic control actions of protection relays, we assume that individual system components fail in a deterministic manner according to a component-specific state equation. Specifically, we assume that a component, such as a transmission line $l \in \mathcal{L}$, fails if the value of a certain function $\Theta_l : \mathbb{R}^m \rightarrow \mathbb{R}$ at the current state $(x, y)$ exceeds a critical value $\Theta_l^{max}$. Furthermore, we assume that after failure, a component remains failed over the entire dispatch horizon of the ACOPF.

The algebraic state functions $\Theta$ can be used to model various component failures, such as under-voltage conditions at buses or line surges in transformers. For the remainder of the paper, however, we assume that component failures occur only due to current overloads in transmission lines $l = (i, j) \in \mathcal{L}$ beyond an emergency rating $I_l^{max}$: (here, $\Theta_l^{max} := (I_l^{trip})^2$)

$$\Theta_l(x, y) = |Y_{ij}|^2 (V_i^2 + V_j^2 - 2V_i V_j \cos(\theta_i - \theta_j))$$

This is partly because it simplifies the exposition, and partly because the corresponding failure model has already been studied and validated against real cascade data in [18], [19]. We note, however, that the methodology does not preclude us from considering other failure mechanisms.
In addition, we shall assume that (i) the network is lossless\(^1\), i.e., \( G = 0 \), (ii) transmission line failures are induced only by fluctuations in active and reactive power demand and active power generation (rather than exogenous events), and therefore, (iii) only the subset \( L' \) of transmission lines that connect to at least one non-generator bus have a nonzero likelihood of failing over the dispatch horizon of the ACOPF.

C. Analytical model of line failures

We model the grid’s electro-mechanical behavior using a system of stochastic differential equations (SDE), by introducing a scalar noise parameter \( \tau > 0 \) into the following variant\(^2\) of the standard, structure-preserving model [21]–[25]:

\[
dx_l^T = (J - S) \nabla_x \mathcal{H}(x_l^T, y_l) dt + \sqrt{2\tau \Sigma_l} dW_l.
\]  

(2)

Here, \( x_l^T \in \mathbb{R}^d \) denotes the system state at time \( t \), \( J \) and \( S \) represent appropriate system matrices [18], \( W_l \) is a \( d \)-dimensional vector of independent Wiener processes, and \( \mathcal{H} \) is the following energy function [18], [19], [21] that is obtained as the first integral of the deterministic dynamics\(^3\):

\[
\mathcal{H}(x, y) := \frac{1}{2} (V \circ e^{i\theta})^\ast Y (V \circ e^{i\theta}) + \sum_{l \in \mathcal{L}'} p_{l,\text{net}}^T \theta + q_{l,\text{net}}^T \log(V)
\]  

(3)

It is worth pointing out that (3) is closely related to the widely-used transient stability analysis energy function [20]; similar to the latter, it is used to measure distance to instability, we use (3) to assess the probability of the stochastically perturbed dynamics reaching a triggering surface.

We note that the vector \( y \) is assumed to have been fixed to appropriate values \( a \) priori, and we shall return to choosing a value for \( y \) in the subsequent section when we describe the ACOPF formulation. For now, observe that the structure of the energy function (3) ensures that, for any specified value of \( y \), any point \( x \in \mathbb{R}^d \) that satisfies \( \nabla_x \mathcal{H}(x, y) = 0 \) is a solution to the lossless power flow equations:

\[
p_{\text{net},i} + \sum_{j \in \mathcal{N}} B_{ij} V_i V_j \sin(\theta_i - \theta_j) = 0, \quad i \in \mathcal{N},
\]  

(4)

\[
q_{\text{net},i} - \sum_{j \in \mathcal{N}} B_{ij} V_i V_j \cos(\theta_i - \theta_j) = 0, \quad i \in \mathcal{N}.
\]  

(5)

In particular, any local energy minimizer,

\[
\bar{x}(y) \in \arg \min_{x \in \mathbb{R}^d} \mathcal{H}(x, y)
\]  

(6)

defines a feasible solution to (4), (5) and serves as a stable equilibrium for the network dynamics defined by \( y \). To simplify notation, we drop the dependence of \( \bar{x} \) on \( y \) in the remainder of the paper.

Under the SDE model (2) initialized at \( x_l^0 = \bar{x} \), the failure probability of a line \( l \in \mathcal{L}' \) (considered in isolation from the rest of the network) can be quantified by its so-called \textit{first passage time}. The latter is defined to be the first time at which the system state \( x_l^T \) triggers the failure condition (1) for line \( l \). Assuming \( \Theta_l(x, y) < \Theta_l^{\text{max}} \), this is equivalent to

\[
T_l^*(y) := \inf \{ t > 0 : \Theta_l(x_l^T, y) \geq \Theta_l^{\text{max}} \}.
\]  

(7)

