

<b>Titre:</b>	Réduction de contraintes guidée par l'apprentissage pour le problème de répartition économique optimale de la production sous contraintes de sécurité
<b>Auteur:</b>	Victor Darleguy
<b>Date:</b>	2025
<b>Type:</b>	Mémoire ou thèse / Dissertation or Thesis
<b>Référence:</b>	Darleguy, V. (2025). Réduction de contraintes guidée par l'apprentissage pour le problème de répartition économique optimale de la production sous contraintes de sécurité [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie. <a href="https://publications.polymtl.ca/67830/">https://publications.polymtl.ca/67830/</a>

## Document en libre accès dans PolyPublie

Open Access document in PolyPublie

**URL de PolyPublie:** <https://publications.polymtl.ca/67830/>  
PolyPublie URL:

**Directeurs de recherche:** Antoine Lesage-Landry  
Advisors:

**Programme:** Génie énergétique  
Program:

**POLYTECHNIQUE MONTRÉAL**

affiliée à l'Université de Montréal

**Réduction de contraintes guidée par l'apprentissage pour le problème de  
répartition économique optimale de la production sous contraintes de sécurité**

**VICTOR DARLEGUY**

Département de génie électrique

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*  
Génie énergétique

Août 2025

**POLYTECHNIQUE MONTRÉAL**

affiliée à l'Université de Montréal

Ce mémoire intitulé :

**Réduction de contraintes guidée par l'apprentissage pour le problème de répartition économique optimale de la production sous contraintes de sécurité**

présenté par **Victor DARLEGUY**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

a été dûment accepté par le jury d'examen constitué de :

**Antoine LEGRAIN**, président

**Antoine LESAGE-LANDRY**, membre et directeur de recherche

**Hussein SUPRÊME**, membre externe

## DÉDICACE

*À ma mère, pour son soutien inaltérable.*

*À Fred, pour avoir toujours été à mes côtés.*

*À ma famille et à mes amis, dont la patience fut plus longue que mes silences.*

*Et à ceux qui ne sont plus là, et que je n'oublierai jamais.*

## REMERCIEMENTS

Je souhaite remercier chaleureusement mon partenaire industriel pour m'avoir offert l'opportunité de mener ce projet dans un environnement aussi stimulant qu'exigeant. J'ai eu la chance d'être intégré à une équipe aussi compétente qu'accueillante et profondément humaine.

Je tiens à adresser une mention particulière à Nadir, pour sa confiance, son soutien constant et sa bienveillance tout au long de ce travail.

Je souhaite également exprimer ma profonde gratitude à mon directeur de recherche, Antoine Lesage-Landry, pour sa disponibilité, la qualité de son accompagnement et la richesse de ses conseils tout au long de cette maîtrise.

## RÉSUMÉ

Les réseaux électriques modernes font face à une complexité croissante en raison de leur grande échelle, de l'intégration massive de ressources renouvelables intermittentes et de la nécessité d'une prise de décision en temps réel, compatible avec les mécanismes de marché. Dans ce contexte, le problème de répartition économique optimale de la production sous contraintes de sécurité (*security-constrained economic dispatch*, SCED) joue un rôle central pour garantir à la fois l'efficacité économique et la fiabilité opérationnelle. Dans sa version reposant sur les matrices des facteurs de distribution de transfert de puissance (*power transfer distribution factors*, PTDF), le SCED est largement utilisé pour sa simplicité de résolution. Toutefois, cette formulation souffre d'un nombre excessif de contraintes, notamment en cas de contingence, ce qui engendre une redondance importante, laquelle alourdit les temps de résolution et induit une représentation trop rigide des opérations réelles du réseau. Bien que la pratique industrielle repose largement sur l'usage de contraintes souples, autorisant des non-respects de limites contrôlés à un coût explicite, la majorité des travaux académiques suppose des contraintes strictes, limitant ainsi la capacité des modèles à capturer la flexibilité opérationnelle du marché.

Ce mémoire de maîtrise propose une méthodologie fondée sur l'identification dynamique d'un sous-ensemble minimal de contraintes pertinentes, adapté à chaque profil de charge, et formulée pour le cas réaliste des contraintes souples. La méthode repose sur trois composantes essentielles. D'abord, une hiérarchie imbriquée de contraintes non redondantes est construite par suppression récursive de redondances structurelles, définissant une suite croissante de régions admissibles. Ensuite, un modèle de classification supervisée de type *light gradient boosting machine* (LightGBM) est entraîné pour prédire, à partir d'un profil de charge nodale, le niveau hiérarchique minimal suffisant pour reproduire la solution exacte du SCED pénalisé. Enfin, un mécanisme de reprise adaptative garantit la correction des erreurs éventuelles de prédiction, en réinjectant progressivement les contraintes omises jusqu'à l'obtention d'une solution conforme. Cette stratégie permet de combiner l'efficacité algorithmique d'une réduction structurelle hors ligne avec la flexibilité d'un ajustement guidé par l'apprentissage lors de la phase de déploiement. Elle offre une alternative robuste et interprétable aux méthodes purement heuristiques de présélection de contraintes, en s'appuyant sur des garanties de faisabilité et une évaluation postérieure explicite de la validité des prédictions. L'ensemble de la chaîne de traitement, de la construction hiérarchique à la phase d'inférence supervisée, est conçu pour être intégré dans des contextes de résolution intensifs, typiques des applications en temps réel sur les marchés de l'électricité. Les résultats expérimentaux confirment l'efficacité

du cadre proposé. La comparaison entre la résolution du SCED complet et sa version restreinte au niveau hiérarchique exact montre qu'il est possible de réduire significativement le nombre de contraintes sans altérer la solution optimale. La hiérarchie construite s'avère ainsi structurellement pertinente. Par ailleurs, l'intégration du modèle LightGBM permet de prédire le niveau requis avec une précision de 94,3%, tout en maintenant le taux de sous-prédiction en dessous de 3,5%. Enfin, la méthode réduit en moyenne de 19,7% le temps de calcul par rapport à la formulation complète.

Ces résultats suggèrent que la méthode parvient à concilier gains de performance et garanties de robustesse, tout en maintenant un niveau de précision compatible avec les exigences opérationnelles d'environnements à contraintes de temps.

## ABSTRACT

Modern power systems face increasing complexity due to their large scale, the massive integration of intermittent renewable resources, and the need for real-time decision-making compatible with market-based mechanisms. In this context, the security-constrained economic dispatch (SCED) problem plays a pivotal role in ensuring both economic efficiency and operational reliability. In its formulation based on power transfer distribution factors (PTDF), SCED is widely adopted for its computational simplicity. However, this formulation suffers from an excessive number of constraints, especially under contingency scenarios, resulting in significant redundancy that hampers solution times and leads to an overly rigid representation of real-world operations. While industry practice often relies on soft constraints—allowing controlled violations at explicit penalty costs—most academic contributions assume hard constraints, thereby limiting the ability of such models to capture the operational flexibility inherent to electricity markets.

This master's thesis introduces a methodology grounded in the dynamic identification of a minimal subset of relevant constraints, tailored to each load profile, and formulated under the realistic assumption of soft security constraints. The proposed method comprises three core components. First, a nested hierarchy of non-redundant constraints is constructed via recursive elimination of structurally redundant elements, defining a sequence of increasingly relaxed feasible regions. Second, a supervised classification model based on light gradient boosting machine (LightGBM) is trained to predict, from a given nodal load profile, the minimal hierarchical level required to reproduce the exact solution of the penalized SCED. Third, an adaptive recovery mechanism ensures correction of potential misclassifications by progressively reinjecting omitted constraints until a compliant solution is obtained.

This strategy combines the offline algorithmic efficiency of structural constraint reduction with the online flexibility of machine-learning-guided adjustment. It provides a robust and interpretable alternative to heuristic constraint preselection methods, leveraging feasibility guarantees and posterior validation of prediction accuracy. The entire pipeline—from hierarchical construction to supervised inference—is designed to support intensive computational settings typical of real-time electricity market applications.

Experimental results confirm the effectiveness of the proposed framework. The comparison between the full SCED solution and its restricted counterpart at the exact hierarchical level demonstrates that a substantial reduction in the number of constraints can be achieved without altering the optimal solution. The constructed hierarchy thus proves to be structurally

relevant. Furthermore, the integration of the LightGBM model enables the required level to be predicted with an accuracy of 94.3%, while maintaining the underestimation rate below 3.5%. Finally, the method yields an average reduction of 19.7% in computational time compared to the complete formulation.

These findings suggest that the proposed method succeeds in reconciling computational efficiency and robustness guarantees, while maintaining a level of accuracy consistent with the stringent operational requirements of time-constrained environments.

## TABLE DES MATIÈRES

DÉDICACE . . . . .	iii
REMERCIEMENTS . . . . .	iv
RÉSUMÉ . . . . .	v
ABSTRACT . . . . .	vii
TABLE DES MATIÈRES . . . . .	ix
LISTE DES TABLEAUX . . . . .	xii
LISTE DES FIGURES . . . . .	xiii
LISTE DES SIGLES ET ABRÉVIATIONS . . . . .	xv
<b>CHAPITRE 1 INTRODUCTION . . . . .</b>	<b>1</b>
1.1 Le réseau électrique nord-américain : un marché libéralisé complexe . . . . .	1
1.1.1 Du monopole public au marché libéralisé . . . . .	2
1.1.2 Acteurs et architecture des marchés régionaux nord-américains . . . . .	3
1.1.3 Rôle des participants de marché et instruments financiers . . . . .	7
1.2 Mécanismes de marché et formation des prix : du modèle au marché . . . . .	8
1.2.1 Écoulement de puissance optimal . . . . .	10
1.2.2 Le modèle d'écoulement de puissance linéarisé . . . . .	11
1.2.3 Matrice PTDF et contingences . . . . .	13
1.2.3.1 Construction de la matrice des facteurs de distribution de transfert de puissance . . . . .	14
1.2.3.2 Modélisation des contingences et des contraintes de sécurité .	17
1.2.4 Répartition économique optimale de la production sous contraintes de sécurité . . . . .	21
1.2.5 Prix nodaux et arbitrage : dérivation des LMP . . . . .	23
1.3 Défis et opportunités actuels du SCED dans les marchés . . . . .	25
1.4 Question de recherche et objectifs du mémoire . . . . .	28
<b>CHAPITRE 2 REVUE DE LA LITTÉRATURE . . . . .</b>	<b>30</b>
2.1 Méthodes de décomposition et de calcul parallèle pour le SCED à grande échelle	30

2.1.1	Méthodes de décomposition de Benders . . . . .	31
2.1.2	Méthodes de décomposition par multiplicateurs alternés . . . . .	43
2.1.3	Autres méthodes de parallélisation . . . . .	48
2.2	Méthodes d'apprentissage supervisé pour la prédiction de contraintes actives .	49
2.2.1	Apprentissage régressif de contraintes actives . . . . .	50
2.2.1.1	Réseaux neuronaux pour la régression directe des optima . . .	51
2.2.1.2	Garanties de faisabilité et mécanismes de réparation intégrés .	52
2.2.1.3	Architectures hybrides classification-régression pour la prédic- tion structurée . . . . .	54
2.2.2	Méthodes de classification pour l'identification de contraintes actives .	55
2.2.2.1	Classification supervisée par réseaux neuronaux profonds . . .	56
2.2.2.2	Classification multiples des ensembles actifs optimaux par réseaux neuronaux . . . . .	57
2.2.2.3	Méthodes géométriques et classificateurs à séparation polyto- pique . . . . .	59
2.3	Méthodes de réduction de contraintes dans la formulation SCOPF . . . . .	61
2.3.1	Fondements de la détection de redondance . . . . .	61
2.3.2	Méthodes déterministes de réduction de contraintes . . . . .	63
2.3.3	Réduction de contraintes par apprentissage machine . . . . .	71
2.3.3.1	Réduction progressive des contraintes par classement ou score de priorité . . . . .	71
2.3.3.2	Hiérarchies implicites guidées par structure réseau ou topologie	73
2.3.3.3	Réduction paramétrée ou pilotée par représentation latente .	76
2.4	Synthèse critique et lien avec la méthode proposée . . . . .	78
<b>CHAPITRE 3 MÉTHODE DE RÉDUCTION DE CONTRAINTE GUIDÉE PAR L'APPRENTISSAGE SUPERVISÉ . . . . .</b>		81
3.1	Présentation générale de l'approche . . . . .	81
3.2	Outils de construction hiérarchique et de prédiction . . . . .	83
3.2.1	Hiérarchie de contraintes et espaces de faisabilité imbriqués . . . . .	83
3.2.2	Modèle de classification supervisée avec LightGBM . . . . .	90
3.3	Chaîne de traitement d'apprentissage supervisé . . . . .	95
3.3.1	Génération du jeu de données . . . . .	95
3.3.2	Stratégie d'exploration adaptative pour la génération de profils . . . .	97
3.3.3	Prétraitement, apprentissage et calibration des modèles . . . . .	98
3.3.4	Procédure d'inférence et de résolution . . . . .	100

3.4 Conclusion . . . . .	100
<b>CHAPITRE 4 RÉSULTATS NUMÉRIQUES . . . . .</b>	<b>102</b>
4.1 Parallélisation de la détection de contraintes redondantes . . . . .	102
4.2 Résultats de la construction hiérarchique de contraintes . . . . .	105
4.3 Résultats du modèle de prédiction de niveau hiérarchique de contrainte . . . . .	109
4.4 Procédure complète de résolution guidée par la prédiction de niveau hiérarchique	116
4.5 Conclusion des résultats . . . . .	124
<b>CHAPITRE 5 CONCLUSION . . . . .</b>	<b>125</b>
5.1 Synthèse des travaux . . . . .	125
5.2 Limitations de la solution proposée . . . . .	126
5.3 Améliorations futures . . . . .	128
<b>RÉFÉRENCES . . . . .</b>	<b>130</b>

**LISTE DES TABLEAUX**

Tableau 4.1	Comparaison entre SCED complet et formulation hiérarchiquement restreinte . . . . .	106
Tableau 4.2	Évaluation globale du modèle de prédiction du niveau hiérarchique $k^*$ .	110
Tableau 4.3	Performance du modèle par niveau hiérarchique réel $k^*$ . . . . .	110
Tableau 4.4	Bilan agrégé de la procédure prédictive à seuil $\tau = 0.5$ . . . . .	123

## LISTE DES FIGURES

Figure 1.1	Principaux opérateurs de marché régionaux indépendants (RTO/ISO) aux États-Unis. Cette carte, publiée par la <i>Federal Energy Regulatory Commission</i> (FERC), reflète le découpage fonctionnel du réseau nord-américain en zones opérées par des entités indépendantes assurant la coordination des marchés de gros d'électricité. Source : [1].	4
Figure 1.2	Protocole de coordination temps réel entre l'IESO et le NYISO. Le cycle inclut la soumission et la validation des horaires d'échange, la transmission des prévisions de charge, l'optimisation SCED, et l'ajustement des consignes à l'approche du temps réel. Source : [2].	6
Figure 1.3	Courbe de mérite – Alberta Power Exchange, 12 novembre 1999 à 13h06. Source : [3].	9
Figure 1.4	Illustration de la matrice $\Phi$ pour le réseau <code>t_case3_most.m</code> . Le noeud 3 est un générateur (G), les noeuds 1 et 2 des charges (D), et le noeud 1 est le noeud de référence (R). Les $X_{ij}$ sont les valeurs de réactances associées aux lignes. Les coefficients sont colorés selon le noeud de destination dans la matrice.	17
Figure 3.1	Projection 2D du polytope admissible $\mathcal{F}$ sur pour le réseau à 39 bus.	96
Figure 4.1	Réduction moyenne du temps de calcul par niveau hiérarchique $k^*$ .	108
Figure 4.2	Courbes de convergence par niveau avec arrêt anticipé ( <i>early stopping</i> )	112
Figure 4.3	Niveau 2 — Valeurs SHAP par variable. Longueur des barres = importance moyenne $ \phi $ ; flèche ( $\uparrow$ en vert et $\downarrow$ en rouge) = signe moyen de la contribution. Top 10 des variables classées par importance décroissante.	113
Figure 4.4	Niveau 3 — Valeurs SHAP par variable. Longueur des barres = importance moyenne $ \phi $ ; flèche ( $\uparrow$ en vert et $\downarrow$ en rouge) = signe moyen de la contribution. Top 10 des variables classées par importance décroissante.	114
Figure 4.5	Niveau 4 — Valeurs SHAP par variable. Longueur des barres = importance moyenne $ \phi $ ; flèche ( $\uparrow$ en vert et $\downarrow$ en rouge) = signe moyen de la contribution. Top 10 des variables classées par importance décroissante.	115
Figure 4.6	Taux d'exactitude de la prédiction initiale $\hat{k}^*$ en fonction du seuil de décision $\tau$ , avant activation du mécanisme de reprise.	117
Figure 4.7	Impact du seuil $\tau$ sur les erreurs de classification avant reprise.	118
Figure 4.8	Analyse des performances de prédiction par le niveau $k^*$ , en fonction du seuil de décision $\tau$ .	119

Figure 4.9	Proportion de profils nécessitant au moins une reprise adaptative, en fonction du seuil de décision $\tau$ . . . . .	121
Figure 4.10	Compromis entre l'exactitude de la prédiction initiale ( $\hat{k}^* = k^*$ ) et le gain de temps moyen obtenu par rapport à la résolution complète. . . .	122

## LISTE DES SIGLES ET ABRÉVIATIONS

ACOPF	Écoulement de puissance optimal en courant alternatif ( <i>alternating current optimal power flow</i> )
ADMM	Méthode des multiplicateurs alternés directionnels ( <i>alternating direction method of multipliers</i> )
CAISO	Opérateur de système indépendant de Californie ( <i>California independent system operator</i> )
CCGA	Algorithme de génération de colonnes et de contraintes ( <i>column-and-constraint generation algorithm</i> )
D-ED	Répartition économique dynamique ( <i>dynamic economic dispatch</i> )
DA	Marché à terme journalier ( <i>day-ahead</i> )
DCT	Coût-Charge-Topologie ( <i>demand-cost-topology</i> )
DFACTS	Systèmes flexibles distribués de transmission en courant alternatif ( <i>distributed flexible AC transmission systems</i> )
ED	Répartition économique ( <i>economic dispatch</i> )
EFB	Regroupement exclusif de caractéristiques ( <i>exclusive feature bundling</i> )
FERC	Commission fédérale de régulation de l'énergie ( <i>federal energy regulatory commission</i> )
FTR	Droits financiers de transport ( <i>financial transmission rights</i> )
GBDT	Arbres de décision à gradient renforcé ( <i>gradient boosted decision trees</i> )
GNN	Réseaux de neurones à graphes ( <i>graph neural networks</i> )
GOSS	Échantillonnage unilatéral basé sur le gradient ( <i>gradient-based one-side sampling</i> )
HILP	Fort impact, faible probabilité ( <i>high-impact low-probability</i> )
IESO	Opérateur indépendant du réseau électrique ( <i>independent electricity system operator</i> )
IMML	Lemme de modification de matrice inverse ( <i>inverse matrix modification lemma</i> )
IPM	Méthode de points intérieurs ( <i>interior point method</i> )
ISO	Opérateurs de système indépendants ( <i>independent system operators</i> )

L2O	Apprentissage pour l'optimisation ( <i>learning-to-optimize</i> )
LBLC	Base linéaire, contingence linéaire ( <i>linear base linear contingency</i> )
LDA	Analyse discriminante linéaire ( <i>linear discriminant analysis</i> )
LightGBM	Machine de gradient allégé ( <i>light gradient boosted machine</i> )
LMP	Prix marginaux localisés ( <i>locational marginal prices</i> )
LODF	Facteurs de distribution de l'effet de perte de ligne ( <i>line outage distribution factors</i> )
LP	Programme linéaire ( <i>linear program</i> )
LP-test	Test de programmation linéaire ( <i>linear programming test</i> )
LSI	Interventions chirurgicales latentes ( <i>latent surgical interventions</i> )
LT	Long terme ( <i>long-term</i> )
MISO	Opérateur de système indépendant du Centre-Ouest ( <i>Midcontinent independent system operator</i> )
NBLC	Base non linéaire, contingence linéaire ( <i>nonlinear base linear contingency</i> )
NBNC	Base non linéaire, contingence non linéaire ( <i>nonlinear base nonlinear contingency</i> )
NYISO	Opérateur de système indépendant de New York ( <i>New York independent system operator</i> )
OPF	Écoulement de puissance optimal ( <i>optimal power flow</i> )
P-UCD	Version partielle de la découverte de contraintes générales ( <i>partial UCD</i> )
PJM	Interconnexion de la Pennsylvanie et du New Jersey ( <i>Pennsylvania and New Jersey interconnection</i> )
PTDF	Facteurs de distribution de transfert de puissance ( <i>power transfer distribution factors</i> )
ROPFG	Écoulement de puissance optimal réduit avec contraintes sur les générateurs ( <i>reduced optimal power flow with generator constraints</i> )
ROPFL	Écoulement de puissance optimal réduit avec contraintes sur les lignes ( <i>reduced optimal power flow with line constraints</i> )
RT	Temps réel ( <i>real-time</i> )
RTC	Engagement en temps réel ( <i>real-time commitment</i> )

RTO	Organisations régionales de transport ( <i>regional transmission organizations</i> )
SCED	Répartition économique optimal de la production sous contrainte de sécurité ( <i>security-constrained economic dispatch</i> )
SCOPF	Écoulement de puissance optimal sous contrainte de sécurité ( <i>security-constrained optimal power flow</i> )
SCUC	Engagement unitaire sous contrainte de sécurité ( <i>security-constrained unit commitment</i> )
SF	Facteurs de décalage ( <i>shift factors</i> )
ST	Court terme ( <i>short-term</i> )
TCC	Contrats de congestion du réseau ( <i>transmission congestion contracts</i> )
TCDC	Courbes de demande pour contraintes de transport ( <i>transmission constraint demand curves</i> )
UCD	Découverte de contraintes générales ( <i>umbrella constraint discovery</i> )

## CHAPITRE 1 INTRODUCTION

La transition vers un paradigme énergétique bas carbone, combinée à la montée en puissance des ressources distribuées et à l'accroissement de l'incertitude opérationnelle, confère aux marchés de l'électricité un rôle central dans la coordination efficiente des investissements et des décisions d'exploitation. Dans cet environnement caractérisé par des interactions complexes entre agents, technologies et contraintes opérationnelles, l'analyse fine des mécanismes de marché devient une condition préalable à toute amélioration algorithmique.