By exploiting results from large deviations theory [27], [28], it was shown in [18] that, as \( \tau \to 0 \), \( T_l^*(y) \) is an exponential random variable whose mean satisfies the following relation:

\[
\lim_{\tau \to 0} \tau \log(\mathbb{E}[T_l^*(y)]) = \min_{x, \Theta_l(x_l^T, y) = \Theta_l^{\text{max}}} \mathcal{H}(x, y) - \mathcal{H}(\bar{x}, y)
\]  

(8)

A point at which the minimum is obtained is called the \textit{most likely failure point} and is defined as

\[
x_l^*(y) := \arg \min_{x \in \mathbb{R}^d} \{ \mathcal{H}(x, y) : \Theta_l(x, y) = \Theta_l^{\text{max}} \},
\]  

(9)

where we have implicitly assumed \( x_l^* \) to be unique\(^4\). As before, we drop the dependence of \( x_l^* \) on \( y \) to simplify notation.

For finite values of \( \tau \), we can thus use relation (8) to compute a log-approximate \textit{failure rate} (i.e., the reciprocal of the mean failure time). In addition to this log-approximation, a subexponential correction to the failure rate was also obtained in [18] yielding the following expressions:

\[
\lambda_l^*(y) := \text{pf}_l^1(y) \times \text{ef}_l(y),
\]  

(10)

\[
\text{pf}_l^1(y) := \text{pf}_l^0(y) \times \left( 1 + \frac{\tau}{\mathcal{H}(x_l^* - y) - \mathcal{H}(\bar{x}, y)} \right),
\]  

(11)

\[
\text{pf}_l^0(y) := \left\| \nabla_x \mathcal{H}(x_l^*, y) \right\|^2 \sqrt{\frac{\det \nabla^2_x \mathcal{H}(x_l^*, y)}{2\pi \tau C_l^*(y)}},
\]  

(12)

\[
\text{ef}_l(y) := \exp \left[ -\frac{\mathcal{H}(x_l^*, y) - \mathcal{H}(\bar{x}, y)}{\tau} \right],
\]  

(13)

where \( C_l^*(y) \) is a factor accounting for the curvature of the failure surface in the vicinity of \( x_l^* \), and is given by:

\[
C_l^*(y) := \nabla^2_x \mathcal{H}(x_l^*, y) \text{adj} (X_l(y)) \nabla_x \mathcal{H}(x_l^*, y),
\]  

(14)

\[
X_l(y) := \nabla^2_x \mathcal{H}(x_l^*, y) - \mu_l^T \nabla^2_x \Theta_l(x_l^*, y),
\]  

(15)

where \( \mu_l^T \in \mathbb{R} \) is the optimal Lagrange multiplier in the constrained optimization problem (9).

In summary, the distribution of failure times for line \( l \in \mathcal{L}' \) can be well-approximated by \( T_l^*(y) \sim \text{Exp}(\lambda_l^*(y)) \), and therefore, the probability of observing line \( l \) fail in the first \( t \) seconds can be given by

\[
P(T_l^*(y) \leq t) = 1 - \exp[-\lambda_l^*(y)t].
\]  

(16)

III. Failure Probability-constrained AC Optimal Power Flow Formulation

A. Conceptual formulation

The probability of transmission line failure (16) is a function of \( y = \{ (V_i)_{i \in \mathcal{N} \setminus \mathcal{N}_1}, \theta_i, p_{g_i}, q_{g_i} \} \), the voltage magnitudes at generator buses, the voltage angle at the slack bus and the active and reactive power generations. The choice of \( y \) also determines the equilibrium operating point \( \bar{x}(y) \) through the power flow equations (4), (5); we shall denote these simply

\(^4\) This is not a strong assumption based on the validation in [18].
as $x = (\{V_i\}_{i \in \mathcal{N}}, \{\theta_i\}_{i \in \mathcal{N}' \setminus \{x\}})$ and along with the vector $y$, they constitute decision variables in our Failure Probability-constrained ACOPF (FP-ACOPF) formulation:

$$
\begin{align}
\min_{x,y} & \sum_{k \in \mathcal{G}} c_k(p_{g,k}) \\
\text{subject to} \quad & (4) - (5), \\
& p_{g,k}^{\min} \leq p_{g,k} \leq p_{g,k}^{\max}, \quad k \in \mathcal{G}, \\
& q_{g,k}^{\min} \leq q_{g,k} \leq q_{g,k}^{\max}, \quad k \in \mathcal{G}, \\
& V_{\min}^{\max} \leq V_i \leq V_{\max}^{\max}, \quad i \in \mathcal{N}, \\
& \Theta_l(x, y) \leq (I_l^{\lim})^2, \quad l \in \mathcal{L}, \\
& \lambda_l^\star(y) \leq \lambda_l^{\lim} := -I_l^{-1} \log(1 - \epsilon_l^{\lim}), \quad l \in \mathcal{L}'.
\end{align}
$$

The objective function (17) minimizes the cost of generation where $c_k$ is a function representing the cost of generating active power $p_{g,k}$ at generator $k \in \mathcal{G}$. The constraints (18), (19) and (20) ensure that the power generations and voltages magnitudes stay within predefined limits. The constraint (21) limits the amount of current flow on the transmission lines, where we have distinguished the flow limit $I_l^{\lim}$ from the value $I_l^{\text{trip}}$ used in the failure model (1). The latter quantity is the value at which protection relays would automatically disconnect the line, and it is typically much larger than the former, which is determined based on thermal considerations.