Ce mémoire s'inscrit dans cette perspective en proposant une méthode adaptée aux besoins des opérateurs pour allouer efficacement les ressources de production d'électricité en respectant les contraintes physiques du réseau électrique. Ce chapitre propose une lecture structurée du fonctionnement technique des marchés nord-américains, en articulant les dimensions institutionnelles, économiques et algorithmiques, afin d'identifier les sources de complexité qui justifient l'intégration de méthodes d'optimisation avancées de résolution.

### 1.1 Le réseau électrique nord-américain : un marché libéralisé complexe

Les marchés de l'électricité en Amérique du Nord, en particulier aux États-Unis et à l'ouest du Canada, fonctionnent selon des cadres libéralisés où la production et la consommation d'électricité sont coordonnées par des mécanismes concurrentiels. Ces marchés visent à assurer une allocation efficace des ressources, une formation transparente des prix et une exploitation fiable du réseau électrique. L'architecture du marché repose sur des entités régionales indépendantes, appelées *Independent System Operators* (ISO) ou *Regional Transmission Organizations* (RTO), qui sont responsables de garantir l'équilibre offre-demande, d'organiser les enchères journalières (*day-ahead*, DA) et en temps réel (*real-time*, RT), ainsi que d'assurer la fiabilité opérationnelle du réseau.

Cette organisation décentralisée et multi-juridictionnelle s'inscrit dans un contexte technique en pleine mutation, marqué par une complexité croissante des réseaux électriques. L'essor rapide des sources de production renouvelable, notamment l'éolien et le solaire, introduit une variabilité temporelle et spatiale accrue, qui rend les écoulements de puissance plus volatils et moins prévisibles. Parallèlement, la multiplication des scénarios de défaillance à prendre en compte, comme l'exige le critère de sécurité  $N-1$ , où  $N$  réfère au nombre de lignes de transport du réseau et  $N-1$  au fait qu'il doit être en mesure d'opérer malgré l'absence d'une ligne, alour-

dit considérablement les calculs requis pour garantir la fiabilité du réseau. Cette complexité est renforcée par une asymétrie d'information structurelle entre les opérateurs de marché, tels que les fournisseurs, et les gestionnaires de réseau, les premiers ne disposant que d'une vision partielle de l'état du réseau et de ses contraintes physiques. Dans ce contexte, la qualité de la décision de production de la puissance ou d'arbitrages sur les marchés, dépend fortement de la capacité à intégrer des représentations technico-économiques précises des réseaux électriques.

Ces enjeux motivent le cadre d'analyse et de développement retenu dans ce mémoire, centré sur la formulation et la résolution efficace du problème de répartition économique optimale de la production sous contraintes de sécurité (*security-constrained economic dispatch*, SCED). Ce problème d'optimisation constitue un pilier majeur des marchés de l'électricité, et sert à déterminer la répartition économique optimale de la production (*economic dispatch*, ED) tout en respectant les limites opérationnelles du réseau, y compris en situation de contingence. Pour améliorer l'efficacité computationnelle de la résolution du SCED dans des contextes à grande échelle, ce travail propose une approche hybride, combinant des outils issus de l'optimisation convexe et des techniques d'apprentissage automatique supervisé. L'objectif est de guider intelligemment la sélection de l'ensemble des contraintes souples, afin de réduire la taille du problème à résoudre pour augmenter l'efficacité de sa résolution, sans compromettre l'exactitude des solutions obtenues.

### **1.1.1 Du monopole public au marché libéralisé**

Historiquement, le réseau électrique nord-américain était organisé selon un modèle de monopole public intégré verticalement, dans lequel une seule entité assurait conjointement la production, le transport, la distribution et la coordination du réseau électrique. Ce modèle demeure en vigueur dans certaines juridictions, notamment au Québec, où l'ensemble du réseau est exploité par la société d'état Hydro-Québec. Cette structure centralisée, longtemps justifiée par les économies d'échelle et la nature critique de l'approvisionnement de l'électricité, a progressivement cédé la place à des architectures de marché plus ouvertes. Ce mouvement de transformation institutionnelle, connu sous le nom de restructuration des marchés de l'électricité (*electricity market restructuring*), a été formalisé aux États-Unis par la publication de l'*Order No. 888* de la *Federal Energy Regulatory Commission* (FERC) en 1996. Cet ordre réglementaire imposait aux gestionnaires de réseau de transport un accès ouvert et non discriminatoire à leurs infrastructures, et appelait explicitement à la dissociation fonctionnelle des activités verticales historiques. L'objectif était de « briser le monopole sur les réseaux de transport de puissance détenus par les services publics intégrés afin de favoriser la

concurrence dans la production »<sup>1</sup> [4]. Cette réforme a aussi conduit à la création d'organismes de coordination indépendants, les ISO et RTO, responsables de l'organisation du marché et de la supervision de la fiabilité du réseau.

Cette libéralisation visait avant tout une amélioration de l'efficacité allocative par l'introduction de signaux de prix reflétant les coûts marginaux réels du réseau, tout en renforçant la transparence des mécanismes de marché. Elle a également transféré une part significative des risques économiques liés à la volatilité des prix, à la variabilité de la demande ou aux aléas techniques, vers les producteurs, fournisseurs et courtiers d'électricité. Ce nouveau cadre a favorisé l'essor de produits dérivés énergétiques comme les droits de transport financiers (*financial transmission rights*, FTR) ou les contrats futures sur l'électricité, ainsi que le développement d'outils d'analyse quantitative sophistiqués pour la modélisation des prix, la gestion de portefeuille et l'optimisation du ED [3].

### 1.1.2 Acteurs et architecture des marchés régionaux nord-américains

Les marchés de gros de l'électricité aux États-Unis reposent sur une architecture fortement décentralisée, organisée autour des ISO ou des RTO. À ce jour, neuf zones de marché structurées autour de RTO ou d'ISO couvrent plus des deux tiers de la charge électrique des États-Unis [1]. Leur principale fonction consiste à organiser les marchés de l'énergie et des services de réserve, en s'appuyant sur des procédures standardisées de collecte des offres, de coordination des écoulements de puissance, d'allocation des ressources et de règlement financier. Les RTO, à la différence des ISO, ont une reconnaissance fédérale plus explicite, une supervision interétatique du réseau de transport, et des responsabilités élargies en matière de planification du réseau.

---

1. Federal Energy Regulatory Commission, Order No. 888, April 24, 1996

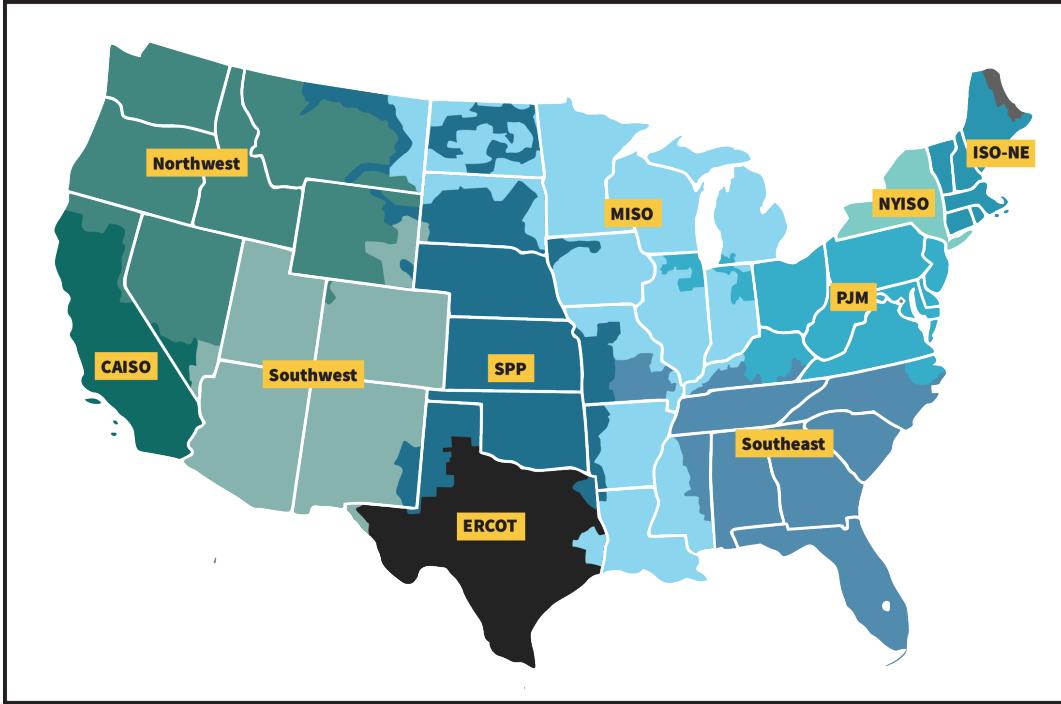


FIGURE 1.1 Principaux opérateurs de marché régional indépendants (RTO/ISO) aux États-Unis. Cette carte, publiée par la *Federal Energy Regulatory Commission* (FERC), reflète le découpage fonctionnel du réseau nord-américain en zones opérées par des entités indépendantes assurant la coordination des marchés de gros d'électricité. Source : [1].

Par exemple, la *Midcontinent Independent System Operator* (MISO) ou la *Pennsylvania and New Jersey interconnection* (PJM) sont des RTO opérant sur des territoires multi-étatiques, tandis que la *California Independent System Operator* (CAISO) ou la *New York Independent System Operator* (NYISO), bien que fonctionnellement similaires, restent classés comme ISO et sont limités à des périmètres juridictionnels plus restreints. La figure 1.1 illustre la répartition géographique de ces entités sur le territoire nord-américain. Les ISO et RTO opèrent de manière techniquement homogène : ils agrègent les offres de génération et de consommation, évaluent la sécurité du réseau en fonction de la topologie courante et des prévisions de charge, activent les réserves de manière coordonnée, et déterminent les prix marginaux locaux (*locational marginal prices*, LMP) pour chaque nœud du réseau, à partir de la résolution d'un ED. Ils assurent également le règlement des transactions et la transparence des données publiées, au service de la liquidité et de la compétitivité des marchés régionaux [5, 6].

Ce fonctionnement des marchés d'électricité en Amérique du Nord repose sur une boucle décisionnelle continue et réactive, pilotée par les ISO et RTO. Toutes les cinq minutes, les opérateurs exécutent un processus d'optimisation du SCED, qui vise à déterminer l'allocation

économiquement optimale de la génération sur le réseau en fonction de la charge, tout en satisfaisant certaines contraintes physiques (capacités de ligne, rampes de production), les contraintes économiques (liées à la demande et aux prix de l'électricité) et les critères de sécurité (tels que les marges  $N - 1$  et  $N - k$  définies plus loin). À chaque itération, ce calcul produit des consignes opérationnelles (*setpoints*) à destination des unités de production, et génère les prix d'équilibre nodaux des LMP, interprétables comme multiplicateurs de Lagrange associés à certaines contraintes du SCED [2].

Ce processus temps réel s'appuie en amont sur un marché fondamental : le marché DA (*day-ahead market*, DAM). Ce marché établit, la veille pour le lendemain, un plan d'engagement économique et opérationnel des unités de production et des échanges interzonaux, basé sur la résolution d'un problème de planification de l'engagement unitaire sous contrainte de sécurité (*security-constrained unit commitment*, SCUC). Contrairement au SCED, qui opère à court terme avec des consignes continues, le SCUC inclut des variables binaires représentant les décisions de mise en marche et arrêt, les rampes initiales, les durées minimales de fonctionnement, ou les contraintes de démarrage. Le résultat du DAM est un horaire de production optimisé, accompagné de prix horaires localisés, LMP, qui servent de référence financière pour le règlement des transactions planifiées.

Les engagements issus du DAM sont juridiquement contraignants et servent de base à la comparaison avec les réalisations temps réel dans le cadre du mécanisme de double règlement (*two-settlement system*). Dans ce contexte, les différents types d'acteurs c'est-à-dire les producteurs, consommateurs, agrégateurs, et les courtiers soumettent des offres d'énergie ou de réserve, arbitrent entre les horizons temporels (DAM et RTM), et prennent position sur les marchés de capacité ou de gestion des congestions, tels que les FTR ou les certificats de congestion de transmission (*transmission congestion contracts*, TCC)<sup>2</sup>. Le rôle différencié de chacun de ces participants est formalisé par les règles de marché définies par l'ISO ou le RTO concerné [7].

Ce cycle décisionnel, bien que commun à tous les ISO et RTO, requiert une coordination particulièrement fine lorsque plusieurs opérateurs sont impliqués dans une même interconnexion transfrontalière. C'est le cas, par exemple, entre le *New York independent system operator* (NYISO) et le *independent electricity system operator* (IESO), qui ont mis en place un protocole spécifique de coordination horaire et intra-horaire des échanges de puissance, connu sous le nom de *NY90 interchange scheduling protocol*. La figure 1.2 présente la chronologie détaillée de ce protocole, depuis la publication des engagements DA jusqu'à la confirmation

---

2. Les FTR et TCC désignent des mécanismes similaires de couverture du risque de congestion. Leur appellation diffère selon la juridiction : le terme FTR est couramment utilisé dans la plupart des marchés des États-Unis, tandis que l'expression TCC est spécifique au marché de New York (NYISO).

finale des écoulements de puissance à la frontière.

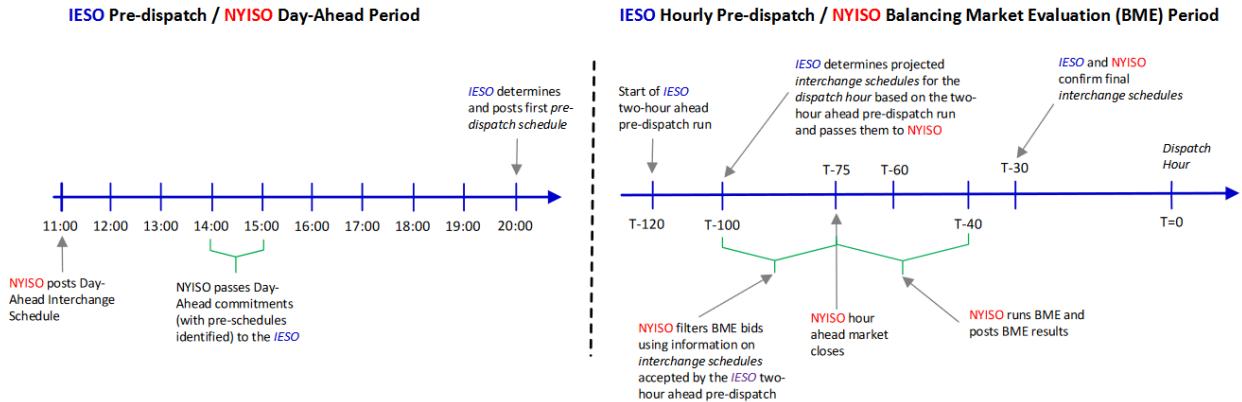


FIGURE 1.2 Protocole de coordination temps réel entre l'IESO et le NYISO. Le cycle inclut la soumission et la validation des horaires d'échange, la transmission des prévisions de charge, l'optimisation SCED, et l'ajustement des consignes à l'approche du temps réel. Source : [2].

Ce mécanisme s'articule autour d'un enchaînement rigoureux de fenêtres de planification glissantes. Par exemple, le protocole de coordination temps réel entre l'IESO et le NYISO s'articule comme suit pour un instant de temps  $T$  :

- $T - 120$  minutes : L'IESO publie les premiers résultats du calcul des **pré-répartition**<sup>3</sup> sur un horizon de 27 heures, et transmet les horaires d'échange projetés au NYISO. Les horaires projetés sont marqués du code NY90 dans les systèmes IESO, indiquant leur prise en compte dans la coordination bilatérale.
- $T - 100$  minutes à  $T - 75$  minutes : Le NYISO filtre les offres admissibles pour l'engagement unitaire sous contrainte de sécurité (*real-time commitment*, RTC) ayant une incidence sur les interconnexions IESO–NYISO. Seules les offres correspondant aux transactions acceptées par l'évaluation anticipée (à deux heures) du moteur de **pré-répartition** de l'IESO sont conservées.
- $T - 75$  minutes à  $T - 40$  minutes : Le NYISO exécute l'engagement unitaire sous contrainte de sécurité RTC, ajuste automatiquement les identifiants d'échange électroniques (*e-tags*) en fonction des résultats du RTC, puis notifie l'IESO des transactions rejetées, totalement ou partiellement. Le NYISO publie ensuite l'horaire horaire d'échange prévu.
- $T - 30$  minutes : L'IESO confirme les horaires définitifs d'échange avec le NYISO.

3. La *pré-répartition* correspond à une planification horaire anticipée effectuée à l'échelle sur un horizon glissant de plusieurs heures (typiquement 27h). Il s'agit d'une prévision optimisée des échanges interprovinciaux et de l'état du réseau, fondée sur des offres/besoins déclarés, mais non contraignante.

Si nécessaire, des réductions finales sont appliquées aux horaires d'échange, et les participants concernés reçoivent une notification automatique. Le NYISO publie alors les résultats définitifs du RTC.

- Ajustements supplémentaires entre  $T - 100$  et  $T - 75$  minutes : Afin de préserver la fiabilité du réseau, l'IESO se réserve la possibilité :
  - (a) d'inclure, par ordre de mérite économique, certaines transactions validées par l'évaluation RTC du NYISO mais non retenues dans la **pré-répartition IESO** ;
  - (b) ou d'ajuster les horaires d'échange indépendamment du protocole IESO–NYISO, si les conditions de fiabilité le justifient.
- Révisions de données d'exploitation : Toute révision ou ajout de données de programmation à l'intérieur de la fenêtre des deux heures précédant l'heure de répartition est restreinte. L'IESO peut néanmoins les autoriser en cas de nécessité pour la fiabilité interne, ou à la demande du NYISO si cela contribue à résoudre une contrainte affectant la fiabilité du réseau new-yorkais.
- Application des rampes de transaction : La mise en œuvre des transactions d'interconnexion IESO–NYISO respecte la rampe par défaut de 10 minutes définie par le NERC, centrée sur le début de l'heure de répartition.