Finally, constraint (22) imposes an upper bound on the failure rate of line $l \in \mathcal{L}'$ failing over a time horizon $t_H$ is less than some operator-prescribed limit $\epsilon_l^{\lim}$ (say 1%):

$$
\mathbb{P}(T_l^\star(y) \leq t_H) \leq \epsilon_l^{\lim}, \quad l \in \mathcal{L}'.
$$

We highlight that constraint (23) can be used as a safe approximation for limiting the probability of (i) non-independent failures (using Bonferroni’s inequality), as well as (ii) cascading failures involving more than one line, over the horizon $t_H$. Indeed, while it is possible to constrain the exact probability of any cascading failure sequence of lines to stay below a certain risk threshold (since failure times are exponential random variables), the left-hand side of (23) can also be viewed as an upper bound on the probability of all such failure sequences that are initiated with the tripping of line $l$.

The presented model is a standard ACOPF formulation with the exception of the failure rate constraint (22). We now present an efficient reformulation of these rate constraints that can be incorporated into standard optimization solvers.

### B. Reformulation of failure rate constraints: Key steps

Constraining the failure rate via (22) is nontrivial because it involves (i) the solution of a nested nonconvex optimization problem (9) to obtain $x_l^\star(y)$, and (ii) constraints involving determinants of large matrices, namely $\nabla^2_{xx} \mathcal{H}(x, y)$ and $X_l(y)$, see (12) and (14). This section shows that these can be circumvented by exploiting the low-rank property of the line failure function $\Theta_l$, and a partial Taylor expansion of the energy function $\mathcal{H}$ around the equilibrium operating point $x$.

1. **Low rank factorization of failure function**: We shall capitalize on the fact that, the Hessian of the failure function $\Theta_l$ admits a low-rank factorization of the form:

$$
\nabla^2_{xx} \Theta_l(x, y) = Q_l(x, y) K_l(x, y) Q_l^\top(x, y),
$$

where $Q_l(x, y)$ is a $d \times r_l$ matrix, with $r_l \ll m$ and $K_l(x, y)$ is a $r_l \times r_l$ diagonal matrix. Observe that this low-rank structure is natural whenever $\Theta_l$ is a function of a small number of state variables, as is the case when $\Theta_l$ models the failure of transmission lines, system buses, generators or transformers.

Although such a low-rank factorization can be computed numerically (e.g., using an eigenvalue decomposition), it can also be obtained analytically in several cases, such as in the examples of Section II-B. To illustrate this, Appendix A presents explicit closed-form expressions for $Q_l$ and $K_l$ when $\Theta_l$ is the line overcurrent function (1). In particular, it shows that $r_l = 3$ for a line $l = (i, j)$ where both $i, j \in \mathcal{N}'$ are non-generator buses, whereas $r_l = 2$ otherwise. The fact that the rank $r_l \leq 3$ has nothing to do with the specific functional form of (1); in fact, $r_l \leq 3$ will always hold whenever $\Theta_l$ is a function only of $V_i, V_j$ and the angle difference $\theta_i - \theta_j$.

2. **Taylor approximation of energy function**: The value of the energy function at the most likely failure point, $\mathcal{H}(x_l^\star, y)$, is approximated using a quadratic Taylor polynomial centered around the equilibrium operating point $x$ that is determined by the ACOPF formulation:

$$
\mathcal{H}(x_l^\star, y) = \mathcal{H}(x, y) + \frac{1}{2}(x_l^\star - x)^\top \nabla^2_{xx} \mathcal{H}(x, y)(x_l^\star - x)
$$

where we have ignored the first-order term since the FP-ACOPF constrains $x$ to satisfy the power flow equations (4), (5), which are equivalent to $\nabla^2_x \mathcal{H}(x, y) = 0$. We shall comment on the accuracy of (25) in the next subsection.

### C. Reformulation of the 'most likely failure point' problem

Equations (24) and (25) can be used to reformulate the nested nonconvex most likely failure point optimization problem (9). In particular, the following proposition shows that they can be used to generate tractable algebraic constraints under which the solution of (9) can be characterized by its first- and second-order optimality conditions.

**Proposition 1.** **Fix a line** $l \in \mathcal{L}'$, and assume that $x, x_l^\star \in \mathbb{R}^d$ and $y \in \mathbb{R}^{m-d}$ are such that:

1. $\nabla^2_{xx} \mathcal{H}(x, y) \succ 0$, and
2. the Taylor approximation (25) is applicable.