La précision temporelle de ce protocole permet une coordination fine des décisions des deux opérateurs, tout en préservant l'intégrité du SCED dans chacun des marchés. Elle illustre également la nécessité d'une structure opérationnelle fortement synchronisée, capable de réagir en continu aux conditions du réseau, aux révisions de prévision, et aux comportements stratégiques des acteurs. Ce type de fonctionnement à haute fréquence impose des exigences considérables en matière de calculs de haute performance, de robustesse algorithmique, et de qualité des données échangées.

### 1.1.3 Rôle des participants de marché et instruments financiers

Les participants du marché constituent les agents économiques centraux du marché de l'électricité. Ils interviennent à la fois sur les segments physiques et financiers, en soumettant des offres d'énergie ou de réserve dans le DAM et le RTM, et en prenant des positions spéculatives ou de couverture à l'aide de produits dérivés. Cette catégorie regroupe des producteurs, des fournisseurs, des agrégateurs, mais aussi des entités purement financières. Leur activité soutient la liquidité des marchés et contribue à l'efficacité des signaux de prix, tout en introduisant des dynamiques stratégiques complexes.

Les instruments utilisés par ces acteurs se répartissent en plusieurs classes. Les offres dites *physiques* impliquent un engagement ferme de livraison ou de consommation d'électricité

à un point spécifique du réseau. Par contraste, le courtage virtuel (*virtual bidding*) permet de prendre des positions purement financières dans le DAM, sans livraison physique, afin d'arbitrer les écarts attendus avec le RTM. Ce mécanisme contribue à la convergence des prix entre les deux marchés, tout en exposant les acteurs à un risque de règlement différentiel [8].

D'autres instruments comme les FTR permettent aux participants de se prémunir contre les coûts de congestion en s'adossant à la différence de prix entre deux nœuds dans le DAM. Les FTR génèrent un écoulement de puissance de revenus proportionnel à l'intensité de la congestion entre un point d'injection et un point de soutirage. Enfin, les contrats d'écart (*basis swaps*) sont des instruments dérivés permettant aux participants de se couvrir contre les écarts de prix entre deux localisations, par exemple, un point d'injection et un hub régional, ou entre deux produits temporellement liés. Ces contrats sont conçus pour gérer le risque de base, c'est-à-dire l'incertitude liée à la fluctuation de ces écarts dans le temps. Ces outils, bien que définis contractuellement, s'inscrivent dans une structure de marché fortement contrainte par la physique du réseau électrique, renforçant le lien entre modélisation économique et résolution de problèmes d'optimisation sous contraintes [8].

## 1.2 Mécanismes de marché et formation des prix : du modèle au marché

Les marchés de gros de l'électricité en Amérique du Nord reposent sur une architecture fortement structurée, dans laquelle les ISO et RTO déterminent l'allocation des ressources via la résolution successive de problèmes d'optimisation. Ces problèmes traduisent la recherche d'un équilibre économico-technique, dans lequel la production est planifiée de manière à satisfaire la demande tout en respectant les contraintes physiques du réseau. Le cœur de cette approche repose sur la modélisation du SCED à travers l'écoulement de puissance optimal (*optimal power flow, OPF*), qui vise à minimiser le coût global de production du réseau sous contraintes d'équilibre nodal, de capacité des lignes et de limites opérationnelles des générateurs [9].

Dans un cadre théorique idéal, les interactions offre-demande sur un marché libéralisé peuvent être représentées graphiquement par l'intersection d'une courbe d'offre agrégée (ou *merit order*) croissante et d'une courbe de demande décroissante. Le point d'intersection détermine un prix d'équilibre unique, appelé *system marginal price*. Ce mécanisme constitue la base du fonctionnement des marchés de l'électricité de type *uniform-price*, où un prix unique est déterminé par l'intersection entre la courbe d'offre agrégée (*merit order*) et la demande du réseau, comme illustré à la figure 1.3, issue du marché de l'électricité en Alberta [3]. Nous observons notamment la forte inélasticité de la demande, représentée par une courbe presque verticale.

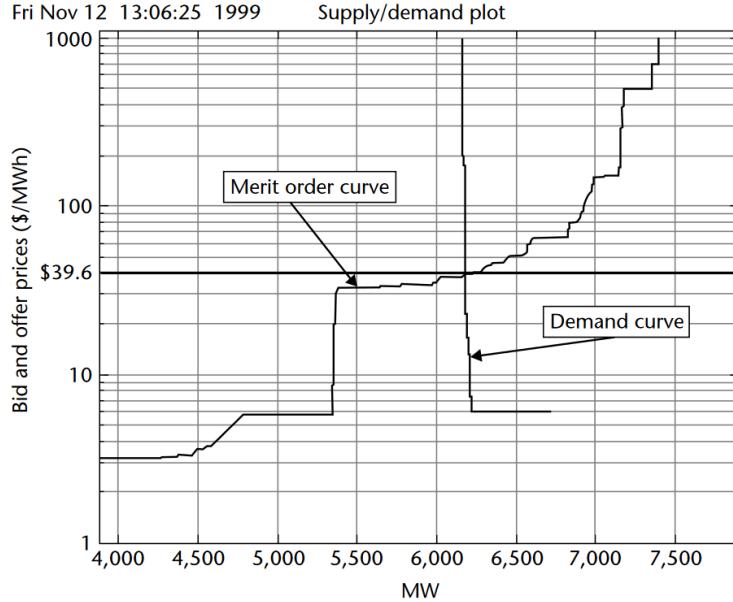


FIGURE 1.3 Courbe de mérite – Alberta Power Exchange, 12 novembre 1999 à 13h06.  
Source : [3].

Cependant, cette représentation agrégée omet deux éléments cruciaux : (i) la demande est généralement inélastique à court terme, ce qui aplatis la courbe de demande et (ii) le réseau est contraint physiquement, de sorte que tous les producteurs ne peuvent injecter librement, même si leur coût est compétitif. Ces limitations rendent insuffisants les modèles globaux de marché et justifient le recours à des mécanismes dits de *clearing nodal*.

Dans ce cadre, les décisions de production sont déterminées en résolvant un SCED, formulation simplifiée et convexe de l'OPF, dans laquelle les relations physiques non linéaires sont linéarisées. Le problème ainsi formulé produit, pour chaque instant, un plan de répartition économique optimal de la production  $p_i^*$  pour chaque unité, ainsi qu'un ensemble de LMP associés à chaque nœud du réseau. Ces LMP reflètent le coût associé à l'augmentation d'une unité de puissance à un point donné du réseau [9,10]. Le LMP au noeud  $n$  peut être décomposé en trois composantes fondamentales qui sont le coût marginal de production, le coût marginal de la congestion et le coût marginal des pertes ohmiques dans le réseau électrique. Cette décomposition constitue le fondement du mécanisme de prix d'équilibre (*clearing*) des marchés noraux et du calcul des paiements aux producteurs dans les marchés nord-américains. Nous détaillerons la dérivation de LMP et leur interprétation dans une prochaine partie.

### 1.2.1 Écoulement de puissance optimal

L'OPF consiste à déterminer la manière optimale d'exploiter un réseau électrique donné. Le problème complet de l'OPF, i.e., dans sa forme exacte, est dénoté comme écoulement de puissance optimal en courant alternatif (*alternating current optimal power flow*, ACOPF) et prend la forme d'un problème quadratique non convexe, car il modélise fidèlement les lois physiques de l'écoulement de puissance, et ce, à l'aide de variables complexes.

Considérons un réseau électrique modélisé par un graphe orienté  $(\mathcal{N}, \mathcal{L})$ , où :

- $\mathcal{N} \subset \mathbb{N}$  désigne l'ensemble des noeuds du réseau et chaque noeud  $i \in \mathcal{N}$  représente un point de connexion physique tel qu'un site de production, de consommation, ou une sous-station.
- $\mathcal{L} \subseteq \mathcal{N} \times \mathcal{N}$  désigne l'ensemble des arêtes orientées (ou lignes de transport) et chaque arête  $(i, j) \in \mathcal{L}$  représente l'écoulement de puissance du noeud  $i$  vers le noeud  $j$ .

Pour chaque noeud  $i \in \mathcal{N}$ , et chaque ligne  $(i, j) \in \mathcal{L}$ , introduisons les variables suivantes :

- $v_i \in \mathbb{C}$  : tension complexe au noeud  $i \in \mathcal{N}$  ;
- $p_i \in \mathbb{R}$  : puissance active nette injectée au noeud  $i \in \mathcal{N}$  ;
- $q_i \in \mathbb{R}$  : puissance réactive nette injectée au noeud  $i \in \mathcal{N}$  ;
- $p_{ij} \in \mathbb{R}$  : composante active de l'écoulement de puissance sur la ligne  $(i, j) \in \mathcal{L}$  ; un écoulement de puissance positif  $p_{ij} > 0$  indique un transfert de  $i$  vers  $j$  ;
- $q_{ij} \in \mathbb{R}$  : composante réactive de l'écoulement de puissance sur la ligne  $(i, j) \in \mathcal{L}$ .

Les paramètres du modèle sont :

- $y_{ij} = g_{ij} + j b_{ij}$  : admittance complexe de la ligne  $(i, j) \in \mathcal{L}$ , où  $g_{ij} \in \mathbb{R}$  et  $b_{ij} \in \mathbb{R}$  sont respectivement la conductance et la susceptance de la ligne, et  $j$  désigne l'unité imaginaire ;
- $|v_i| \in [v_i, \bar{v}_i]$  : borne admissible sur la magnitude de la tension au noeud  $i \in \mathcal{N}$  ;
- $p_i \in [p_i, \bar{p}_i]$ ,  $q_i \in [q_i, \bar{q}_i]$  : bornes admissibles sur les puissances active et réactive injectées au noeud  $i \in \mathcal{N}$  ;
- $p_{ij}^2 + q_{ij}^2 \leq \bar{s}_{ij}^2$  : limite thermique sur la puissance apparente transitant sur la ligne  $(i, j) \in \mathcal{L}$ .

Le problème ACOPF s'écrit alors :

$$\min_{(\mathbf{p}, \mathbf{q}, \mathbf{v})} f(\mathbf{p}) \quad (1.1a)$$

$$\text{s.c. } p_{ij} + jq_{ij} = v_i(v_i^* - v_j^*)y_{ij}^*, \quad \forall (i, j) \in \mathcal{L} \quad (1.1b)$$

$$p_i = \sum_{j:(i,j) \in \mathcal{L}} p_{ij}, \quad \forall i \in \mathcal{N} \quad (1.1c)$$

$$q_i = \sum_{j:(i,j) \in \mathcal{L}} q_{ij}, \quad \forall i \in \mathcal{N} \quad (1.1d)$$

$$p_{ij}^2 + q_{ij}^2 \leq \bar{s}_{ij}^2, \quad \forall (i,j) \in \mathcal{L} \quad (1.1e)$$

$$\underline{v}_i \leq |v_i| \leq \bar{v}_i, \quad \forall i \in \mathcal{N} \quad (1.1f)$$

$$\underline{p}_i \leq p_i \leq \bar{p}_i, \quad \forall i \in \mathcal{N} \quad (1.1g)$$

$$\underline{q}_i \leq q_i \leq \bar{q}_i, \quad \forall i \in \mathcal{N} \quad (1.1h)$$

La fonction objectif (1.1a) minimise le coût total de production lié à la puissance active. La contrainte (1.1b) décrit les écoulements de puissance complexes sur chaque ligne de transport. Les bilans de puissance active et réactive sont respectivement exprimés dans (1.1c) et (1.1d). La contrainte (1.1e) constraint la puissance apparente circulant sur chaque ligne. Les contraintes de tension nodale sont imposées par (1.1f), via une borne inférieure et supérieure sur la magnitude de la tension. Les limites admissibles sur les puissances injectées sont encadrées dans (1.1g) pour la puissance active, et (1.1h) pour la puissance réactive.

Même en supposant une fonction objectif  $f$  convexe (souvent linéaire en pratique), la non-convexité des contraintes (1.1b) et (1.1f) rendent le problème non convexe et difficile à résoudre globalement. Cette complexité structurelle justifie le recours à des approximations, des relaxations ou des formulations alternatives dans les marchés de l'électricité [9, 10].

### 1.2.2 Le modèle d'écoulement de puissance linéarisé

Afin de contourner la non-convexité du modèle ACOPF, nous introduisons une approximation convexe, connue sous le nom d'écoulement de puissance linéarisé ou DCOPF. Ce modèle repose sur plusieurs hypothèses simplificatrices :

- La puissance réactive est négligée : nous omettons les variables  $q_i$  et  $q_{ij}$  ;
- Le réseau opère autour de sa tension nominale et nous fixons les modules  $|v_i| = 1$  de la tension ;
- La conductance  $g_{ij}$  des lignes est négligeable devant leur susceptance  $b_{ij}$ , i.e.  $g_{ij} \approx 0$ .

Dans ce cadre, les écoulements de puissance active sur chaque ligne  $(i,j) \in \mathcal{L}$  deviennent une fonction linéaire de l'écart de phase entre les tensions natales :

$$p_{ij} = b_{ij}(\theta_i - \theta_j),$$

où  $\theta_i \in [0, 2\pi]$  désigne l'angle de phase de la tension au noeud  $i \in \mathcal{N}$ , exprimé en radians.

Dans ce modèle, la limite thermique  $\bar{s}_{ij}$  désigne la capacité apparente maximale de la ligne  $(i, j)$ , exprimée en MVA. Il s'agit de la norme du vecteur de puissance apparente, i.e.,

$$\bar{s}_{ij} = (\bar{p}_{ij}^2 + \bar{q}_{ij}^2)^{1/2},$$

où  $\bar{p}_{ij}$  et  $\bar{q}_{ij}$  sont les bornes maximales admissibles sur les puissances active et réactive transitant sur la ligne.

La formulation du problème DCOPF (1.2) est donnée par :

$$\min_{(\theta, p)} f(p) \quad (1.2a)$$

$$\text{s.c. } p_{ij} = b_{ij}(\theta_i - \theta_j), \quad \forall (i, j) \in \mathcal{L} \quad (1.2b)$$

$$\sum_{j:(i,j) \in \mathcal{L}} p_{ij} = p_i, \quad \forall i \in \mathcal{N} \quad (1.2c)$$

$$\underline{p}_i \leq p_i \leq \bar{p}_i, \quad \forall i \in \mathcal{N} \quad (1.2d)$$

$$|p_{ij}| \leq \bar{s}_{ij}, \quad \forall (i, j) \in \mathcal{L}. \quad (1.2e)$$

La fonction objectif (1.2a) représente le coût total de production de puissance active à minimiser. La contrainte (1.2b) modélise les écoulements de puissance active sur chaque ligne de transport, comme fonction linéaire de l'écart de phase. La contrainte (1.2c) impose l'équilibre nodal des puissances injectées. Les bornes admissibles sur les injections sont spécifiées dans (1.2d), tandis que (1.2e) contraint la puissance maximale pouvant transiter sur chaque ligne.

Chaque noeud  $i \in \mathcal{N}$  est associé à une injection ou une absorption nette  $p_i$ , positive pour un générateur, négative pour une charge. Si la fonction objectif  $f$  est linéaire, cette formulation constitue un programme linéaire (*linear program*, LP).

Le DCOPF constitue un modèle convexe canonique pour la répartition économique optimale de la production dans les réseaux électriques. Il impose strictement les contraintes d'équilibre nodal ainsi que les limites d'écoulement de puissance sur un réseau supposé intègre, c'est-à-dire exempt de toute défaillance. Cette modélisation repose sur l'hypothèse d'une topologie parfaitement disponible, ne tenant pas compte des aléas structurels, notamment les défaillances de lignes ou de générateurs, appelée contingences, qui sont pourtant au cœur des exigences de sécurité définies par les opérateurs de transport, telles que les critères  $N - 1$  ou  $N - k$  présentés ultérieurement.

### 1.2.3 Matrice PTDF et contingences

Dans un contexte opérationnel, l'hypothèse d'intégrité absolue du DCOPF rend le modèle excessivement rigide. Elle justifie l'adoption d'une extension plus réaliste, dans laquelle certaines contraintes techniques peuvent être relâchées à coût pénalisé. Cette relaxation permet de préserver l'interprétabilité économique du modèle, en cohérence avec les logiques de marché, tout en introduisant une flexibilité indispensable à la gestion des incertitudes.

Dans cette optique, les marchés nord-américains ont adopté une formulation du problème de répartition économique reposant sur la matrice des facteurs de distribution de transfert de puissance (*power transfer distribution factors*, PTDF) aussi appelée facteurs de distribution de transfert de puissance (*shift factors*, SF) dans certains documents d'exploitation des ISO [6, 7]. L'utilisation de la formulation PTDF présente plusieurs avantages conceptuels et numériques [11] :

- **Interprétabilité physique** : chaque coefficient de la matrice PTDF exprime la sensibilité de l'écoulement de puissance active sur une ligne de transport à une injection marginale de puissance active en un nœud donné (dans le cadre d'un DCOPF, la puissance réactive étant omise, cette interprétation repose exclusivement sur les équilibres actifs) ;
- **Réduction dimensionnelle** : les variables d'angle de phase  $\theta_i$  sont éliminées, ce qui diminue le nombre de variables et améliore la mise à l'échelle ;
- **Contraintes explicites** : les limites thermiques sont imposées directement sur les écoulements de puissance déduits des injections nodales ;
- **Adéquation avec les processus de marché** : la structure PTDF permet un traitement rapide et modulaire des contraintes actives, notamment pour les mécanismes de redéploiement économique.

Dans les pratiques de marché, notamment chez MISO [6, 7], les PTDF sont mobilisés pour caractériser l'impact relatif de chaque ressource sur une contrainte active du réseau électrique. Lorsqu'une ligne de transport devient congestionnée, les opérateurs s'appuient sur les coefficients de la matrice PTDF pour identifier les générateurs ayant le plus fort effet de décharge sur cette contrainte, c'est-à-dire ceux dont l'activation marginale permet de réduire le dépassement de manière efficace.

Cette analyse est intégrée au processus de redéfinition des courbes de demande de contrainte de transport (*transmission constraint demand curves*, TCDC), qui traduisent la pénalité économique associée au non-respect d'une limite d'écoulement de puissance du réseau. Si la TCDC initialement définie ne permet pas de soulager la contrainte de façon satisfaisante, les opérateurs peuvent la modifier temporairement (procédure dite d'*override*), en s'appuyant

explicitement sur les coefficients de la matrice PTDF pour orienter la re-répartition économique de manière ciblée, tout en préservant la sécurité du réseau [7].

### 1.2.3.1 Construction de la matrice des facteurs de distribution de transfert de puissance

Soit  $\mathbf{B}_{\text{bus}} \in \mathbb{R}^{|\mathcal{N}|^2}$  la matrice de susceptance nodale (aussi appelée matrice de Kirchhoff) associée au réseau électrique, définie pour tout couple  $(i, j) \in \mathcal{N} \times \mathcal{N}$  par :

$$[\mathbf{B}_{\text{bus}}]_{ij} = \begin{cases} \sum_{k:(i,k) \in \mathcal{L}} b_{ik}, & \text{si } i = j \\ -b_{ij}, & \text{si } (i, j) \in \mathcal{L} \text{ et } i \neq j \\ 0, & \text{sinon.} \end{cases}$$

La matrice  $\mathbf{B}_{\text{bus}}$  est symétrique, à diagonale strictement dominante, mais singulière. Cette singularité reflète l'indétermination absolue des angles de phase en l'absence de référence : son noyau est de dimension un, engendré par le vecteur constant  $\mathbf{1}_n$  de dimension  $n$ . Par conséquent, l'inversion directe de  $\mathbf{B}_{\text{bus}}$  n'est pas possible. Pour construire une matrice inversible, nous définissons la matrice de réactance nodale  $\hat{\mathbf{B}}_{\text{bus}}$  selon la procédure suivante.