If $x_l^\star$ and $\mu_l^\star \geq 0$ satisfy (26)–(28), then $x_l^\star$ is a local solution of the optimization problem (9).

$$
\begin{align}
\nabla^2_{xx} \mathcal{H}(x, y)(x_l^\star - x) &= \mu_l^\star \nabla_x \Theta_l(x_l^\star, y) \\
\Theta_l(x_l^\star, y) &= \Theta_l^{\max} \\
\mu_l^\star \rho(A_l) &< 1,
\end{align}
$$

where $A_l := K_l(x_l^\star, y) Q_l(x_l^\star, y) [\nabla^2_{xx} \mathcal{H}(x, y)]^{-1} Q_l(x_l^\star, y)$ is a $r_l \times r_l$ matrix.

**Proof.** See Appendix B.

Equations (26) and (27) are the first-order optimality conditions, whereas the spectral inequality (28) is derived from the
second-order optimality condition of the nested problem (9). It is noteworthy that the matrix \( A_l \) in the left-hand side of the latter inequality is a small \( r_l \times r_l \) matrix, and hence it is possible to reformulate the inequality in a completely algebraic form without requiring a numerical eigenvalue computation.

When \( r_l = 2 \), it can be verified (e.g., using the characteristic polynomial of \( A_l \)) that inequality (28) is equivalent to:

\[
\mu_l^2 \operatorname{tr}(A_l) - (\mu_l^2)^2 \det(A_l) < 1,
\]

where the trace and determinant can be computed in algebraic closed-form, since \( A_l \) is a \( 2 \times 2 \) matrix.

When \( r_l = 3 \), it is nontrivial to get a tractable reformulation, since \( A_l \) need not be symmetric. We consider two cases.

- If \( K_l(x_l^*, y) \geq 0 \), then \( \mu_l^2 \rho(A_l) < 1 \) is equivalent to \( \mu_l^2 \rho(A_l) < 1, \hat{A}_l := K_l^T Q_l^{-1} [\nabla_x^2 \mathcal{H}(x, y)]^{-1} Q_l K_l^T \). This is a symmetric matrix, and hence, we can use Sylvester’s criterion to enforce that all \( r_l \) leading principal minors of \( I - \mu_l^2 A_l \) (which are computable in algebraic closed-form) must be positive. A similar trick can be used if \( K_l(x_l^*, y) \) is negative definite.

- If \( K_l(x_l^*, y) \) is indefinite, then it is difficult to enforce the spectral inequality, and we settle for a relaxation which enforces a constraint on the sum of its \( r_l \) eigenvalues:

\[
\mu_l^2 \operatorname{tr}(A_l) < r_l,
\]

where again the trace can be computed in closed-form.

We close with some remarks on the assumptions of Proposition 1. The first assumption requires that the Hessian of the energy function at the equilibrium point \( x \) is positive definite. Note that it is satisfied whenever \( x \) is obtained via the energy minimization problem (6). Although using the power flow equations (4), (5) as a surrogate for the latter cannot guarantee this condition, we found that it was almost never violated in our experiments. With regards to the second assumption, the Taylor approximation (25) is guaranteed to be accurate only when \( x_l^* \) is near \( x \). We provide empirical evidence in Section IV that the approximation is not severe.

D. Treatment of high-dimensional determinants

Using equations (10) and (11), the failure rate constraint (22) can be equivalently written as:

\[
\lambda_l^T(y) = pf_l^1(y) \cdot ef_l(y) \cdot \left(1 + \frac{\tau}{\mathcal{H}(x_l^*, y) - \mathcal{H}(x, y)}\right) \leq \lambda_l^\text{lim}.
\]

The formula for the prefactor \( pf_l^1(y) \) that appears in the above constraint involves determinants of (the typically large) \( d \times d \) matrices, \( \nabla_{xx}^2 \mathcal{H}(x, y) \) and \( X_l(y) \), via (12) and (14) respectively. The presence of these determinants can slow down computation of the rate constraints and their gradients, and may result in ill-conditioning. Fortunately, equations (24) and (25) can be used to circumvent these issues. The following proposition uses the matrix-determinant lemma [29] to obtain a reformulation of the prefactor that avoids large determinants.

**Proposition.** Fix a line \( l \in \mathcal{L} \), and assume that \( x, x_l^* \in \mathbb{R}^d \) and \( y \in \mathbb{R}^{m-d} \) satisfy the conditions of Proposition 1. If \( x_l^* \) and \( \mu_l^2 \geq 0 \) satisfy (26)–(28), then the prefactor and energy factor in equation (10) admit the following reformulation:

\[
pf_l^1(y) = pf_l^1(y) \left(1 + \frac{2\tau}{\mu_l^2 \alpha_l}\right)
\]

\[
pf_l^0(y) = \left(\frac{\mu_l^*}{\mu_l^*} \right)^{3/2} \left\| \nabla_x \Theta_l(x_l^*, y) \right\|^2_{2} \sqrt{2\pi r \alpha_l \det(W_l)} + \beta_l)
\]

\[
ef_l(y) = \exp \left(-\frac{\mu_l^2 \alpha_l}{2\tau}\right).
\]

where \( W_l := I - \mu_l^2 A_l \) is a \( r_l \times r_l \) matrix, \( A_l \) is as in Proposition 1, and the scalars \( \alpha_l := (x_l^* - x)^T \nabla_x \Theta_l(x_l^*, y) \) and \( \beta_l := (x_l^* - x)^T Q_l(x_l^*, y) \operatorname{adj}(W_l) K_l(x_l^*, y) Q_l^T(x_l^*, y) (x_l^* - x) \).