Soit  $i_0 \in \mathcal{N}$  un nœud désigné comme nœud de référence (*slack bus*), choisi en accord avec les conventions du marché (typiquement le nœud auquel est attaché le prix de référence pour les LMP). La matrice  $\hat{\mathbf{B}}_{\text{bus}} \in \mathbb{R}^{|\mathcal{N}|^2}$  est construite comme suit :

1. Nous formons la matrice réduite  $\tilde{\mathbf{B}}_{\text{bus}} \in \mathbb{R}^{(|\mathcal{N}|-1)^2}$  en supprimant de  $\mathbf{B}_{\text{bus}}$  la ligne et la colonne correspondant au nœud  $i_0$  :

$$\tilde{\mathbf{B}}_{\text{bus}} = \mathbf{B}_{\text{bus}}[\mathcal{N} \setminus \{i_0\}, \mathcal{N} \setminus \{i_0\}].$$

2. Nous inversons cette matrice symétrique définie positive pour obtenir la pseudo inverse de la matrice de susceptance nodale utilisé pour la formation de la matrice PTDF, que nous notons  $\Phi$  :

$$\tilde{\mathbf{X}}_{\text{bus}} = \tilde{\mathbf{B}}_{\text{bus}}^\dagger. \quad (1.3)$$

3. Pour reconstruire le vecteur complet des angles de phase  $\boldsymbol{\theta} \in \mathbb{R}^{|\mathcal{N}|}$ , nous définissons une matrice étendue  $\hat{\mathbf{B}}_{\text{bus}} \in \mathbb{R}^{|\mathcal{N}|^2}$  en insérant une ligne et une colonne de zéros à la

position  $i_0$ , ce qui revient à imposer  $\theta_{i_0} = 0$  :

$$[\hat{\mathbf{B}}_{\text{bus}}]_{ij} = \begin{cases} [\tilde{\mathbf{B}}_{\text{bus}}^\dagger]_{ij}, & \text{si } i, j \in \mathcal{N} \setminus \{i_0\} \\ 0, & \text{si } i = i_0 \text{ ou } j = i_0. \end{cases} \quad (1.4)$$

La matrice de susceptance des lignes, notée  $\mathbf{B}_{\text{line}} \in \mathbb{R}^{|\mathcal{L}| \times |\mathcal{N}|}$ , permet de représenter les écoulements de puissance de (1.2b) sous forme vectorielle en fonction des angles de phase noraux. Elle est définie comme suit.

Chaque ligne  $(i, j) \in \mathcal{L}$  est orientée arbitrairement de son extrémité  $i \in \mathcal{N}$  vers  $j \in \mathcal{N}$ , et est associée à une susceptance positive  $b_{ij} > 0$ . La matrice  $\mathbf{B}_{\text{line}}$  est construite ligne par ligne, selon la correspondance ligne-nœud suivante :

$$[\mathbf{B}_{\text{line}}]_{(i,j),k} = \begin{cases} b_{ij}, & \text{si } k = i \text{ (nœud d'origine de la ligne } (i, j)) \\ -b_{ij}, & \text{si } k = j \text{ (nœud d'arrivée de la ligne } (i, j)) \\ 0, & \text{sinon.} \end{cases} \quad \forall (i, j) \in \mathcal{L}, k \in \mathcal{N} \quad (1.5)$$

Ainsi, chaque ligne  $(i, j) \in \mathcal{L}$  est représentée par une ligne creuse de  $\mathbf{B}_{\text{line}}$ , contenant exactement deux coefficients non nuls :  $+b_{ij}$  au nœud d'origine  $i$ , et  $-b_{ij}$  au nœud d'arrivée  $j$ . Cette structure permet d'exprimer de manière compacte les écoulements de puissance active sur l'ensemble des lignes à partir du vecteur des angles de phase noraux  $\boldsymbol{\theta} \in \mathbb{R}^{|\mathcal{N}|}$  comme suit :

$$p_{ij} = \sum_{k \in \mathcal{N}} [\mathbf{B}_{\text{line}}]_{(i,j),k} \theta_k \quad \forall (i, j) \in \mathcal{L} \quad (1.6)$$

où  $p_{ij} \in \mathbb{R}$  est la puissance active qui s'écoule sur la ligne orienté de  $i$  vers  $j$ .

Il convient de noter que, contrairement à la matrice de réactance nodale  $\hat{\mathbf{B}}_{\text{bus}}$ , la matrice  $\mathbf{B}_{\text{line}}$  est généralement rectangulaire, et plus précisément de type mince (plus de lignes que de nœuds) dans les réseaux maillés typiques.

À partir de (1.2c) et (1.4), le vecteur des angles de phase noraux peut s'exprimer comme une combinaison linéaire des puissances actives injectées :

$$\boldsymbol{\theta} = \hat{\mathbf{B}}_{\text{bus}} \mathbf{p}, \quad (1.7)$$

où  $\boldsymbol{\theta} \in \mathbb{R}^{|\mathcal{N}|}$  désigne le vecteur des angles de phase (en radians),  $\mathbf{p} \in \mathbb{R}^{|\mathcal{N}|}$  le vecteur des puissances actives nettes injectées aux nœuds, et  $\hat{\mathbf{B}}_{\text{bus}}$  la matrice de réactance nodale réduite

(pseudo-inverse de la matrice de susceptance nodale), introduite en (1.4).

La relation (1.7) s'écrit, composante par composante, sous la forme indexée suivante :

$$\theta_i = \sum_{j \in \mathcal{N}} [\hat{\mathbf{B}}_{\text{bus}}]_{ij} p_j, \quad \forall i \in \mathcal{N}, \quad (1.8)$$

ce qui explicite le couplage entre les angles de phase et les injections actives sur l'ensemble du réseau, via les coefficients de  $\hat{\mathbf{B}}_{\text{bus}}$ . En réinjectant (1.8) dans (1.6), nous obtenons une expression directe de la puissance active  $p_{ij}$  s'écoulant sur la ligne orientée  $(i, j) \in \mathcal{L}$ , en fonction des puissances injectées aux nœuds  $k \in \mathcal{N}$  :

$$p_{ij} = \sum_{k \in \mathcal{N}} [\mathbf{B}_{\text{line}}]_{(i,j),k} \theta_k = \sum_{k \in \mathcal{N}} [\mathbf{B}_{\text{line}}]_{(i,j),k} \sum_{r \in \mathcal{N}} [\hat{\mathbf{B}}_{\text{bus}}]_{k\ell} p_r. \quad (1.9)$$

Cette formulation met en évidence la linéarité des écoulements de puissance en fonction des injections, et prépare la définition matricielle de la matrice de sensibilité  $\Phi$  :

$$\Phi = \mathbf{B}_{\text{line}} \hat{\mathbf{B}}_{\text{bus}}.$$

Dès lors, les contraintes d'écoulement (1.2e) peuvent être exprimées sous forme matricielle :

$$-\bar{\mathbf{s}} \leq \Phi \mathbf{p} \leq \bar{\mathbf{s}}, \quad (1.10)$$

où  $\bar{\mathbf{s}} \in \mathbb{R}^{|\mathcal{L}|}$  est le vecteur des limites thermiques admissibles pour chaque ligne. La version indicielle compacte du DCOPF sous forme PTDF, dans laquelle les angles de phase sont éliminés, s'écrit alors :

$$\min_{\mathbf{p}} f(\mathbf{p}) \quad (1.11a)$$

$$\text{s.c. } \sum_{i \in \mathcal{N}} p_i = 0 \quad (1.11b)$$

$$\underline{p}_i \leq p_i \leq \bar{p}_i, \quad \forall i \in \mathcal{N} \quad (1.11c)$$

$$-\bar{s}_{ij} \leq \sum_{k \in \mathcal{N}} \Phi_{ij,k} p_k \leq \bar{s}_{ij}, \quad \forall (i, j) \in \mathcal{L}, \quad (1.11d)$$

où la fonction objective (1.11a) minimise le coût total de production active, (1.11b) impose le bilan global des injections dans le réseau, en assurant que la somme des puissances nettes injectées est nulle, (1.11c) encadrent les puissances injectées à chaque nœud entre leurs bornes inférieure et supérieure et enfin, les contraintes d'écoulement (1.11d) expriment la limite thermique admissible sur chaque ligne de transport, en fonction des injections natales via la

matrice  $\Phi$ .

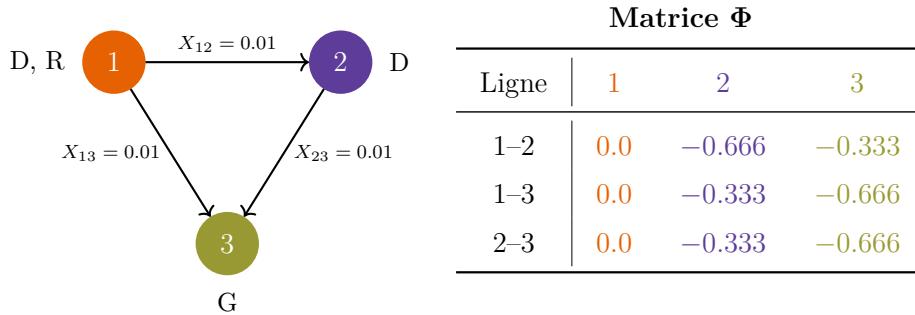


FIGURE 1.4 Illustration de la matrice  $\Phi$  pour le réseau `t_case3_most.m`. Le nœud 3 est un générateur (G), les nœuds 1 et 2 des charges (D), et le nœud 1 est le nœud de référence (R). Les  $X_{ij}$  sont les valeurs de réactances associées aux lignes. Les coefficients sont colorés selon le nœud de destination dans la matrice.

Utilisons la figure 1.4 pour illustrer et interpréter la matrice les coefficients de la matrice  $\Phi$ . Tout d'abord, la colonne associée au nœud de référence (ici, le nœud 1) est nulle, ce qui reflète le fait qu'une variation d'injection nette à ce nœud n'a aucun effet relatif sur les écoulements de puissance, l'angle de phase  $\theta_1$  étant fixé à zéro. En revanche, une injection nette de puissance au nœud 2 se traduit par un écoulement de puissance négatif sur les lignes 1–2, 1–3, et 2–3, indiquant un transfert de puissance partiel vers le nœud de référence et vers les autres charges. De même, une augmentation de la génération au nœud 3 induit un écoulement de puissance dirigé vers les charges, principalement via la ligne 1–3.

Les coefficients négatifs indiquent un écoulement orienté dans le sens opposé à l'orientation définie pour la ligne. La répartition des coefficients  $-0,666$  et  $-0,333$  traduit une distribution topologiquement cohérente des écoulements dans un réseau maillé : une même injection se répartit entre plusieurs chemins de manière proportionnelle aux réactances relatives. Enfin, la structure de cette matrice souligne que le choix du nœud de référence influe sur les valeurs absolues des coefficients, sans toutefois modifier les écoulements physiques réels, lesquels dépendent uniquement des injections nettes et des caractéristiques électriques du réseau.

### 1.2.3.2 Modélisation des contingences et des contraintes de sécurité

Dans les marchés d'électricité libéralisés, la planification de la production ne peut se limiter à satisfaire les injections nettes attendues sous conditions normales. Le respect de critères de sécurité opérationnelle impose d'anticiper des événements indésirables, appelés contingences, susceptibles d'affecter l'état du réseau électrique. Une contingence désigne typiquement la défaillance soudaine d'un élément critique du réseau, tel qu'une ligne de transport, un

transformateur ou un groupe de production [2].

Afin de garantir la continuité de service et l'adéquation de puissance du réseau, les ISO et RTO appliquent le critère de sécurité  $N-k$ , selon lequel le réseau doit rester opérationnel après la perte simultanée de  $k$  composants indépendants parmi les  $N$  lignes actives. La version la plus couramment utilisée est le critère  $N-1$ , qui impose la résilience du système face à toute perte unique crédible d'un composant. Cette exigence est formalisée dans les procédures d'exploitation en temps réel de marchés tels que NYISO, IESO ou MISO [2, 7].

Dans le cadre du SCED, ces scénarios de contingence sont modélisés explicitement par un ensemble de contraintes thermiques, valides dans chaque état post-contingence. Dans le contexte des marchés électriques, un état pré-contingence désigne la configuration opérationnelle du réseau avant la survenue d'une défaillance. Les contraintes de sécurité y sont évaluées sous l'hypothèse de fonctionnement nominal, et les répartitions économiques sont calculées pour rester robustes à toute perte admissible. Un état post-contingence, en revanche, correspond à la topologie du réseau après la perte d'une ligne spécifiée dans un scénario de contingence. Le SCED impose que les écoulements de puissance dans ces états modifiés restent compatibles avec les limites thermiques des équipements restants. L'introduction de ces contraintes supplémentaires augmente drastiquement la dimension du problème d'optimisation, justifiant le recours à des outils analytiques spécifiques, notamment la construction de matrices de sensibilité aux contingences, qui seront présentées ci-après.

La modélisation explicite des états post-contingence repose sur l'usage de la matrice des facteurs de distribution de l'écoulement en cas de défaillance (*line outage distribution factors*, LODF). Cette matrice permet d'évaluer analytiquement la variation de l'écoulement de puissance sur une ligne  $(i, j) \in \mathcal{L}$ , à la suite de la perte soudaine d'une autre ligne  $(k, l) \in \mathcal{L}$ . Les LODF permettent d'exprimer l'impact relatif d'une défaillance sur la redistribution des écoulements physiques, sans le recalcule complet des facteurs la matrice PTDF. Cette propriété est particulièrement utile dans un contexte opérationnel ou de simulation, car reconstruire la matrice des facteurs de transfert  $\Phi$  nécessite l'inversion partielle ou totale de la matrice de susceptance nodale  $\mathbf{B}_{\text{bus}}$ , une opération dont le coût croît significativement avec la taille du réseau [12] (cette problématique sera approfondie plus tard). Notons  $\Lambda$  la matrice LODF. Concrètement, si  $p_{ij}^{(0)}$  désigne l'écoulement sur la ligne  $(i, j)$  en régime pré-contingence, alors son écoulement post-contingence, noté  $p_{ij}^{(c)}$ , après l'indisponibilité de la ligne  $(k, l)$ , est donné par [13] :

$$p_{ij}^{(c)} = p_{ij}^{(0)} + \Lambda_{ij,kl} p_{kl}^{(0)}, \quad (1.12)$$

Ce facteur traduit donc une sensibilité topologique et physique de l'écoulement sur  $(i, j)$  vis-à-vis de la perte de  $(k, l)$ . L'ensemble des coefficients  $\Lambda_{ij,kl}$  peut être organisé sous forme

matricielle, et combiné avec la matrice  $\Phi$  pour former une version ajustée de cette dernière, propre à chaque scénario de contingence. Cette matrice modifiée, notée  $\Phi^{(c)}$ , permet de représenter l'état post-contingence  $c \in \mathcal{C}$  comme une simple application linéaire des injections nodales [13], où  $\mathcal{C} \subset \mathbb{N}$  est l'ensemble des scénarios de contingence.

Soit  $\mathcal{L}_c \subset \mathcal{L}$  l'ensemble des lignes en panne dans le scénario de contingence  $c \in \mathcal{C}$  et  $\mathbf{A} \in \{-1, 0, 1\}^{|\mathcal{L}| \times |\mathcal{N}|}$  la matrice d'incidence orientée du réseau où chaque ligne  $\mathbf{A}_{(i,j)}$  correspond à une ligne de transport  $(i, j) \in \mathcal{L}$  et vérifie :

$$[\mathbf{A}_{(i,j)}]_n = \begin{cases} +1 & \text{si } n = i \text{ (nœud d'envoi)} \\ -1 & \text{si } n = j \text{ (nœud de réception)} \\ 0 & \text{sinon} \end{cases} \quad (1.13)$$

Pour chaque ligne  $(i, j) \notin \mathcal{L}_c$ , la  $\Lambda$  associée est donnée par (1.14) [13] :

$$\Lambda_{(i,j),\mathcal{L}_c} = \left( \mathbf{X}_{(i,j)}^{-1} \mathbf{A}_{(i,j)} \hat{\mathbf{B}}_{\text{bus}} \mathbf{A}_{\mathcal{L}_c}^\top \right) \left( \mathbf{I} - \mathbf{X}_{\mathcal{L}_c}^{-1} \mathbf{A}_{\mathcal{L}_c} \hat{\mathbf{B}}_{\text{bus}} \mathbf{A}_{\mathcal{L}_c}^\top \right)^{-1}, \quad (1.14)$$

où :

- $\mathbf{X} \in \mathbb{R}^{|\mathcal{L}| \times |\mathcal{L}|}$  est une matrice diagonale dont chaque terme diagonal  $\mathbf{X}_{(i,j),(i,j)}$  est égal à  $1/b_{ij}$ , i.e. l'inverse de la susceptance de la ligne  $(i, j) \in \mathcal{L}$ ;
- $\mathbf{A}_{(i,j)} \in \mathbb{R}^{1 \times |\mathcal{N}|}$  est la ligne de la matrice d'incidence orientée  $\mathbf{A}$  définie par (1.13), associée à la ligne de transport  $(i, j)$ ;
- $\mathbf{A}_{\mathcal{L}_c} \in \mathbb{R}^{|\mathcal{L}_c| \times |\mathcal{N}|}$  est la sous-matrice de  $\mathbf{A}$  définie par (1.13) constituée des lignes correspondant aux lignes en panne  $(k, l) \in \mathcal{L}_c$ ;
- $\mathbf{X}_{\mathcal{L}_c} \in \mathbb{R}^{|\mathcal{L}_c| \times |\mathcal{L}_c|}$  est la sous-matrice diagonale de  $\mathbf{X}$  correspondant aux lignes de  $\mathcal{L}_c$ ;
- $\mathbf{I}$  est la matrice identité de dimension  $|\mathcal{L}_c|$  par  $|\mathcal{L}_c|$ ;
- $\hat{\mathbf{B}}_{\text{bus}} \in \mathbb{R}^{|\mathcal{N}| \times |\mathcal{N}|}$  désigne la matrice de susceptance nodale inverse définie par (1.4).

Cette formulation explicite permet de calculer le facteur de la matrice  $\Lambda$  pour tout scénario  $c \in \mathcal{C}$  à partir de la structure du réseau et de la matrice  $\hat{\mathbf{B}}_{\text{bus}}$ , sans réinversion globale. Dans le cas particulier d'une contingence simple, i.e.,  $|\mathcal{L}_c| = 1$ , la matrice  $\Lambda$  s'exprime comme un scalaire :

$$\Lambda_{(i,j),(k,l)} = \frac{b_{ij} \mathbf{A}_{(i,j)} \hat{\mathbf{B}}_{\text{bus}} \mathbf{A}_{(k,l)}^\top}{1 - b_{kl} \mathbf{A}_{(k,l)} \hat{\mathbf{B}}_{\text{bus}} \mathbf{A}_{(k,l)}^\top}.$$

La matrice  $\Lambda$  ainsi obtenues peuvent être utilisées pour former une matrice  $\Phi$  post-contingence sous contraintes de sécurité  $N - 1$  ou  $N - k$ , notée  $\Phi^{(c)}$ , qui ajuste la matrice de transfert de base en intégrant les effets redistributifs de la défaillance [13]. Formellement, cette matrice

peut être exprimée comme une perturbation affine de la matrice pré-contingence :

$$\Phi^{(c)} = \Phi^{(0)} + \Delta^{(c)}, \quad (1.15)$$

où  $\Delta^{(c)} \in \mathbb{R}^{|\mathcal{L}| \times |\mathcal{N}|}$  est une matrice de correction dépendant du scénario de contingence  $c \in \mathcal{C}$ , obtenue à partir des facteurs de la matrice  $\Lambda$ .

Plus explicitement, soit  $\Phi_{\mathcal{L}_c}^{(0)} \in \mathbb{R}^{|\mathcal{L}_c| \times |\mathcal{N}|}$  la sous-matrice de  $\Phi^{(0)}$  contenant uniquement les lignes affectées. La matrice  $\Phi^{(c)}$  s'écrit alors comme :

$$\Phi^{(c)} = \Phi^{(0)} + \begin{bmatrix} \Lambda_{(1),\mathcal{L}_c} \Phi_{\mathcal{L}_c}^{(0)} \\ \Lambda_{(2),\mathcal{L}_c} \Phi_{\mathcal{L}_c}^{(0)} \\ \vdots \\ \Lambda_{(|\mathcal{L}|),\mathcal{L}_c} \Phi_{\mathcal{L}_c}^{(0)} \end{bmatrix}, \quad (1.16)$$

où chaque ligne  $(i,j) \notin \mathcal{L}_c$  de la matrice  $\Phi^{(c)}$  est ajustée par une combinaison linéaire pondérée des lignes affectées dans  $\Phi^{(0)}$ , selon les coefficients  $\Lambda_{(i,j),\mathcal{L}_c}$ . Cette structure permet de propager analytiquement l'effet d'une défaillance topologique sur l'ensemble des écoulements de puissance attendus, sans recalcule explicite des angles noraux ni inversion supplémentaire de la matrice de susceptance [12].