**Proof.** See Appendix C.

Unlike equations (12) and (14), equation (32) involves the determinant and adjugate of the small \( r_l \times r_l \) matrix \( W_l \). This allows efficient computation of the scalar \( \beta_l \), and hence, of the overall failure rate and its gradient. Moreover, it avoids the ill-conditioning that plagues the original constraint, and allows expressing it in an algebraic optimization system.

E. Efficient practical implementation

We now highlight other key algorithmic enhancements that are necessary to improve the practical performance of the FP-ACOPF. First, since the absolute value of the failure rates \( \lambda_l^T(y) \) can be very small, we found it beneficial to implement the rate constraint (22) in its log-form. Using equations (10), and (31)–(33), this is equivalent to:

\[
\frac{3}{2} \log(\mu_l^2) + \log \left(\left\| \nabla_x \Theta_l(x_l^*, y) \right\|^2_{2}\right) - \frac{1}{2} \log(\alpha_l \det(W_l) + \beta_l) + \log \left(1 + \frac{2\tau}{\mu_l^2 \alpha_l}\right) - \frac{\mu_l^2 \alpha_l}{2\tau} \leq \log(\sqrt{2\pi r \lambda_l^\text{lim}}).
\]

All intermediate terms involved in the left-hand side of (34) are computable in closed form, except for the matrix \( A_l \) which appears in the definition of \( W_l := I - \mu_l^2 A_l \). A close examination of the formula for \( A_l \) in Proposition 1 reveals that this matrix can be efficiently computed by first solving for the intermediate matrix \( Z_l = \mathbb{R}^{d \times r_l} \) as follows:

\[
\nabla_{xx}^2 \mathcal{H}(x, y) \cdot Z_l = Q_l(x_l^*, y),
\]

and then setting \( A_l = K_l(x_l^*, y) Q_l^T(x_l^*, y) Z_l \). We therefore introduce explicit decision variables \( Z_l \), along with (35) as constraints in the FP-ACOPF formulation. Notably, this allows the complete rate constraint (34), including all of its intermediate expressions, to be computed algebraically.

Implementing the rate constraint (22) for a particular line \( l \in \mathcal{L} \) thus involves introducing \( dr_l + d + 1 \) additional variables, namely \( Z_l \in \mathbb{R}^{r \times r_l} \), \( x_l^* \in \mathbb{R}^d \) and \( \mu_l^2 \geq 0 \) and \( dr_l + d + 3 \) additional constraints, namely (26)–(28) and (34)–(35). In practice, the rate constraints (34) are binding for only a small fraction of critical lines. Therefore, adding \( O(dr_l) \) constraints for the remaining non-critical lines unnecessarily increases model complexity. To limit this growth, we add additional variables and constraints in an iterative manner, only for those lines \( l \in \mathcal{L} \) whose rate constraints (34) are found to be violated. Our algorithm can be described as follows.
Step 1: Set $\hat{\mathcal{L}} = \emptyset$.

Step 2: Define decision variables $x \in \mathbb{R}^d$, $y \in \mathbb{R}^{m-d}$, and $\{(x^*_l, \mu^*_l, Z_l) \in \mathbb{R}^d \times \mathbb{R}_{>0} \times \mathbb{R}^{d \times r_l} \}$ for each $l \in \hat{\mathcal{L}}$. Compute an optimal solution $\hat{z}$ of the problem:

\[
\text{minimize} \quad (17) \\
\text{subject to} \quad (4), (5), (18) - (21), \\
(26) - (28), (34), (35) \quad \forall l \in \hat{\mathcal{L}}.
\]

Step 3: For all $l \in \mathcal{L}' \setminus \hat{\mathcal{L}}$ (possibly in parallel):

a) Use $\hat{z}$ to compute candidate values of $x^*_l, \mu^*_l, Z_l$ by solving equations (26)-(28), and (35).

b) If inequality (34) is violated using $\hat{z}$ and the computed values of $x^*_l, \mu^*_l, Z_l$, update $\hat{\mathcal{L}} = \mathcal{L} \cup \{l\}$ and save the computed values as warm-starts.

Step 4: If $\hat{\mathcal{L}}$ was updated, go to Step 2; else, stop and output $(\hat{x}, \hat{y})$ as the optimal FP-ACOPF solution.