Dans un contexte multi-contingence, les limites thermiques applicables aux lignes de transport varient d'un scénario à l'autre. En effet, la perte d'une ligne  $(k,l) \in \mathcal{L}$  entraîne une reconfiguration du réseau qui modifie non seulement les coefficients de sensibilité des écoulements de puissance, mais également les contraintes opérationnelles associées. Par exemple, certaines lignes initialement redondantes, car sans impact sur la solution optimale, peuvent devenir critiques après défaillance d'un autre élément du réseau. À l'inverse, d'autres lignes peuvent, sous conditions d'exploitation spécifiques, être temporairement autorisées à dépasser leur limite nominale : c'est le cas des limites d'urgences (*emergency ratings*) définies par MISO, qui désignent les niveaux admissibles de circulation de puissance active en régime transitoire post-contingence, notamment pour éviter un délestage immédiat ou pour maintenir la stabilité du système durant les premières secondes suivant l'événement [7]. Cette flexibilité opérationnelle est souvent intégrée sous la forme de limites de contingence spécifiques, notées  $\bar{s}^{(c)}$  pour chaque scénario  $c \in \mathcal{C}$ .

La contrainte de sécurité imposée à chaque scénario s'exprime alors sous forme matricielle comme suit :

$$-\bar{s}^{(c)} \leq \Phi^{(c)} \mathbf{p} \leq \bar{s}^{(c)}, \quad (1.17)$$

où  $\mathbf{p} \in \mathbb{R}^{|\mathcal{N}|}$  est le vecteur des puissances actives injectées aux noeuds. L'ensemble de ces

contraintes, valides pour chaque  $c \in \mathcal{C}$ , définit l'espace de faisabilité sécurisé du problème.

Le problème d'écoulement de puissance optimal sous contrainte de sécurité (*security-constrained optimal power flow*, SCOPF) s'obtient alors comme extension directe du modèle DCOPF standard. Ce dernier généralise le modèle DCOPF en exigeant que la solution soit réalisable non seulement dans l'état pré-contingence, mais également dans tous les états post-contingence considérés. Sa formulation compacte, fondée sur les matrices  $\Phi^{(c)}$ , s'écrit :

$$\min_{\mathbf{p}} \quad f(\mathbf{p}) \quad (1.18a)$$

$$\text{s.c.} \quad \sum_{i \in \mathcal{N}} p_i = 0 \quad (1.18b)$$

$$\underline{p}_i \leq p_i \leq \bar{p}_i, \quad \forall i \in \mathcal{N} \quad (1.18c)$$

$$-\bar{s}_{ij}^{(0)} \leq \sum_{k \in \mathcal{N}} \Phi_{ijk}^{(0)} p_k \leq \bar{s}_{ij}^{(0)}, \quad \forall (i, j) \in \mathcal{L} \quad (1.18d)$$

$$-\bar{s}_{ij}^{(c)} \leq \sum_{k \in \mathcal{N}} \Phi_{ijk}^{(c)} p_k \leq \bar{s}_{ij}^{(c)}, \quad \forall (i, j) \in \mathcal{L}, \quad \forall c \in \mathcal{C} \quad (1.18e)$$

Le modèle (1.18) conserve les contraintes d'équilibre (1.18b) et de production (1.18c), tout en imposant le respect des limites thermiques nominales (1.18d) et contingentes (1.18e). Ces dernières s'appliquent à l'aide des matrices  $\Phi^{(c)}$ , définies pour chaque scénario  $c \in \mathcal{C}$ , ainsi que des bornes  $\bar{s}^{(c)}$  propres à chaque état post-contingence.

Dans cette formulation, une contingence unique correspond typiquement à la mise hors service d'une unique ligne de transport  $(i, j) \in \mathcal{L}$ , tandis qu'une contingence multiple recouvre soit la perte simultanée de plusieurs lignes, soit la défaillance d'un nœud  $i \in \mathcal{N}$ , laquelle entraîne de fait même la perte de l'ensemble des lignes incidentes à ce nœud. Cette modélisation permet d'anticiper de manière préventive les effets de telles défaillances sur les écoulements de puissance. Elle se heurte toutefois à une croissance combinatoire du nombre de contraintes associées, rendant d'autant plus pertinent le développement de méthodes de résolution efficaces.

#### 1.2.4 Répartition économique optimale de la production sous contraintes de sécurité

En réalité, des marchés de l'électricité comme celui de MISO intègrent un très grand nombre de contraintes opérationnelles dans leur formulation de SCED. Ces contraintes excèdent largement le seul respect des limites thermiques post-contingence. Le modèle mis en œuvre dans le DA de MISO, tel que documenté dans [6] et [7], inclut par exemple :

- des contraintes physiques de rampes, de délais de démarrage, d'interruptibilité et de modulation de réserve ;
- des contraintes économiques liées aux engagements et à la gestion des offres horaires de puissance et de réserve ;
- des contraintes de sécurité relatives aux exigences zonales en réserve tournante, réserve supplémentaire et régulation de fréquence, modulées par des courbes de demande dépendant des conditions de rareté ;
- ou encore, des règles spécifiques dites de «bonnes pratiques opérationnelles» (*good utility practices*) qui, bien que non strictement physiques, sont modélisées comme contraintes souples afin d'assurer la faisabilité de l'exploitation prévisionnelle.

Cette formulation étendue conduit à un problème d'optimisation de très grande taille lorsqu'appliquée à des réseaux réels. Dans ce contexte, les limites thermiques sur l'écoulement de puissance (1.18d), ainsi que les limites de contingence (1.18e), sont modélisées comme des contraintes « souples » via des variables de dépassement pénalisées, permettant de garantir une solution réalisable même en conditions tendues, au prix d'ajustements explicites dans l'objectif économique du marché [6]. À titre d'exemple, le marché de MISO applique des pénalités de dépassement explicites pouvant atteindre 1500\$/MW, tant pour les contraintes thermiques nominales que contingentes [14]. Cette modélisation assure une hiérarchisation explicite entre les objectifs économiques et les impératifs de sécurité du réseau, tout en garantissant la continuité et la cohérence économique des prix nodaux dans les situations limites, y compris lorsque certaines contraintes physiques sont temporairement relâchées.

Dans le présent mémoire, nous considérons une formulation simplifiée dérivée de ce cadre opérationnel : elle isole les contraintes les plus structurelles du problème afin de mettre en évidence les défis computationnels associés au grand nombre de scénarios de contingence.

La version pénalisée du modèle (1.18), désignée ici comme SCED, introduit des variables de dépassement notées  $\xi_{ij,+}^{(c)}$  et  $\xi_{ij,-}^{(c)}$  associées aux limites thermiques de chaque ligne  $(i, j) \in \mathcal{L}$  pour chaque scénario  $c \in \mathcal{C}$  en incluant le scénario nominal. Sa formulation s'écrit :

$$\min_{\mathbf{p}, \boldsymbol{\xi}_-, \boldsymbol{\xi}_+} f(\mathbf{p}) + \sum_{(i,j) \in \mathcal{L}} M \left( \xi_{ij,+}^{(c)} + \xi_{ij,-}^{(c)} \right) \quad (1.19a)$$

$$\text{s.c.} \quad \sum_{i \in \mathcal{N}} p_i = 0 \quad (1.19b)$$

$$\underline{p}_i \leq p_i \leq \bar{p}_i, \quad \forall i \in \mathcal{N} \quad (1.19c)$$

$$\sum_{k \in \mathcal{N}} \Phi_{ijk} p_k \leq \bar{s}_{ij} + \xi_{ij,+}, \quad \forall (i, j) \in \mathcal{L}, \quad (1.19d)$$

$$-\sum_{k \in \mathcal{N}} \Phi_{ijk} p_k \leq \bar{s}_{ij} + \xi_{ij,-}, \quad \forall (i,j) \in \mathcal{L}, \quad (1.19e)$$

$$\xi_{ij,+} \geq 0, \quad \xi_{ij,-} \geq 0, \quad \forall (i,j) \in \mathcal{L}, \quad (1.19f)$$

$$\min_{\mathbf{p}, \xi_-, \xi_+} f(\mathbf{p}) + \sum_{c \in \mathcal{C}} \sum_{(i,j) \in \mathcal{L}} M \left( \xi_{ij,+}^{(c)} + \xi_{ij,-}^{(c)} \right) \quad (1.20a)$$

$$\text{s.c.} \quad \sum_{i \in \mathcal{N}} p_i = 0 \quad (1.20b)$$

$$\underline{p}_i \leq p_i \leq \bar{p}_i, \quad \forall i \in \mathcal{N} \quad (1.20c)$$

$$\sum_{k \in \mathcal{N}} \Phi_{ijk}^{(c)} p_k \leq \bar{s}_{ij}^{(c)} + \xi_{ij,+}^{(c)}, \quad \forall (i,j) \in \mathcal{L}, \forall c \in \mathcal{C} \quad (1.20d)$$

$$-\sum_{k \in \mathcal{N}} \Phi_{ijk}^{(c)} p_k \leq \bar{s}_{ij}^{(c)} + \xi_{ij,-}^{(c)}, \quad \forall (i,j) \in \mathcal{L}, \forall c \in \mathcal{C} \quad (1.20e)$$

$$\xi_{ij,+}^{(c)} \geq 0, \quad \xi_{ij,-}^{(c)} \geq 0, \quad \forall (i,j) \in \mathcal{L}, \forall c \in \mathcal{C}, \quad (1.20f)$$

où  $M > 0$  est l'amplitude de la pénalité. Le modèle (1.20) conserve les contraintes structurelles du SCOPF (1.18), mais autorise un dépassement contrôlé des limites d'écoulement sur les lignes de transport via les termes  $\xi_{ij,\pm}^{(c)}$ , dont l'impact est reflété dans l'objectif par des coefficients de pénalité  $M$ . Ces coefficients peuvent être choisis de manière réaliste à partir des valeurs utilisées dans les marchés opérationnels : par exemple, MISO applique des pénalités pouvant atteindre \$1500/MW pour les dépassements thermiques [14].

Soit  $\mathcal{G}$  l'ensemble des générateurs du réseau électrique. La fonction  $f(\mathbf{p})$  représente le coût de production, généralement modélisé comme une fonction quadratique convexe de la forme :

$$f(\mathbf{p}) = \sum_{i \in \mathcal{G}} a_i p_i^2 + b_i p_i + c_i$$

où  $a_i \geq 0$ ,  $b_i \in \mathbb{R}$  et  $c_i \in \mathbb{R}$  sont les coefficients de coût spécifiques à chaque générateur  $i \in \mathcal{G}$ . Cette approximation quadratique constitue un standard de par sa représentativité aux tendances des courbes de coût réelles des unités thermiques, tout en garantissant la convexité du problème.

### 1.2.5 Prix nodaux et arbitrage : dérivation des LMP

Dans le cadre du modèle de SCED (1.20), les LMP sont interprétés comme les multiplicateurs de Lagrange associés à la contrainte d'équilibre de puissance nodale (1.20b). Plus précisément,

si nous désignons par  $\lambda \in \mathbb{R}$  le multiplicateur global associé à la contrainte d'équilibre (1.20b), et par  $\mu_{ij}^{(c)} \in \mathbb{R}_+$  les multiplicateurs associés aux contraintes d'écoulement (1.20d)–(1.20e), alors le LMP au noeud  $i \in \mathcal{N}$  du scénario  $c \in \mathcal{C}$  à l'instant  $t$  peut s'exprimer comme :

$$\text{LMP}_i^{(c)} = \lambda - \sum_{(i,j) \in \mathcal{L}} \mu_{ij}^{(c)} \Phi_{ij,i}^{(c)}.$$

Cette expression, dérivée d'un modèle linéarisé sans pertes, traduit une interprétation économique du dual : le LMP correspond au coût marginal d'approvisionnement d'une unité supplémentaire de demande au noeud  $i$ , tenant compte du coût de l'énergie produite ( $\lambda$ ) et des contraintes de congestion locales. Dans une formulation plus complète incluant les pertes électriques, une composante supplémentaire reflétant l'effet marginal des pertes serait ajoutée.

Dans le cas où les contraintes d'écoulement sont rigides (*hard constraints*), chaque contrainte saturée génère un multiplicateur strictement positif, ce qui permet d'interpréter la composante congestion des LMP comme les coûts marginaux liés aux limitations du réseau. Cependant, dans les formulations opérationnelles où les contraintes thermiques nominales et post-contingence sont modélisées comme souples, alors, dès lors qu'une contrainte est non respectée, i.e.,  $\xi_{ij,\pm}^{(c)} > 0$ , elle n'est plus active au sens dual : la contrainte physique elle-même est relâchée, et le multiplicateur de Lagrange associé à la contrainte de positivité  $\mu_{ij}^{(c)}$  devient nul. Ainsi, une ligne effectivement congestionnée au sens physique ne génère alors aucun prix de congestion dans les LMP. Inversement, un multiplicateur positif peut provenir d'une contrainte qui serait active uniquement du fait de la pénalisation. Pour résoudre cette incohérence entre signal de congestion physique et prix de marché, MISO implémente une procédure en deux étapes [6] :

1. Dans une première passe, le problème est résolu avec contraintes souples, ce qui permet d'identifier les pénalités effectives  $\xi_{ij}^{*(c)}$  nécessaires à la faisabilité ;
2. Dans une seconde passe, un nouveau SCED est résolu, en remplaçant les contraintes initiales par des contraintes rigides assouplies à  $\bar{s}_{ij}^{(c)} + \xi_{ij}^{*(c)} + \delta_{ij}^{(c)}$ , avec  $\delta_{ij}^{(c)}$  un relâchement marginal (ex. : 5%), afin de désactiver les variables de pénalité. La contrainte de congestion post-contingence devient alors :

$$\sum_{k \in \mathcal{N}} \Phi_{ijk}^{(c)} p_k \leq \bar{s}_{ij}^{(c)} + \xi_{ij}^{*(c)} + \delta_{ij}^{(c)}, \quad \forall (i, j) \in \mathcal{L}, \forall c \in \mathcal{C}.$$

Cette deuxième résolution, sans activation des pénalités, garantit que les multiplicateurs de Lagrange  $\mu_{ij}^{(c)}$  sont à nouveau économiquement valides, et que les LMP obtenus reflètent fidèlement les congestions structurelles du réseau. Cette méthode permet donc de préserver la cohérence économique des signaux de prix tout en maintenant la robustesse opérationnelle du

modèle de marché.

Les composantes de congestion des LMP traduisent le coût marginal d'une contrainte active sur le réseau. Lorsqu'une ligne de transport atteint sa limite thermique, i.e., devient active, les injections de puissance au nœud en amont et les retraits au nœud en aval ne sont plus substituables sans coût. Le marché reflète alors cette rareté physique sous la forme d'un différentiel de prix entre les nœuds concernés, appelé prix de congestion (*shadow price*). Ce différentiel crée une incitation directe à l'arbitrage spatial. Un participant au marché capable d'acheter de la puissance à un nœud à bas prix et de la revendre à un nœud à prix plus élevé, tout en respectant les équilibres de puissance et la topologie du réseau, peut en tirer un profit proportionnel à la différence de LMP. Toutefois, ces opportunités d'arbitrage physique demeurent fortement limitées. D'une part, les congestions sont variables car elles dépendent des profils de charge, des injections intermittentes de production renouvelable et des aléas topologiques liés aux contingences. D'autre part, les conditions d'accès au réseau, les incertitudes sur les prix futurs et les restrictions d'interconnexion peuvent empêcher un acteur de mettre effectivement en œuvre un transfert physique profitable. Ainsi, bien que les différentiels de LMP signalent des opportunités de gains, la réalisation effective de ces arbitrages reste incertaine, coûteuse, et sujette à des restrictions structurelles et réglementaires.

Pour offrir une protection contre ces états de congestion, les marchés électriques ont développé des instruments financiers dédiés tels que les FTR. Un FTR associé à un chemin directionnel ( $i \rightarrow j$ ) donne à son détenteur le droit de percevoir, ex post, la différence de prix nodaux entre les nœuds source et puits. Son revenu est donné par :

$$\text{Revenue} = (\text{LMP}_j - \text{LMP}_i)Q,$$

où  $Q \in \mathbb{R}$  désigne la quantité de puissance transférée sous forme contractuelle. Le signe de  $Q$  encode la direction économique du transfert : un FTR avec  $Q < 0$  sur le chemin ( $i \rightarrow j$ ) équivaut à un FTR positif sur le chemin opposé ( $j \rightarrow i$ ). D'un point de vue économique, les FTR s'interprètent comme la contrepartie duale des contraintes actives d'écoulement de puissance (1.20d)–(1.20e), et la valeur agrégée des FTR est ainsi égale à la somme des produits entre les multiplicateurs de Lagrange de ces contraintes [15].

### 1.3 Défis et opportunités actuels du SCED dans les marchés

Comme discuté à la section 1.1.2, les ISO et les RTO sont tenus de résoudre à une très haute fréquence des problèmes de SCED, notamment dans le cadre des marchés en temps réel, où un SCED est exécuté toutes les 5 minutes. Cette exigence temporelle, déjà structurellement

contraignante, est amplifiée par un phénomène plus fondamental : la montée en incertitude à toutes les échelles de décision. L'intégration croissante des sources d'énergie renouvelable, par nature intermittentes et spatialement corrélées, ainsi que la présence d'une asymétrie informationnelle inhérente aux marchés concurrentiels, introduisent une pluralité de scénarios plausibles que les opérateurs et autres participants doivent envisager, simuler ou anticiper. Ces incertitudes couvrent aussi bien les injections, la demande, la disponibilité des lignes, que les règles des marchés elles-mêmes [16]. En pratique, il ne s'agit donc plus de résoudre un SCED unique, mais bien un grand nombre de variantes, associées à différents états anticipés du réseau, pour dériver des décisions robustes, évaluées selon des critères d'espérance ou de regret maximal [17]. L'accroissement de cette fréquence temporelle couplée à la multiplicité des scénarios opérationnels accentue fortement la pression sur les pipelines de résolution, tant au niveau du volume d'instances à traiter que de leur diversité structurelle.

Toutefois, ces résolutions successives se heurtent à une complexité structurelle profonde, liée à la nature même des contraintes de sécurité imposées dans le modèle (1.20). Les contraintes d'écoulement nominal et post-contingence y sont formulées à l'aide des matrices  $\Phi$  et  $\Phi^{(c)}$ , dont la structure dense impose une connectivité quasi-totale entre les injections nodales et les lignes surveillées. Cette densité a pour effet d'induire un couplage généralisé entre les variables, et par conséquent, de transformer chaque contrainte thermique en une inégalité dense, affectant potentiellement l'ensemble du vecteur des productions. Dans les formulations purement PTDF, ce phénomène peut même rendre les fichiers de problème extrêmement volumineux à générer et à lire, et ralentir significativement les étapes de présolution [18]. Le problème est aggravé lorsque le nombre de lignes sujettes à des contingences est important, ou que nous cherchons à représenter un critère de sécurité de type  $N-1$  ou  $N-k$ , qui impose l'introduction explicite d'un ensemble  $\mathcal{C}$  de scénarios post-contingence, chacun associé à une matrice  $\Phi^{(c)}$  distincte et à ses propres contraintes. De plus, dans une formulation à contraintes souples, chaque inégalité liée aux contraintes de lignes thermiques contingentes active potentiellement une variable de pénalité  $\xi_{ij}^{(c)}$ , elle-même dualisée, ce qui amplifie significativement la taille du problème d'optimisation, tant du côté primal (nombre de variables) que du côté dual (nombre de multiplicateurs associés).

Or, chaque contrainte thermique contingente introduit deux nouvelles inégalités dans le modèle (1.20), correspondant aux bornes supérieure et inférieure de l'écoulement sur la ligne considérée après contingence. Ainsi, même pour un critère de sécurité limité au cas  $N-1$ , le nombre total de contraintes de congestion post-contingence croît en  $\mathcal{O}(|\mathcal{C}||\mathcal{L}|)$ , soit typiquement plusieurs dizaines de milliers d'inégalités dans les réseaux de taille réaliste. De plus, ces contraintes doivent être réévaluées à chaque pas de temps dans les problèmes SCED lors de l'horizon roulant, ce qui rend la résolution globale rapidement intractable à mesure

que nous augmentons le nombre de scénarios ou l'horizon temporel [12].