IV. NUMERICAL SIMULATIONS

We now present results and insights obtained from several experiments.\(^5\) The IEEE 118-bus test system with line limits $I^\text{lim}$ obtained from PGLib \([30]\), is used in all experiments. We set branch and shunt conductances to zero to satisfy the voltage magnitudes and angles, respectively. The IEEE 118-bus test system with line limits $I^\text{lim}$ obtained from PGLib \([30]\), is used in all experiments. The numerical performance of the FP-ACOPF is largely unaffected by the choice of the noise parameter $\tau$, the line failure threshold $I^\text{trip}$, and the network loading $(p_d, q_d)$ levels. Therefore, Section IV-A uses ‘baseline’ values of $\tau = 10^{-4}$, $I^\text{trip} = 1.02 I^\text{max}$ and $(p_d, q_d)$ values from MATPOWER \([32]\). Their values can, however, strongly affect cascading behavior; therefore, Section IV-B includes a discussion of cascading sensitivity to their values.

A. Computational efficiency of FP-ACOPF

For a fixed dispatch point $y$, the true line failure rates $\lambda^T\gamma(y)$ are computed by solving problem (9) to obtain the most likely failure point $x^*_T$ and its multiplier $\mu^*_T$. However, the FP-ACOPF computes an estimate, which we denote as $\hat{\lambda}^T(y)$, by solving (26)-(28) instead of (9). Fig. 1a shows the corresponding approximation error for the subset of lines with the highest failure rates. We observe that the relative log-error in the approximation is less than $10^{-2}$. Fig. 1b shows that lines with higher utilization of their flow capacity, defined as $\sqrt{T \gamma(y)} / I^\text{lim} \times 100\%$, do not necessarily have higher failure rates, thus illustrating that the magnitude of line flow is not an effective surrogate for its failure rate. For example the line with the second-largest capacity utilization has a failure rate in the middle of the range, whereas lines with the largest failure rates have capacity utilization close to $50\%$. For accuracy, we must thus work directly with the failure rates via (10).

\(^5\)Our Julia code is available at: github.com/jacob-roth/OPF. All optimization problems were solved using Ipopt via JuMP. The runs were performed on 32 threads of an Intel Xeon Gold 6140 CPU at 2.30GHz with 512 GB RAM.

\(^6\)For some context on this assumption, the mean branch resistance/reactance ratio is approximately 0.2. Also, the mean difference between the ACOPF solutions computed with and without resistances is 0.004 p.u. and 0.04 for the voltage magnitudes and angles, respectively.

Table I summarizes the computational time and generation cost of the failure probability-constrained and $N-k$ security-constrained\(^7\) dispatch points. We observe that the FP-ACOPF solves within an order of magnitude of the time it takes to solve the $N-0$ ACOPF. On the other hand, the $N-1$ model is slower by almost three orders of magnitude, although this may be because all contingencies are included in a single optimization problem. In terms of generation cost, including the rate constraints results in a cost increase of less than 1% compared to the $N-0$ model, even with extremely stringent rate limits, without requiring any load shedding.

B. Cascading failure simulations of FP-ACOPF dispatch

To compare the cascade potentials of the failure probability-constrained and $N-k$ security-constrained ACOPF, we perform 1,000 Kinetic Monte Carlo (KMC) simulations \([18]\) per dispatch point. The latter simulates cascading sequences of line failures for a given dispatch point, using the failure rate expression in (10). Similar to other quasi-steady-state models, it assumes that the system ‘re-equilibrates’ immediately following an individual line failure; i.e., the post-failure system state satisfies the power flow equations and hence, is a local minimizer (6) of the energy function, appropriately modified (via the admittance matrix) to account for the degraded network topology. Following each line failure, the slack bus is assumed to maintain power balance in the network, and buses that no longer connect to the slack bus (under the modified network topology) are ‘deactivated’. The

\(^7\)The set of contingencies is comprised of all lines, and post-contingency generation is allowed to vary by at most $0.1 P^\text{max}_g$ at each generator $k \in \mathcal{G}$.
cumulative active demand $p_d$ at deactivated buses serves as a measure of lost load. This enables fair comparison of different dispatch points, on the basis of time and severity until full system collapse or until a very large simulation time ($10^{40}$ sec in our implementation) has elapsed.

Fig. 2 summarizes the simulation performance of the FP-ACOPF and $N-1$ security-constrained ACOPF, relative to the $N-0$ dispatch, for baseline values of $\tau = 10^{-4}$, $I_t^{\text{trip}} = 1.02 I_t^{\text{max}}$ and nominal MATPOWER ($p_d, q_d$). First, in Fig. 2a, we group the 1,000 simulated line failure sequences into individual cascades as per [33] and then report the distribution of cascades as a function of the number of failed lines. We observe not only an approximate power law distribution (which serves as validation), but also—and more importantly—that the FP-ACOPF reduces the frequency of longer, more severe cascades. Unlike the $N-1$ model, it can be tuned via the rate limits $\lambda^{\text{lim}}$, and for sufficiently stringent $\lambda^{\text{lim}} = 10^{-15}$, it can prevent the occurrence of long cascades entirely.

To study this phenomenon more closely, in Figs. 2b and 2c, we plot the observed distribution (over the 1,000 failure sequences) of elapsed time and total demand lost, as a function of number of line failures, respectively. These figures display a distribution of line failure sequences from separate KMC runs, in contrast to Fig. 2a which separates failure sequences into subsequences (cascades) based on the time between successive failures and plots the aggregate count of subsequences of a certain length. In these figures, one method dominates another if its failure time distribution in Fig. 2b lies above, and if its load lost distribution in Fig. 2c lies below that of the other. Fig. 2b shows that, by delaying the system’s first failure time, the FP-ACOPF dispatch points are able to successfully delay—and hence, prevent—subsequent line failures which could lead to system collapse. Interestingly, all dispatch points display a long plateau of rapid consecutive failures following the first few failures, indicating dangerous system instability. The FP-ACOPF models seem to delay the onset of such instability and are thus able to find more ‘stable’ dispatch points; for example, the onset of the plateau in $\lambda^{\text{lim}} = 10^{-15}$ is delayed by almost ten orders of magnitude, compared to $N-1$. Moreover, Fig. 2c shows that the fraction of load lost decreases almost monotonically with $\lambda^{\text{lim}}$, indicating that reduced cascading risk is not accompanied with increased load shedding.

To study the sensitivity of our results to the choice of the noise parameter $\tau$, line failure threshold $I_t^{\text{trip}}$ and loads ($p_d, q_d$), we repeated our analysis by varying them as follows: $\tau \in \{2^{-1}, 1, 2\} \times 10^{-4}$, $I_t^{\text{trip}} \in \{1.02, 1.05\} \times I_t^{\text{max}}$ and $\{p_d, q_d\} \in \{0.9, 1, 1.1\} \times$ nominal MATPOWER values.8,9 We found that the overall results were largely unaffected by the choice of $I_t^{\text{trip}}$, but were sensitive to the noise parameter $\tau$ and load levels ($p_d, q_d$), with cascade severity increasing monotonically with $\tau$ as well as ($p_d, q_d$). Moreover, we observed that the FP-ACOPF generally outperformed the $N-1$ ACOPF dispatch, both in terms of cascade severity and timing, across the range of parameter values we studied, and therefore, do not present detailed results for the sake of brevity. Importantly, this observation alleviates some of the practical concerns over specifying the values of these parameters precisely.

V. CONCLUSION AND FUTURE WORK

This paper takes the first steps towards quantifying and proactively reducing failure risk, and, implicitly, cascading risk in operational dispatch. Our empirical results demonstrate that the failure rate limit $\lambda^{\text{lim}}$ serves as an accurate, monotone and tunable metric of cascade severity, and the FP-ACOPF can strongly outperform classical $N-k$ models in terms of both the onset as well as severity of line failures, without incurring significantly larger computational or economic costs.

We envision future work along several directions. From a methodological viewpoint, we need to extend the failure probability model and the corresponding optimization model to lossy networks. From a modeling viewpoint, more general multiple-component failure rate constraints, as well as extensions that combine classical $N-1$ models with rate constraints at the inner level need to be investigated. From a practical viewpoint, the model parameters (e.g., $\tau$) need to be calibrated and the model itself needs to be further validated and compared against established cascading simulators on real network data, with some initial steps already being taken in [18]. We believe these extensions can open up several other use cases for our method including long-term planning decisions, line capacity allocations, and contingency screening.

APPENDIX

A. Closed-form expressions of low-rank factors

Consider the line overcurrent function (1) for a line $l = (i, j)$. First, consider the case where both $i, j \in \mathcal{N}^l$ are non-generator buses, so that $V_i, V_j, \theta_i, \theta_j$ are all part of the state vector $x$. In this case, $r_1 = 3$ and only the rows of $Q_l$ corresponding to these components have non-zero elements. Suppose $1_w \in \mathbb{R}^d$ is the canonical unit vector with 1 in the $w$-component and 0 otherwise. Then, it can be verified that $\nabla^2_{x,x} Q_l = Q_l K_l Q_l^\top$, where $K_l = \text{diag} (1, 1, -1)$ and the matrix $Q_l$ is as follows:

1) $Q_1 = 1_w = 0 \in \mathbb{R}^3$ for all $w \notin \{V_i, V_j, \theta_i, \theta_j\}$,

2) $Q_1 = \begin{bmatrix} 0 & V_j s_{ij} & -V_j s_{ij} & -V_j s_{ij} \\ 0 & V_i + V_j c_{ij} & -V_i - V_j c_{ij} & -V_i - V_j c_{ij} \\ 0 & \sqrt{V_i^2 + V_j^2 + V_i V_j c_{ij}} & -\sqrt{V_i^2 + V_j^2 + V_i V_j c_{ij}} & -\sqrt{V_i^2 + V_j^2 + V_i V_j c_{ij}} \\ c_{ij} := \cos(\theta_i - \theta_j), & s_{ij} := \sin(\theta_i - \theta_j) \end{bmatrix}$.