Plus fondamentalement, l'effet de cette croissance combinatoire sur la complexité effective du modèle est amplifié par la densité des matrices  $\Phi^{(c)}$ , qui confère à chaque inégalité une portée globale sur les variables de décision. La région réalisable résultante devient alors définie par un très grand nombre d'hyperplans denses, fortement interdépendants, ce qui limite sévèrement les possibilités d'élimination structurelle de contraintes ou de factorisation parcimonieuse dans les méthodes de résolution. Bien que nous puissions démontrer que seul un sous-ensemble de ces inégalités est effectivement nécessaire pour décrire le polytope [12, 19, 20] lors de la résolution du SCED, cette réduction n'est généralement pas connue a priori et dépend fortement de l'état anticipé du réseau. Cette densité structurelle affecte directement la résolution numérique, notamment dans les méthodes de points intérieurs (*interior point method*, IPM), où chaque itération implique la factorisation d'un système KKT couplant les variables primales et duales [21]. Or, la complexité de cette étape dépend de la structure des blocs associés aux contraintes, dont la taille et la densité conditionnent la possibilité d'exploiter des algorithmes creux. En présence de matrices  $\Phi^{(c)}$  denses, le solveur perd toute opportunité de tirer parti de la structure topologique clairsemée du réseau [22]. Il en résulte un coût de factorisation plus élevé, une croissance des temps de présolution, et une convergence plus lente, même dans des cas où la topologie physique sous-jacente reste relativement simple.

Ce constat met en lumière un dilemme fondamental entre robustesse opérationnelle et efficacité algorithmique. D'une part, la garantie de faisabilité décisionnelle pour l'ensemble des scénarios de contingence impose, en théorie, l'inclusion simultanée de l'ensemble des contraintes thermiques post-contingence, ainsi que des variables de pénalité associées. Cette inclusion exhaustive alourdit significativement la structure du problème, tant au niveau primal, par l'introduction de variables supplémentaires, qu'au niveau dual, par la croissance du système KKT induit. Elle se traduit inévitablement par une dégradation des performances des solveurs. D'autre part, de nombreux travaux ont mis en évidence que, malgré leur grand nombre, seule une fraction réduite des contraintes de contingence est réellement nécessaire pour caractériser l'ensemble réalisable [12], de même qu'une fraction encore plus marginale de ces contraintes sont effectivement actives [23]. En effet, pour une large classe de profils d'injection, la majorité de ces inégalités demeurent strictement inactives, et ne participent ni à la définition locale du point optimal, ni à la structure globale du polytope admissible. Ce déséquilibre structurel suggère l'existence d'un potentiel substantiel de réduction dimensionnelle : l'identification d'un sous-ensemble minimal de contraintes strictement nécessaires permettrait, en principe, d'alléger la formulation sans compromettre ni la faisabilité, ni la qualité économique et ni la sécurité du réseau électrique associé à la solution.

## 1.4 Question de recherche et objectifs du mémoire

Dans ce contexte de croissance dimensionnelle, de formulation à contraintes souples et de redondance structurelle massive, ce mémoire explore la question suivante :

*Comment identifier, pour une instance donnée du problème de SCED, un sous-ensemble minimal de contraintes de contingence pertinentes, suffisant à garantir l'optimalité et la sécurité opérationnelle du réseau tout en minimisant la surcharge computationnelle imposée au solveur ?*

Ce problème soulève des enjeux tant géométriques qu'algorithmiques. Il s'agit d'identifier les contraintes réellement nécessaires à la définition de l'ensemble admissible utile dans une formulation à contraintes souples, c'est-à-dire les inégalités qui, même en l'absence de saturation à l'optimum, conditionnent structurellement la solution du problème. L'objectif est de construire une représentation réduite du modèle, de complexité maîtrisée, tout en préservant l'exactitude de la solution optimale du SCED.

Pour répondre à cette problématique, le mémoire propose une approche hybride fondée sur trois composantes complémentaires :

1. **Une réduction hiérarchique structurée des contraintes de contingence**, obtenue par élimination récursive des inégalités redondantes dans la formulation pénalisée du SCED, conduisant à une décomposition imbriquée en niveaux de contraintes ;
2. **L'exploitation explicite du relâchement des contraintes par pénalisation**, garantissant la faisabilité dans chaque instance tout en autorisant un dépassement contrôlé des limites thermiques en cas de congestion extrême ;
3. **La prédiction supervisée du niveau hiérarchique optimal de réduction à activer**, à partir du profil de charge nodale, via un modèle de classification cumulative exploitant la hiérarchie de contraintes extraite en amont.

L'articulation de ces trois composantes vise à concevoir une stratégie de résolution adaptative, capable d'ajuster dynamiquement le périmètre de contraintes considérées en fonction des caractéristiques de chaque instance, tout en maintenant les garanties de faisabilité exigées dans les environnements temps réel des marchés électriques libéralisés. L'ensemble de la méthode repose sur un apprentissage effectué hors ligne à partir de simulations, de manière à permettre un déploiement en ligne efficace.

Ce travail s'inscrit dans un contexte industriel où les mécanismes de marché et les contraintes opérationnelles réelles jouent un rôle central. Un effort substantiel a été consacré à comprendre

en profondeur les règles de fonctionnement des marchés nord-américains, en particulier la logique de planification, le calcul des LMP, et les traitements des contraintes souples. Cette compréhension a directement motivé la structure détaillée de la revue de littérature, qui ne se limite pas à une synthèse académique, mais cherche à établir un pont rigoureux entre les formulations mathématiques et les pratiques concrètes mises en œuvre par les opérateurs de marché. La section 2, en particulier, constitue une contribution originale de ce mémoire et fait l'objet d'un article scientifique en cours de rédaction.

Les chapitres 2, 3 et 4 présentés dans ce mémoire constituent la base de deux articles scientifiques actuellement en cours de rédaction. D'une part, nous proposerons une revue de littérature structurée sur les stratégies de résolution efficaces du SCED, comblant un vide dans l'état de l'art en contextualisant les approches de décomposition, d'élagage et de réduction de contraintes dans les formulations pénalisées. D'autre part, nous formalisons une méthode de réduction de contraintes guidée par apprentissage supervisé, spécifiquement adaptée aux formulations du SCED intégrant des variables de pénalité, et capable de conserver l'exactitude de la solution. La longueur du présent mémoire reflète ainsi cette double contribution.

## CHAPITRE 2 REVUE DE LA LITTÉRATURE

Dans ce chapitre, une revue de la littérature portant sur les approches de résolution efficaces pour le SCED et, par extension, le SCOPF est présentée. Nous nous concentrons plus particulièrement sur les méthodes de décomposition adaptées aux problèmes à très haute dimension, les techniques de réduction de contraintes dans les formulations à contingences multiples, et les approches d'apprentissage supervisé visant la sélection active de contraintes pertinentes.

### 2.1 Méthodes de décomposition et de calcul parallèle pour le SCED à grande échelle

Les formulations réalistes des problèmes de SCED et de SCOPF donnent lieu à des modèles de grande dimension, souvent trop complexes pour être résolus efficacement de manière monolithique. En particulier, la prise en compte explicite de contingences multiples, de contraintes de sécurité associées à de nombreuses lignes de transport, ou encore l'extension à des horizons multipériodes, conduit à une explosion du nombre de contraintes et de variables. Ces enjeux motivent le recours à des méthodes de décomposition, qui visent à exploiter la structure interne du problème afin d'en améliorer la mise à l'échelle.

L'idée générale des méthodes de décomposition consiste à reformuler le problème initial sous la forme d'un ensemble de sous-problèmes plus simples, souvent résolus indépendamment, coordonnés par un problème maître ou par des mécanismes de dualisation. Ces approches permettent non seulement de réduire l'empreinte mémoire de la résolution, mais également de tirer parti des architectures parallèles pour accélérer significativement les temps de calcul [24].

Dans le contexte des SCED et SCOPF, plusieurs structures de couplage naturelles peuvent être exploitées, notamment la séparation explicite entre conditions normales et conditions de contingence [25, 26], permettant une décomposition par scénario, ou bien la distinction entre variables de commande globales (communes à tous les scénarios) et variables locales spécifiques à chaque configuration post-contingence [27]. Ces propriétés structurantes ont donné lieu à diverses approches, notamment la méthode de Benders [24], particulièrement bien adaptée à la structure du SCOPF [25], ainsi que des variantes reposant sur la méthode des multiplicateurs alternés (*alternating direction method of multipliers*, ADMM) [28–30], qui permettent une parallélisation naturelle et un recours efficace aux techniques de démarrage à chaud (*warm-start*), i.e., à la réutilisation d'une solution approximative comme point de

départ pour accélérer la convergence des itérations ultérieures. D'autres travaux ont également proposé des techniques spécialisées fondées sur des approximations structurelles des termes de pénalité associés aux contingences [31], ou combinant plusieurs mécanismes de coordination dans des architectures distribuées [27].

Les sections suivantes examinent ces différentes familles de méthodes, en mettant l'accent sur leur capacité à traiter efficacement des instances de grande taille dans un cadre opérationnel, tout en conservant les garanties de faisabilité requises par les exploitants de réseau.

### 2.1.1 Méthodes de décomposition de Benders

La méthode de décomposition de Benders [24] constitue une approche classique et bien établie pour traiter des problèmes d'optimisation présentant une structure de couplage entre deux familles de variables.

Soit  $\mathcal{U} \subseteq \mathbb{R}^m$ ,  $m \in \mathbb{N}$ , l'ensemble admissible des variables compliquantes, dont la spécification, i.e., la considération comme paramètres fixés dans une approche décomposée, rend le problème plus facile à traiter, souvent liées à des décisions globales ou entière. Soit  $\mathcal{Z}(\mathbf{u}) \subseteq \mathbb{R}^n$ ,  $n \in \mathbb{N}$ , l'ensemble admissible des variables locales ou secondaires, pour une configuration donnée  $\mathbf{u} \in \mathcal{U}$ . Définissons  $\mathbf{u} \in \mathbb{R}^m$  comme les variables maîtres, prises en compte dans le problème principal, et  $\mathbf{z} \in \mathbb{R}^n$ , comme les variables locales, spécifiques aux sous-problèmes paramétrés par la valeur courante de  $\mathbf{u}$ .

Le problème d'optimisation global s'écrit alors sous la forme suivante :

$$\min_{\mathbf{u} \in \mathcal{U}} \quad \mathbf{c}^\top \mathbf{u} + \alpha(\mathbf{u}) \quad (2.1a)$$

$$\text{où } \alpha(\mathbf{u}) := \min_{\mathbf{z} \in \mathcal{Z}(\mathbf{u})} \mathbf{d}^\top \mathbf{z}, \quad (2.1b)$$

où le vecteur  $\mathbf{c} \in \mathbb{R}^m$  représente les coefficients de coût associés aux variables compliquantes, tandis que  $\mathbf{d} \in \mathbb{R}^n$  encode les coûts relatifs aux variables locales. L'objectif du problème principal (2.1a) est donc de minimiser un coût total composé d'un terme explicite  $\mathbf{c}^\top \mathbf{u}$  et d'un terme implicite  $\alpha(\mathbf{u})$ , évalué via la résolution du sous-problème (2.1b).

La définition (2.1b) désigne ici la fonction valeur optimale associée au sous-problème paramétré par  $\mathbf{u}$ , i.e., la meilleure valeur du coût correctif  $\mathbf{d}^\top \mathbf{z}$  que nous obtenons en résolvant le sous-problème de plus petite dimension restreint à  $\mathcal{Z}(\mathbf{u})$ .

La fonction valeur  $\alpha(\mathbf{u})$  n'étant généralement pas disponible sous forme explicite, l'algorithme procède par approximations successives, à l'aide de coupes (plans de support) construites à partir des solutions des sous-problèmes. À chaque itération  $k \in \mathcal{K}$ , où  $\mathcal{K} \subset \mathbb{N}$  désigne

l'ensemble des indices des itérations précédentes, le problème maître résout une relaxation de (2.1) enrichie par les coupes obtenues à partir des sous-problèmes résolus en  $\mathbf{u}^{(j)}$ , pour tout  $j < k$  :

$$\min_{\mathbf{u}, \eta} \quad c^\top \mathbf{u} + \eta \quad (2.2a)$$

$$\text{s.c.} \quad \eta \geq \boldsymbol{\lambda}^{(k)\top}(\mathbf{u} - \mathbf{u}^{(k)}) + \theta^{(k)}, \quad \forall k \in \mathcal{K} \quad (2.2b)$$

$$\mathbf{u} \in \mathcal{U}, \quad (2.2c)$$

où  $(\boldsymbol{\lambda}^{(k)}, \theta^{(k)})$  désignent respectivement le gradient et l'ordonnée à l'origine d'une coupe affine obtenue à partir de la dualité du sous-problème résolu en  $\mathbf{u}^{(k)}$ . Ces coupes peuvent représenter :

- soit une coupe de faisabilité, si le sous-problème est infaisable en  $\mathbf{u}^{(k)}$ , destinée à exclure cette solution du domaine admissible ;
- soit une coupe d'optimalité, si le sous-problème est réalisable, permettant de raffiner l'approximation de  $\alpha(\mathbf{u})$ .

Dans le cas où le problème maître et les sous-problèmes sont convexes, la convergence vers une solution optimale est garantie après un nombre fini de coupes [24]. Cette stratégie présente deux avantages principaux : elle permet une réduction de la dimension du problème central en évitant l'introduction explicite de toutes les variables du système couplé, et elle autorise une résolution parallèle des sous-problèmes.

La méthode proposée dans [26] repose sur une décomposition explicite du SCOPF en un problème maître, représentant les conditions normales d'exploitation du réseau électrique, et une collection de sous-problèmes, chacun correspondant à un scénario de contingence unique. Le problème maître (2.3) formulé par :

$$\min_{\mathbf{x}_0, \mathbf{u}_0} \quad f_0(\mathbf{x}_0, \mathbf{u}_0) \quad (2.3a)$$

$$\text{s.c.} \quad g_0(\mathbf{x}_0, \mathbf{u}_0) = 0 \quad (2.3b)$$

$$h_0(\mathbf{x}_0, \mathbf{u}_0) \leq \bar{\mathbf{s}}, \quad (2.3c)$$

cherche une solution optimale  $(\mathbf{x}_0^*, \mathbf{u}_0^*)$  sous les contraintes d'égalité (2.3b) et d'inégalité (2.3c) du régime pré-contingence, en minimisant le coût opérationnel (2.3a) associé aux variables de contrôle préventif  $\mathbf{u}_0 \in \mathbb{R}^m$ .

Pour chaque scénario de contingence  $c \in \mathcal{C}$ , un sous-problème défini par (2.4)

$$\min_{\mathbf{x}_c, \boldsymbol{\epsilon}_c} \quad \mathbf{1}^\top \boldsymbol{\epsilon}_c \quad (2.4a)$$

$$\text{s.c.} \quad g_c(\mathbf{x}_c^0, \mathbf{u}_0 + \boldsymbol{\epsilon}_c) = 0 \quad (2.4b)$$

$$h_c(\mathbf{x}_c^0, \mathbf{u}_0 + \boldsymbol{\epsilon}_c) \leq \bar{s}, \quad (2.4c)$$

reçoit la solution  $\mathbf{u}_0$  du problème maître en entrée et évalue sa faisabilité dans la configuration post-contingence  $c$ , en introduisant un vecteur d'ajustement correctif  $\boldsymbol{\epsilon}_c \in \mathbb{R}^m$ . La contrainte (2.4b) impose le respect de la contrainte d'équilibre après la contingence, tandis que (2.4c) en garantit la sécurité.

Cette architecture permet un découplage total entre les sous-problèmes, qui peuvent ainsi être résolus indépendamment, en parallèle, avec un coût computationnel contrôlé. Lorsqu'un sous-problème est infaisable pour une solution donnée  $\mathbf{u}_0$ , une coupe de faisabilité de Benders est générée, puis injectée dynamiquement dans le problème maître afin d'exclure cette solution des itérations futures :

$$\mathbf{1}^\top \boldsymbol{\epsilon}_c + \boldsymbol{\lambda}^\top (\mathbf{u}_0^* - \mathbf{u}_0) \leq 0, \quad (2.5)$$

où  $\boldsymbol{\lambda} \in \mathbb{R}^m$  désigne le vecteur des multiplicateurs de Lagrange associés aux contraintes de couplage du sous-problème de contingence  $c$ , évalués à la solution courante  $\mathbf{u}_0^*$ . Ce vecteur agit comme gradient directionnel dans l'espace des variables maîtres.

Contrairement à d'autres formulations où les sous-problèmes s'écartent de la structure canonique d'écoulement de puissance optimal (notamment en introduisant des injections fictives pour rétablir artificiellement la faisabilité), cette méthode maintient dans toutes les composantes du modèle une structure analogue à celle d'un problème de pré-répartition. Cela permet de limiter le nombre de variables supplémentaires introduites, en particulier, en évitant l'ajout de sources fictives, i.e., de variables artificielles représentant des injections de puissance active  $p_i$  au sein des noeuds  $i \in \mathcal{N}$ , utilisées pour rétablir la faisabilité locale dans les sous-problèmes, ce qui contribue à améliorer la stabilité numérique de l'algorithme de décomposition.

L'article [25] reprend aussi la structure classique de la décomposition de Benders appliquée au SCOPF, mais l'enrichit en considérant, en plus des injections des générateurs, des dispositifs de contrôle de l'écoulement de puissance réactives sur les lignes, appelés dispositifs flexibles de transmission AC distribués (*distributed flexible AC transmission systems*, DFACTS). Contrairement à [26], où seules les variables de génération sont ajustées, ici les réactances effectives des lignes de transport, notées  $X_{ij}^{\text{tot}} = X_{ij} + \Delta X_{ij}$ , où  $\Delta X$  est également traitées comme variables de décision dans le problème maître.

Cette extension modifie sensiblement la structure du SCOPF. D'une part, elle permet une action corrective post-contingence via le réglage dynamique des DFACTS, ce qui renforce la flexibilité du réseau. D'autre part, elle introduit une dépendance non linéaire entre la matrice des facteurs de distribution et les variables de contrôle  $\Delta X$ , ce qui rend le problème globalement non convexe.

Le problème maître (2.6) est formulé comme un SCED enrichi, avec pour variables les injections nodales  $p_i$  et les réglages nominaux  $\Delta X_{ij}$  des DFACTS, menant à :

$$\min_{\mathbf{p}, \Delta \mathbf{x}} \sum_{i \in \mathcal{G}} a_i p_i^2 + b_i p_i + c_i \quad (2.6a)$$

$$\text{s.c.} \quad p_k - d_k = \sum_{n \in \mathcal{N}_k} p_{kn}, \quad \forall k \in \mathcal{N} \quad (2.6b)$$

$$\Delta \underline{X}_{ij} \leq \Delta X_{ij} \leq \Delta \bar{X}_{ij}, \quad \forall (i, j) \in \mathcal{L} \quad (2.6c)$$

$$\underline{p}_i \leq p_i \leq \bar{p}_i, \quad \forall i \in \mathcal{G} \quad (2.6d)$$

$$-\bar{s}_{ij} \leq p_{ij}(\Delta X) \leq \bar{s}_{ij}, \quad \forall (i, j) \in \mathcal{L} \quad (2.6e)$$

Chaque sous-problème de contingence est formulé par la suite comme suit :

$$\min_{\mathbf{R}} \sum_{(i, j) \in \mathcal{L}} R_{ij} \quad (2.7a)$$

$$\text{s.c.} \quad \sum_{n \in \mathcal{N}_k} p_{kn}^c = p_k - d_k, \quad \forall k \in \mathcal{N} \quad (2.7b)$$

$$-\bar{s}_{ij} - R_{ij} \leq p_{ij}^c \leq \bar{s}_{ij} + R_{ij} \quad \forall (i, j) \in \mathcal{L} \quad (2.7c)$$

$$r_{ij} \geq 0, \quad \forall (i, j) \in \mathcal{L}, \quad (2.7d)$$

où :

- $p_k \in \mathbb{R}$  est la puissance active injectée au noeud  $k \in \mathcal{N}$ , dont la valeur optimale est déterminée dans le problème maître ;
- $d_k \in \mathbb{R}$  est la charge au nœud  $k$  ;
- $p_{kn}^c \in \mathbb{R}$  désigne l'écoulement de puissance active entre les noeuds  $k$  et  $n$  dans le scénario de contingence  $c \in \mathcal{C}$  ;
- $p_{ij}^c \in \mathbb{R}$  désigne l'écoulement de puissance active sur la ligne de transport  $(i, j) \in \mathcal{L}$  sous la contingence  $c$  ;
- $R_{ij} \in \mathbb{R}_+$  est une variable artificielle (*slack*) introduite pour quantifier la pénalité admissible des limites de contraintes thermiques sur la ligne  $(i, j)$ .