If $j \in \mathcal{N}^l \setminus \{ \mathcal{N}^l \cup \{i\} \}$ is a non-slack generator bus, then $V_i, \theta_i, \theta_j$ are part of $x$, but not $V_j$. In this case, $r_1 = 2$, $K_l = \text{diag} (1, V_i V_j c_{ij} - V_j^2 s_{ij}^2)$ and, along with condition 1) described above, the matrix $Q_l$ is characterized by:

$Q_l = \begin{bmatrix} 1 & V_i s_{ij} & -V_i s_{ij} \\ 0 & V_i + V_j c_{ij} & -V_i - V_j c_{ij} \\ 0 & V_i s_{ij} & -V_i s_{ij} \\ 1 & -1 \end{bmatrix}$
If \( j = \sigma \) is the slack bus, then \( r_l = 2 \), and the above expressions for \( K_l \) and \( Q_l \) continue to be applicable, except \( \theta_j \) is not part of \( x \) and has no corresponding column in \( Q_l \).

**C. Proof of Proposition 2**

We again drop the subscript \( l \) from all quantities to simplify the notation. Using equations (25) and (26), we obtain:

\[
\mathcal{H}(x^*, y) - \mathcal{H}(x, y) = \frac{1}{2} (x^* - x)^\top \nabla^2_{xx} \mathcal{H}(x, y)(x^* - x) = \frac{1}{2} \mu^* (x^* - x)^\top \nabla_x \Theta(x^*, y) = \frac{1}{2} \mu^* \alpha,
\]

which shows the validity of (31) and (33).

Denote \( \nabla^2 \mathcal{H} := \nabla^2_{xx} \mathcal{H}(x, y) \), \( Q^* := Q(x^*, y) \) and \( K^* := K(x^*, y) \). To show validity of (32), we will show that equation (14) simplifies to \( C^*(y) = \mu^* \det(\nabla^2 \mathcal{H}) (\alpha \det(W) + \beta) \).

The factorization (24) and Taylor approximation (25) imply that (15) is equivalent to \( X(y) = \nabla^2 \mathcal{H} - \mu^* Q^* K^* (Q^*)^\top \). Furthermore, the proof of Proposition 1 shows that \( X(y) \succeq 0 \) and hence, that it is invertible. Therefore, we have

\[
\text{adj}(X(y)) = [X(y)]^{-1} \det(X(y)).
\]

Since \( \nabla^2 \mathcal{H} \succ 0 \), we can use the Woodbury identity to obtain:

\[
[X(y)]^{-1} = [\nabla^2 \mathcal{H}]^{-1} + [\nabla^2 \mathcal{H}]^{-1} \mu^* Q^* W K^* (Q^*)^\top [\nabla^2 \mathcal{H}]^{-1}
\]

Similarly, application of [29, Theorem 18.1.1] gives:

\[
\det(X(y)) = \det(\nabla^2 \mathcal{H}) \det(W)
\]

Equation (25) implies \( \nabla \mathcal{H}^*(y) = \nabla^2 \mathcal{H} \cdot (x^* - x) \), and along with the above equations, this simplifies (14) as follows:

\[
C^*(y) = \nabla \mathcal{H}^*(y) \text{adj}(X(y)) \nabla \mathcal{H}^*(y) = \det(\nabla^2 \mathcal{H}) \det(W) \cdot \left\{ (x^* - x)^\top \nabla^2 \mathcal{H} (x^* - x) + \mu^* (x^* - x)^\top Q^* W K^* (Q^*)^\top (x^* - x) \right\}
\]

\[
= \mu^* \det(\nabla^2 \mathcal{H}) (\alpha \det(W) + \beta)
\]

where the last equality follows from equation (26) and the definition of \( \alpha \) and \( \beta \).
REFERENCES


Anirudh Subramaniam is a postdoctoral researcher in the Mathematics and Computer Science Division at Argonne National Laboratory. He obtained his bachelor’s degree from the Indian Institute of Technology, Bombay and his Ph.D. from Carnegie Mellon University, both in chemical engineering. His research interests are in computational methods for nonlinear and discrete optimization under uncertainty with applications in energy, transportation and process systems.

Jacob Roth received his M.S. degree from the University of Chicago and worked at Argonne National Laboratory as a pre-doctoral appointee for two years. Currently he is a Ph.D. candidate at the University of Minnesota and is interested in computational mathematics.

Mihai Anitescu is a senior computational mathematician in the Mathematics and Computer Science Division at Argonne National Laboratory and a professor in the Department of Statistics at the University of Chicago. He obtained his engineer diploma (electrical engineering) from the Polytechnic University of Bucharest in 1992 and his Ph.D. in applied mathematical and computational sciences from the University of Iowa in 1997. He specializes in the areas of numerical optimization, computational science, numerical analysis and uncertainty quantification in which he has published more than 100 papers in scholarly journals and book chapters. He is on the editorial board of the SIAM Journal on Optimization and he is a senior editor for Optimization Methods and Software, he is a past member of the editorial boards of the Mathematical Programming A and B, SIAM Journal on Uncertainty Quantification, and SIAM Journal on Scientific Computing. He has been recognized for his work in applied mathematics by his selection as a SIAM Fellow in 2019.