Le vecteur  $\mathbf{p} \in \mathbb{R}^{|\mathcal{G}|}$  est considéré comme constant dans le sous-problème, conformément à la politique de correction post-contingence adoptée. Les seules variables ajustables sont les réglages de susceptance  $\Delta X_{ij}$  (non représentés explicitement ici) et les relâches  $R_{ij}$ , dont la somme est minimisée dans l'objectif (2.7a).

Une coupe d'infaisabilité est ensuite construite sous la forme (2.8), analogue à celle de [26],

mais enrichie du rôle des multiplicateurs de Lagrange sur les contraintes de génération :

$$w^* + \boldsymbol{\lambda}^\top (\mathbf{p} - \mathbf{p}^*) \leq 0, \quad (2.8)$$

où :

- $\mathbf{p} \in \mathbb{R}^{|\mathcal{G}|}$  désigne le vecteur des puissances actives générées dans la solution courante du problème maître ;
- $\mathbf{p}^*$  est la solution de génération candidate jugée infaisable pour au moins une contingence ;
- $\boldsymbol{\lambda} \in \mathbb{R}^{|\mathcal{G}|}$  est le vecteur des multiplicateurs de Lagrange associés aux contraintes de bilan nodal (2.6b) ou d'ajustement de puissance dans le sous-problème post-contingence (2.7b) ;
- $w^* \in \mathbb{R}$  est le coût minimal (ou une mesure de l'infaisabilité) obtenu à la résolution du sous-problème associé à cette solution  $\mathbf{p}^*$ .

Les résultats expérimentaux, menés sur les réseaux tests IEEE 30, 39 et 118 nœuds, montrent que cette approche permet de réduire significativement le coût du SCOPF lorsque les DFACTS sont activés. Les auteurs observent également une accélération substantielle par rapport à la méthode brute, en particulier pour les grands réseaux, tout en ne sacrifiant que marginalement l'optimalité (écart inférieur à 0,5 % dans le pire des cas).

Dans [27], les auteurs abordent le problème de SCOPF dans un cadre opérationnel réaliste inspiré des spécifications du concours ARPA-E GO, avec des réseaux de plusieurs milliers de nœuds, et un délai strict de résolution ( $< 10$  min) pour la pré-répartition économique. Le défi porte à la fois sur la mise à l'échelle à un nombre massif de contingences et sur le respect de contraintes complexes, incluant des variables continues, discrètes, et des fonctions non différentiables liées à la réponse dynamique des générateurs.

La formulation de référence du SCOPF est posé comme suit :

$$\min_{\substack{\mathbf{x}^0, \mathbf{s}^0 \\ \{\mathbf{x}^{(c)}, \mathbf{s}^{(c)}\}_{c \in \mathcal{C}}}} c(\mathbf{x}^0) + c^0(\mathbf{s}^0) + \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} c^{(c)}(\mathbf{s}^{(c)}) \quad (2.9a)$$

$$\text{s.c.} \quad \mathbf{f}^0(\mathbf{x}^0) = \mathbf{s}_{f0}^+ - \mathbf{s}_{f0}^- \quad (2.9b)$$

$$\mathbf{g}^0(\mathbf{x}^0) \leq \mathbf{s}_{g0} \quad (2.9c)$$

$$\mathbf{f}^{(c)}(\mathbf{x}^{(c)}) = \mathbf{s}_{f^{(c)}}^+ - \mathbf{s}_{f^{(c)}}^- \quad \forall c \in \mathcal{C} \quad (2.9d)$$

$$\mathbf{g}^{(c)}(\mathbf{x}^{(c)}) \leq \mathbf{s}_{g^{(c)}} \quad \forall c \in \mathcal{C} \quad (2.9e)$$

$$\mathbf{h}^{(c)}(\mathbf{x}^{0,p,q,v}, \mathbf{x}^{(c),p,q,v,\Delta}) = \mathbf{0} \quad \forall c \in \mathcal{C} \quad (2.9f)$$

$$\mathbf{x}^{(c)} \in \mathcal{X}^{(c)}, \quad \mathbf{s}^{(c)} \in \mathcal{S}^{(c)} \quad \forall c \in \{0\} \cup \mathcal{C}, \quad (2.9g)$$

où :

- $\mathbf{x}^{(c)} \in \mathbb{R}^n$  est le vecteur des variables primaires associées au scénario  $c \in \{0\} \cup \mathcal{C}$  ;
- $\mathbf{s}^{(c)} \in \mathbb{R}^{n_s}$  est le vecteur des variables de relâchement (slack), partitionné comme

$$\mathbf{s}^{(c)} = \begin{bmatrix} \mathbf{s}_{f^{(c)}}^+ \\ \mathbf{s}_{f^{(c)}}^- \\ \mathbf{s}_{g^{(c)}} \end{bmatrix}$$

où :

- $\mathbf{s}_{f^{(c)}}^+, \mathbf{s}_{f^{(c)}}^- \in \mathbb{R}^m$  représentent respectivement les pénalités positives et négatives sur les contraintes d'équilibre nodal ;
- $\mathbf{s}_{g^{(c)}} \in \mathbb{R}^{m'}$  représente les pénalités associées aux contraintes d'inégalité, telles que les limites thermiques sur les lignes de transport  $(i, j) \in \mathcal{L}$  ou les bornes de tension nodale.
- $c : \mathbb{R}^{|\mathcal{G}|} \rightarrow \mathbb{R}$  est la fonction de coût de génération définie par morceaux, appliquée au vecteur de production ;
- $c^0 : \mathbb{R}^{n_s} \rightarrow \mathbb{R}$  est la fonction de pénalisation associée aux variables slack du cas nominal ;
- $c^{(c)} : \mathbb{R}^{n_s} \rightarrow \mathbb{R}$  est la fonction de pénalisation associée aux variables slack dans le scénario de contingence  $c \in \mathcal{C}$  ;
- $\mathbf{f}^{(c)} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  désigne les contraintes d'équilibre nodal pour le scénario  $c$  ;
- $\mathbf{g}^{(c)} : \mathbb{R}^n \rightarrow \mathbb{R}^{m'}$  regroupe contraintes d'inégalité dans le scénario  $c$ , incluant les limites thermiques  $\bar{s}_{ij}$  et les bornes de tension  $\underline{v}_i, \bar{v}_i$  ;
- $\mathbf{h}^{(c)} : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^r$  sont les contraintes de cohérence entre l'état nominal et l'état post-contingence ;
- $\mathbf{x}^{0p,q,v} \in \mathbb{R}^{n'}$  et  $\mathbf{x}^{(c),p,q,v,\Delta} \in \mathbb{R}^{n''}$  désignent les sous-vecteurs extraits de  $\mathbf{x}^0$  et  $\mathbf{x}^{(c)}$  correspondant respectivement aux injections actives  $p_i$ , réactives  $q_i$ , tensions  $v_i$  et ajustements  $\Delta p_i$ .

Cette formulation exploite des variables de pénalité pour tolérer des dépassements de contraintes, pénalisées dans la fonction objectif (2.9a). Le couplage structurel entre le régime pré-contingence et chaque configuration post-contingence est formalisé par la contrainte de réponse (2.9f), généralement non linéaire et non différentiable.

Les auteurs proposent une architecture modulaire combinant trois axes de simplification :

- une sélection de contingences fondée sur un critère de taille ou d'impact, calculé à partir d'un point d'exploitation donné ;
- trois formulations approchées du SCOPF :
  - base non linéaire avec contingences non linéaires (*nonlinear base nonlinear contingencies*)

*gency*, NBNC), i.e., ni le cas nominal ni les scénarios de contingence ne sont simplifiés ; la formulation conserve l'intégralité de la modélisation non linéaire du SCOPF ;

- base non linéaire avec contingences linéarisées (*nonlinear base linear contingency*, NBLC), i.e., la formulation conserve un modèle non linéaire pour le régime de base, mais linéarise les scénarios de contingence afin d'en accélérer la résolution ;
- base linéarisée avec contingences linéarisées (*linear base linear contingency*, LBLC), i.e., le cas nominal et les scénarios post-contingence sont tous deux traités par des approximations linéaires, ce qui permet une réduction substantielle du temps de calcul, au prix d'une moindre fidélité physique.
- l'emploi, ou non, d'une décomposition de Benders sur des variantes rendues compatibles par duplication structurée des variables.

Chacune des formulations NBNC, NBLC et LBLC est ensuite rendue décomposable via l'introduction de variables auxiliaires  $\mathbf{y}^{(c)}$ , ce qui permet de découpler les sous-problèmes associés à chaque scénario de contingence  $c \in \mathcal{C}$ . Le découplage repose sur les relations suivantes, illustrées ici dans le cas de la formulation NBNC :

$$\mathbf{x}^{0,p,v} = \mathbf{y}^{(c)} \quad \forall c \in \mathcal{C} \quad (2.10a)$$

$$\tilde{h}^{(c)}(\mathbf{y}^{(c)}, \mathbf{x}^{(c),p,v,\Delta}) = 0 \quad \forall c \in \mathcal{C}, \quad (2.10b)$$

où  $\mathbf{x}^{0,p,v}$  désigne le vecteur de la sous-composante des variables primaires du régime pré-contingence  $\mathbf{x}^0$  associées respectivement aux injections actives  $p_i \in \mathbb{R}$  et aux tensions nodales  $v_i \in \mathbb{C}$ , pour chaque nœud  $i \in \mathcal{N}$ . De même,  $\mathbf{x}^{(c),p,v,\Delta}$  regroupe les composantes de  $\mathbf{x}^{(c)}$  correspondant aux injections  $p_i$ , aux tensions  $v_i$  et aux ajustements correctifs post-contingence  $\Delta p_i$  ou  $\Delta v_i$ , modélisant la réponse admissible du système à la défaillance  $c$ . La contrainte (2.10a) impose une égalité structurelle entre le point de base et sa projection dans les sous-problèmes, tandis que (2.10b) en encode la cohérence via l'approximation DC de l'écoulement de puissance.

Ce mécanisme rend chaque sous-problème parfaitement parallèle, ce qui autorise une exploitation efficace des architectures de calcul massivement distribuées. Par ailleurs, la solution  $\mathbf{x}^{0,(\ell)}$  produite à chaque itération  $\ell = 1, 2, \dots$  du problème maître reste toujours utilisable comme solution de repli, ce qui sécurise les arrêts prématurés dans un cadre temps-réel.

Les tests expérimentaux sur des réseaux ARPA-E de 793 à plus de 2,300 nœuds montrent que :

- les méthodes approximées avec sélection de contingences permettent une réduction de plus de 90% du score par rapport à la solution triviale ;
- l'utilisation combinée de la formulation NBLC et de la décomposition de Benders offre

- les meilleurs compromis entre performance et faisabilité ;
- la solution est générée en moins de 10 minutes, y compris pour des réseaux avec près de 1000 contingences.

Cette approche démontre ainsi la faisabilité d'un SCOPF à très grande échelle, combinant approximations structurelles, réduction de contingences, et décomposition distribuée dans un cadre temps-réel.

Une variante de la décomposition de Benders, introduite dans [32], repose sur une séparation zonale du SCED, avec coordination explicite entre zones par mise à jour des variables dites marginales. Chaque sous-problème local est résolu indépendamment, puis des contraintes de couplage sont imposées itérativement via un algorithme de mise à jour des ensembles de contraintes actives et de variables saturées. Cette méthode garantit la convergence vers une solution optimale sans relaxation, mais reste équivalente, sur le plan structurel, à une décomposition de Benders appliquée à un problème linéaire classique.

Dans [33], les auteurs proposent une méthode de résolution exacte et pouvant être mise à l'échelle du problème de SCOPF avec modélisation explicite de la réponse primaire automatique des générateurs synchronisés, i.e., une variation automatique de la puissance active injectée proportionnelle à un signal global post-contingence dès lors que le générateur n'a pas atteint sa limite supérieure de production. Contrairement aux approches traditionnelles basées sur la décomposition de Benders, leur méthode évite la résolution explicite des sous-problèmes en exploitant deux idées centrales : (i) l'utilisation de structures prétraitées à partir de la formulation PTDF pour détecter les dépassements de limites post-contingence et (ii) la génération itérative de contraintes et de disjonctions activées uniquement en cas de besoin.

Le problème d'origine du SCOPF est un programme linéaire mixte du fait de la nature des contraintes disjonctives de réponse primaire (chaque générateur peut soit suivre le signal global de réponse primaire, soit saturer à sa puissance maximale). La formulation complète est rappelée ci-dessous :

$$\min_{\mathcal{X}, \{\mathcal{X}^{(c)}, \mathbf{x}^{(c)}, n^{(c)}\}_{c \in \mathcal{C}}} \quad \mathbf{c}^\top \mathbf{g} \quad (2.11a)$$

$$\text{s.c.} \quad \mathcal{X} \in \mathcal{E} \quad (2.11b)$$

$$\mathcal{X}^{(c)} \in \mathcal{E}^{(c)}, \quad \forall c \in \mathcal{C} \quad (2.11c)$$

$$\mathcal{Y}^{(c)} \in \mathcal{F}^{(c)}, \quad \forall c \in \mathcal{C}, \quad (2.11d)$$

où  $\mathcal{X}$  regroupe les variables d'état nominales  $(\mathbf{g}, \mathbf{f}, \boldsymbol{\theta})$ ,  $\mathcal{X}^{(c)}$  celles des états post-contingence, et  $\mathcal{Y}^{(c)}$  encode les variables binaires  $\mathbf{x}^{(c)}$  et les signaux  $n^{(c)}$  associés à la réponse primaire.

Afin d'éviter l'explosion combinatoire due aux binaires, les auteurs introduisent une stratégie de génération différée des contraintes disjonctives, via un algorithme de génération de colonnes et de contraintes (*column-and-constraint generation algorithm*, CCGA). Le problème maître initial ne contient ni les contraintes de réseau post-contingence, ni les contraintes disjonctives. À chaque itération :

1. Le programme maître est résolu sur un sous-ensemble actif  $\mathcal{C}' \subset \mathcal{C}$  de contingences critiques ;
2. Une procédure de recherche binaire est utilisée pour calculer, pour chaque  $c \in \mathcal{C}$ , la réponse  $g_{c,i}$  induite par le signal global  $n^{(c)}$  ;
3. Des structures linéaires prétraitées, dérivées des facteurs PTDF et notées  $K_1, K_3 \in \mathbb{R}^{|\mathcal{L}| \times |\mathcal{G}|}$  et  $k_2, k_4 \in \mathbb{R}^{|\mathcal{L}|}$ , permettent d'évaluer directement si une limite thermique est dépassée sous la contingence  $c$ , en testant une inégalité du type  $K_1\mathbf{g} + k_2 \leq \bar{s}_{ij}$ , sans résoudre explicitement de sous-problème ;
4. Les contraintes de sécurité, ainsi que les contraintes disjonctives de réponse primaire, sont alors injectées dynamiquement dans le programme maître, et le processus est itérativement répété.

Cette approche permet de contourner les limitations classiques des sous-problèmes non convexes dans les formulations préventives avec réponse primaire. Les auteurs démontrent également la possibilité d'estimer efficacement un gap d'optimalité certifié en combinant deux processus d'exécution parallèles : l'un explore une suite de modèles approximés (en restreignant le nombre de générateurs susceptibles de répondre), l'autre suit la solution complète.

Selon les résultats numériques la méthode CCGA résout des instances jusqu'à 6468 noeuds et plus de 9 millions de contraintes en moins de 3 heures, alors que les approches de type Benders ou formulation complète échouent sur la plupart des cas (temps limite de 4 heures dépassé). Pour le cas PEGASE-1354, une solution quasi-optimale est obtenue en moins de 10 secondes dès le niveau de priorité  $p = 0$ , avec un écart relatif inférieur à 1,5 % par rapport au coût objectif du SCOPF complet, validant ainsi la qualité de l'approximation hiérarchique proposée.

Dans [34], les auteurs proposent une approche généralisée pour le calcul des coûts marginaux de congestion dans un SCED, tenant compte de l'effet explicite d'un grand nombre de contingences dans un cadre temps réel ou intra-journalier, c'est-à-dire, couvrant les ajustements opératoires effectués après la planification DA mais avant le temps réel, typiquement sur des pas de 5 à 60 minutes [7]. La contribution principale réside dans l'introduction d'une structure factorisée des matrices PTDF post-contingences, qui permet une mise à l'échelle efficace même dans les cas où des milliers de lignes doivent être considérées comme contingentes.

Le problème de base est formulé sous forme de DCOPF, avec un modèle multipériode. Pour chaque période  $t$ , notons  $\mathbf{p}^t \in \mathbb{R}^{|\mathcal{G}|}$  le vecteur des productions, et  $\mathbf{d}^t \in \mathbb{R}^{|\mathcal{N}|}$  la charge. Les contraintes d'équilibre de puissance s'écrivent :

$$\mathbf{B}^{\text{bus}}\boldsymbol{\theta}^t + \mathbf{d}^t - \mathbf{A}\mathbf{p}^t = \mathbf{0}, \quad \forall t \quad (2.12)$$

où  $\mathbf{A} \in \mathbb{R}^{|\mathcal{N}| \times |\mathcal{G}|}$  désigne la matrice d'incidence reliant les générateurs  $g \in \mathcal{G}$  aux noeuds  $i \in \mathcal{N}$ , c'est-à-dire  $A_{ig} = 1$  si le générateur  $g$  est connecté au noeud  $i$ , et 0 sinon.

Les contraintes sur les écoulements de puissance sont formulées sous forme linéaire à l'aide de la matrice PTDF,  $\Phi$ , étendue aux matrices PTDF post-contingences,  $\Phi^{(c)}$ . Soit  $\mathcal{C}$  l'ensemble des contingences. Les contraintes post-contingence sont données par :

$$|\Phi^{(c)}\mathbf{A}\mathbf{p}^t - \Phi^{(c)}\mathbf{d}^t| \leq \bar{s}, \quad \forall c \in \mathcal{C}, \quad \forall t \quad (2.13)$$

La matrice PTDF post-contingence  $\Phi^{(c)}$ , associée au retrait d'une ligne contingente  $c \in \mathcal{C}$ , est exprimée sous une forme factorisée :

$$\Phi^{(c)} = \Phi^{(0)} + \Lambda_c \Phi_{\mathcal{L}_c}^{(0)}, \quad (2.14)$$

où, conformément à (1.16) :

- $\Phi^{(0)} \in \mathbb{R}^{|\mathcal{L}| \times |\mathcal{N}|}$  désigne la matrice  $\Phi$  du cas nominal (sans contingence) ;
- $\Lambda_c \in \mathbb{R}^{|\mathcal{L}| \times 1}$  est la colonne de la matrice  $\Lambda$  associée à la contingence  $c$  ;
- $\Phi_{\mathcal{L}_c}^{(0)} \in \mathbb{R}^{1 \times |\mathcal{N}|}$  correspond à la ligne de la matrice  $\Phi$  nominale associée à la ligne en contingence dans le scénario  $c$ .

Cette expression permet un calcul incrémental de l'impact de chaque contingence sur les écoulements de puissance.

Les auteurs exploitent cette factorisation pour dérivées les LMP associées à chaque contrainte de congestion de manière efficace. Ils montrent que les LMP peuvent être évaluées pour chaque ligne contingente à un coût quasi constant par contingence, sans avoir à reformuler un OPF complet.

L'algorithme global procède ainsi :

1. Résolution du SCED de base avec contraintes nominales (pas de contingence) ;
2. Calcul de tous les non-respects de limites d'écoulement de puissance post-contingence via (2.14), sans résolution d'un second problème ;
3. Génération incrémentale des contraintes réellement non-respectées parmi les  $|\mathcal{C}||\mathcal{L}|$  possibles ;

#### 4. Résolution du problème augmenté, uniquement avec les contraintes activées.

Les résultats numériques de [34] montrent que cette méthode permet de résoudre un SCED avec plus de 2000 contingences en moins de 10 secondes sur un système de taille moyenne (IEEE 300 bus), avec un écart de solution inférieur à 0.1% par rapport à une résolution exhaustive. Cette approche constitue une alternative exacte, non heuristique, aux méthodes de décomposition par sous-problèmes, et se prête naturellement à une intégration dans des simulateurs temps réel.

Finalement, l'article [35] introduit une méthode de résolution du SCOPF reposant sur une combinaison originale entre la décomposition de Benders et le lemme de modification de matrice inverse [36] (*inverse matrix modification lemma*, IMML), issu de l'identité de Sherman–Morrison–Woodbury. L'objectif est de réduire le coût de calcul des écoulements de puissance post-contingence, sans recours à des solveurs externes, en s'appuyant sur des mises à jour analytiques de la matrice inverse de susceptance.

Le SCOPF est formulé avec une distinction explicite entre :

- un état pré-contingence noté  $(\mathbf{x}_0, \mathbf{u}_0)$ ;
- un état post-contingence immédiat noté  $(\mathbf{x}_c, \mathbf{u}_0)$ , obtenu juste après l'ouverture automatique du disjoncteur isolant l'équipement défaillant (*tripping*) ;
- un état post-contingence corrigé par l'action corrective  $\mathbf{u}_c$ .

Le modèle de SCOPF proposé dans [35] repose sur une distinction explicite entre les états pré- et post-contingence, ainsi que sur une modélisation à deux échelles temporelles : une phase immédiate suivant le défaut (*short-term*, ST), et une phase de correction stabilisée (*long-term*, LT). Le problème est formulé comme suit :

$$\min_{\mathbf{x}_0, \mathbf{u}_0, \mathbf{x}_c, \mathbf{u}_c} f(\mathbf{x}_0, \mathbf{u}_0, \mathbf{x}_c, \mathbf{u}_c) \quad (2.15a)$$

$$\text{s.c.} \quad g_0(\mathbf{x}_0, \mathbf{u}_0) = 0 \quad (2.15b)$$

$$h_0(\mathbf{x}_0, \mathbf{u}_0) \leq h_0^{\text{LT}} \quad (2.15c)$$

$$g_c(\mathbf{x}_c, \mathbf{u}_c) = 0, \quad \forall c \in \mathcal{C} \quad (2.15d)$$

$$h_c(\mathbf{x}_c, \mathbf{u}_0) \leq h_c^{\text{ST}}, \quad \forall c \in \mathcal{C} \quad (2.15e)$$

$$h_c(\mathbf{x}_c, \mathbf{u}_c) \leq h_c^{\text{LT}}, \quad \forall c \in \mathcal{C} \quad (2.15f)$$

$$|\mathbf{u}_c - \mathbf{u}_0| \leq \Delta \mathbf{u}_c, \quad \forall c \in \mathcal{C}, \quad (2.15g)$$

où :

- $\mathbf{x}_0 \in \mathbb{R}^n$  est le vecteur des variables d'état en régime pré-contingence ;
- $\mathbf{u}_0 \in \mathbb{R}^m$  représente les variables de contrôle préventif ;

- $\mathbf{x}_c \in \mathbb{R}^n$  est le vecteur des variables d'état post-contingence ;
- $\mathbf{u}_c \in \mathbb{R}^m$  désigne les actions correctives à appliquer après la contingence  $c$  ;
- $f : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$  est la fonction objectif, incluant typiquement le coût de génération et éventuellement des pénalités sur les ajustements post-contingence ;
- $g_0 : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^{|\mathcal{N}|}$ ,  $g_c : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^{|\mathcal{N}|}$  désignent les contraintes d'équilibre nodal en pré- et post-contingence ;
- $h_0 : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^r$ ,  $h_c : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^r$  regroupent les contraintes d'inégalité, incluant les limites thermiques et les bornes opérationnelles ;
- les bornes  $\mathbf{h}_0^{\text{LT}}, \mathbf{h}_c^{\text{ST}}, \mathbf{h}_c^{\text{LT}} \in \mathbb{R}^r$  définissent les seuils admissibles respectivement en régime long terme et court terme ;
- la contrainte (2.15g) impose une borne  $\Delta\mathbf{u}_c \in \mathbb{R}^m$  sur la variation admissible entre contrôle préventif et correctif.

L'originalité de la méthode proposée réside dans l'évaluation analytique des écoulements de puissance post-contingence, sans qu'il soit nécessaire de reconstruire ou de résoudre un sous-problème. Pour cela, les auteurs représentent l'état du réseau sous contingence  $c \in \mathcal{C}$  par les équations d'équilibre nodal associées au modèle linéarisé :

$$(\mathbf{B}_{\text{bus}} + \Delta\mathbf{B}_{\text{bus}}^{(c)}) \boldsymbol{\theta}^{(c)} = \mathbf{p}, \quad (2.16)$$

où  $\mathbf{B}_{\text{bus}} \in \mathbb{R}^{|\mathcal{N}| \times |\mathcal{N}|}$  désigne la matrice de susceptance nodale du réseau dans le cas nominal, et  $\Delta\mathbf{B}_{\text{bus}}^{(c)}$  est la mise à jour induite par le retrait de la ligne  $(i, j)$  dans le scénario de contingence  $c \in \mathcal{C}$ . Cette perturbation structurée s'écrit :

$$\Delta\mathbf{B}_{\text{bus}}^{(c)} = -b_{ij} (\mathbf{e}_i - \mathbf{e}_j) (\mathbf{e}_i - \mathbf{e}_j)^\top, \quad (2.17)$$

où  $b_{ij}$  est la susceptance de la ligne sous contingence, et  $\mathbf{e}_n \in \mathbb{R}^{|\mathcal{N}|}$  est le vecteur canonique unitaire associé au nœud  $n$ , i.e., dont les entrées sont nulles sauf pour la  $i^{\text{ième}}$  qui est 1.

En utilisant le lemme de Sherman–Morrison, la pseudo-inverse réduite de  $\mathbf{B}_{\text{bus}} + \Delta\mathbf{B}_{\text{bus}}^{(c)}$  peut être obtenue analytiquement par :

$$\hat{\mathbf{B}}_{\text{bus}}^{(c)} = \hat{\mathbf{B}}_{\text{bus}} - \frac{\hat{\mathbf{B}}_{\text{bus}} \mathbf{v} \mathbf{v}^\top \hat{\mathbf{B}}_{\text{bus}}}{1/b_{ij} + \mathbf{v}^\top \hat{\mathbf{B}}_{\text{bus}} \mathbf{v}} \quad (\text{S-M})$$

où,  $\mathbf{v} = \mathbf{e}_i - \mathbf{e}_j$ . Cette mise à jour permet de reconstruire les angles de phase sous contingence par  $\boldsymbol{\theta}^{(c)} = \hat{\mathbf{B}}_{\text{bus}}^{(c)} \mathbf{p}$ , sans résolution directe du système modifié.

La matrice PTDF post-contingence  $\Phi^{(c)}$  est alors obtenue par projection :

$$\Phi^{(c)} = \mathbf{B}_{\text{line}} \hat{\mathbf{B}}_{\text{bus}}^{(c)}, \quad (2.18)$$

où  $\mathbf{B}_{\text{line}}$  est la matrice d'incidence orientée des lignes pondérée par les susceptances, utilisée pour calculer les écoulements de puissance sur les lignes :

$$\mathbf{p}_{ij}^{(c)} = \Phi^{(c)} \mathbf{p}.$$

Enfin, les coupes de Benders sont générées dynamiquement à partir des écoulements post-contingence calculés, sans résolution explicite de sous-problèmes. Lorsqu'une limite de contrainte thermique est non-respectée sur une ligne  $(i, j) \in \mathcal{L}$ , une inégalité du type :

$$\sum_{k \in \mathcal{N}} \Phi_{ij,k}^{(c)} \Delta p_k \geq p_{ij}^{(c)} - \bar{s}_{ij} \quad (2.19)$$

est ajoutée au problème maître, où  $\Delta p_k$  désigne la correction de production au nœud  $k$ , et  $\bar{s}_{ij}$  la limite thermique de la ligne.

L'algorithme procède de manière itérative en parcourant les contingences  $c \in \mathcal{C}$ , et en injectant, à chaque étape, les coupes nécessaires dans le programme maître pour garantir la faisabilité post-contingence. Le plan de génération préventif  $\mathbf{u}_0$  est mis à jour en conséquence. Lorsqu'une contingence entraîne une séparation topologique du réseau électrique (îlotage), le modèle considère des hypothèses conservatrices dans lesquelles une coupure totale de la charge est appliquée sur les îlots non alimentés, modélisant ainsi une coupure généralisée d'alimentation électrique (*black-out*) partiel contrôlé.

Les résultats empiriques indiquent une réduction du temps de calcul de plus de 80 % sur le système ACTIVSg500 [37], pour un coût optimal identique à cinq décimales. La structure permet une mise en œuvre modulaire et parallèle.

### 2.1.2 Méthodes de décomposition par multiplicateurs alternés

Le schéma des méthodes de décomposition par multiplicateurs alternés (*alternating direction method of multipliers*, ADMM) s'inscrit dans la famille des Lagrangiens augmentés, initialement conçues pour renforcer la stabilité et la convergence des approches duales classiques, même en l'absence de convexité stricte [28]. Considérons le problème de minimisation générique suivant :

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad (2.20a)$$

$$\text{s.c. } \mathbf{Ax} = \mathbf{b}, \quad (2.20\text{b})$$

où  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  est une fonction convexe propre, et  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $m, n \in \mathbb{N}$ . La méthode du Lagrangien augmenté associe à (2.20) la fonction suivante :

$$\mathcal{L}_\rho(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) + \mathbf{y}^\top (\mathbf{Ax} - \mathbf{b}) + \frac{\rho}{2} \|\mathbf{Ax} - \mathbf{b}\|_2^2, \quad (2.21)$$

où  $\mathbf{y} \in \mathbb{R}^m$  est le vecteur des multiplicateurs de Lagrange, et  $\rho > 0$  un paramètre de pénalité. Cependant, l'ajout du terme quadratique détruit en général la séparabilité du problème. Pour restaurer une structure exploitable, ADMM introduit une séparation explicite des variables. Le problème est reformulé sous la forme :

$$\min_{\mathbf{x}, \mathbf{z}} \quad f(\mathbf{x}) + g(\mathbf{z}) \quad (2.22\text{a})$$

$$\text{s.c. } \mathbf{Ax} + \mathbf{Bz} = \mathbf{c}, \quad (2.22\text{b})$$

où  $f$  et  $g$  sont convexes, et  $\mathbf{A}, \mathbf{B}, \mathbf{c}$  sont des données du problème. L'ADMM procède alors par minimisation alternée de la version augmentée du Lagrangien :

$$\mathcal{L}_\rho(\mathbf{x}, \mathbf{z}, \mathbf{y}) = f(\mathbf{x}) + g(\mathbf{z}) + \mathbf{y}^\top (\mathbf{Ax} + \mathbf{Bz} - \mathbf{c}) + \frac{\rho}{2} \|\mathbf{Ax} + \mathbf{Bz} - \mathbf{c}\|_2^2, \quad (2.23)$$

avec des mises à jour séquentielles :

$$\mathbf{x}^{k+1} := \arg \min_{\mathbf{x}} \mathcal{L}_\rho(\mathbf{x}, \mathbf{z}^k, \mathbf{y}^k) \quad (2.24\text{a})$$

$$\mathbf{z}^{k+1} := \arg \min_{\mathbf{z}} \mathcal{L}_\rho(\mathbf{x}^{k+1}, \mathbf{z}, \mathbf{y}^k) \quad (2.24\text{b})$$

$$\mathbf{y}^{k+1} := \mathbf{y}^k + \rho(\mathbf{Ax}^{k+1} + \mathbf{Bz}^{k+1} - \mathbf{c}). \quad (2.24\text{c})$$

Le vecteur  $\mathbf{r}^k = \mathbf{Ax}^k + \mathbf{Bz}^k - \mathbf{c}$  est appelé résidu primal, mesure le non-respect de la contrainte. Le résidu dual quant à lui, noté  $\mathbf{s}^{k+1} = \rho \mathbf{A}^\top \mathbf{B}(\mathbf{z}^{k+1} - \mathbf{z}^k)$ , rend compte de la variation itérative des décisions. L'algorithme est typiquement arrêté lorsque ces deux quantités sont inférieures à un seuil prédéfini.

Dans la pratique, le paramètre  $\rho$  joue un rôle fondamental dans la vitesse de convergence. Des stratégies d'adaptation dynamique sont souvent mises en œuvre pour équilibrer les résidus

primal et dual [28]. Par exemple, une règle empirique consiste à ajuster  $\rho$  selon :

$$\rho^{k+1} := \begin{cases} \tau_{\text{incr}} \rho^k & \text{si } \|\mathbf{r}^k\|_2 > \mu \|\mathbf{s}^k\|_2, \\ \rho^k / \tau_{\text{decr}} & \text{si } \|\mathbf{s}^k\|_2 > \mu \|\mathbf{r}^k\|_2, \\ \rho^k & \text{sinon,} \end{cases}$$

où  $\tau_{\text{incr}}, \tau_{\text{decr}} > 1$  et  $\mu > 1$  sont des paramètres de réglage. Cette régularisation dynamique améliore considérablement la stabilité numérique et la fiabilité des itérations, en particulier lorsque le conditionnement des contraintes est variable.

L'efficacité du schéma ADMM pour les problèmes d'optimisation distribuée en réseau s'explique par sa capacité à exploiter la structure décomposable des contraintes et sa compatibilité avec les architectures de calcul parallèle. Ces propriétés en font une solution pertinente pour le traitement de problèmes tels que le DCOPF ou le SCOPF, notamment en contexte multi-régional, multi-contingence ou temps réel.

Dans le cas du DCOPF, le problème peut être reformulé en partitionnant les variables de décision selon les zones géographiques, les horizons temporels ou les contingences, et en imposant un couplage entre zones par des contraintes de consensus, i.e., des égalités imposant que certaines variables partagées (par exemple les injections ou les écoulements sur les lignes interzones) prennent la même valeur dans chaque sous-problème local. ADMM permet alors de résoudre chaque sous-problème localement, tout en assurant la coordination globale via la mise à jour des multiplicateurs de Lagrange. Ce principe a été utilisé avec succès dans des configurations de marché décentralisé ou de commande distribué [38–40].

Dans le cadre plus exigeant du SCOPF, ADMM est utilisé soit pour distribuer les scénarios de contingences sur plusieurs nœuds de calcul, soit pour répartir spatialement les zones de contrôle (*region-wise*). Le couplage s'effectue alors à travers les variables de contrôle préventif ou correctif communes à toutes les zones [29, 30]. Ce type de formulation permet d'éviter la résolution centralisée d'un problème de grande taille, au profit de sous-problèmes parallèles de taille réduite, avec convergence assurée sous convexité des composantes.

La revue de littérature proposée dans [41] met en évidence que la majorité des applications ADMM au SCOPF repose sur des formulations linéarisées, en raison de la difficulté à garantir la convergence sur des modèles de ACOPF non convexes. Toutefois, des extensions vers les formulations quadratiques ou semi-définies positives commencent à émerger, notamment en combinaison avec des techniques de relaxation convexes.

Dans [30], les auteurs proposent une méthode fondée sur le schéma primal-dual d'optimisation ADMM. Cette méthode est conçue pour résoudre des SCOPF de très grande taille, incluant

plusieurs centaines de contingences, dans un contexte computationnel distribué.

Le problème initial est formulé comme suit :

$$\min_{\boldsymbol{\theta}_c, \mathbf{p}_c} \sum_{i \in \mathcal{G}} f_i(p_{i,0}) \quad (2.25a)$$

$$\text{s.c. } \mathbf{B}_0^{\text{bus}} \boldsymbol{\theta}_0 + \mathbf{P}_{d,0} - \mathbf{A}_0 \mathbf{p}_0 = \mathbf{0} \quad (2.25b)$$

$$\mathbf{B}_c^{\text{bus}} \boldsymbol{\theta}_c + \mathbf{P}_{d,c} - \mathbf{A}_c \mathbf{p}_c = \mathbf{0}, \quad \forall c \in \mathcal{C} \quad (2.25c)$$

$$|\mathbf{B}_0^f \boldsymbol{\theta}_0| \leq \bar{s} \quad (2.25d)$$

$$|\mathbf{B}_c^f \boldsymbol{\theta}_c| \leq \bar{s}, \quad \forall c \in \mathcal{C} \quad (2.25e)$$

$$\underline{\mathbf{p}} \leq \mathbf{p}_0 \leq \bar{\mathbf{p}} \quad (2.25f)$$

$$\underline{\mathbf{p}} \leq \mathbf{p}_c \leq \bar{\mathbf{p}}, \quad \forall c \in \mathcal{C} \quad (2.25g)$$

$$|\mathbf{p}_0 - \mathbf{p}_c| \leq \Delta_c, \quad \forall c \in \mathcal{C}, \quad (2.25h)$$

où :

- $\boldsymbol{\theta}_c \in \mathbb{R}^{|\mathcal{N}|}$  désigne le vecteur des angles de phase au nœud dans le scénario  $c \in \{0\} \cup \mathcal{C}$ ;
- $\mathbf{p}_c \in \mathbb{R}^{|\mathcal{G}|}$  désigne la puissance active injectée par chaque générateur dans le scénario  $c \in \{0\} \cup \mathcal{C}$ ;
- $\mathbf{P}_{d,c} \in \mathbb{R}^{|\mathcal{N}|}$  représente la charge active nodale dans le scénario  $c$ ;
- $\mathbf{B}_c^{\text{bus}} \in \mathbb{R}^{|\mathcal{N}| \times |\mathcal{N}|}$  est la matrice de susceptance nodale associée au scénario  $c$ ;
- $\mathbf{B}_c^f \in \mathbb{R}^{|\mathcal{L}| \times |\mathcal{N}|}$  est la matrice de susceptance par ligne pour le scénario  $c$ ;
- $\mathbf{A}_c \in \mathbb{R}^{|\mathcal{N}| \times |\mathcal{G}|}$  est la matrice d'incidence reliant les générateurs aux nœuds dans le scénario  $c$ ;
- $\bar{s} \in \mathbb{R}^{|\mathcal{L}|}$  regroupe les limites thermiques  $\bar{s}_{ij}$  des lignes  $(i, j) \in \mathcal{L}$ ;
- $\underline{\mathbf{p}}, \bar{\mathbf{p}} \in \mathbb{R}^{|\mathcal{G}|}$  désignent respectivement les bornes inférieure et supérieure sur la puissance active injectée;
- $\Delta_c \in \mathbb{R}^{|\mathcal{G}|}$  encode les marges admissibles de réajustement entre les puissances préventives  $\mathbf{p}_0$  et correctives  $\mathbf{p}_c$ .

Cette formulation repose sur un modèle linéarisé de l'écoulement de puissance avec variables de contrôle. Pour rendre ce problème compatible avec ADMM, les auteurs introduisent un vecteur auxiliaire  $\mathbf{p}_c^{\text{aux}}$  et une reformulation du couplage par relâchement (*slack*) :

$$\mathbf{p}_0 - \mathbf{p}_c + \mathbf{r}_c = \Delta_c \quad (2.26a)$$

$$\mathbf{0} \leq \mathbf{r}_c \leq 2\Delta_c. \quad (2.26b)$$

Chaque itération d'ADMM met à jour les sous-ensembles de variables en deux blocs :

- le problème de base (pré-contingence) minimise (2.25a) avec régularisation quadratique imposée par les contraintes (2.26a) ;
- chaque scénario post-contingence résout un sous-problème de plus petite taille, de façon localisée (sans coordination directe avec les autres scénarios), et de manière parallèle sur des unités de calcul distinctes (2.25c)–(2.26b).

La coordination s'effectue par les variables duales  $\boldsymbol{\mu}_c$ , interprétées comme des prix d'ajustement. À chaque itération, la mise à jour suit le schéma :

$$\boldsymbol{\mu}_c^{(k+1)} = \boldsymbol{\mu}_c^{(k)} + \mathbf{p}_0^{(k+1)} - \mathbf{p}_c^{(k+1)} + \mathbf{r}_c^{(k+1)} - \boldsymbol{\Delta}_c.$$

Les tests numériques montrent une accélération significative pour les grands réseaux. Sur le cas IEEE 118 nœuds avec 100 contingences, le temps de calcul est réduit de moitié par rapport à une approche centralisée (36 s contre 18 s), avec un facteur de vitesse pouvant atteindre 4.8 pour des solutions approchées à 1% d'erreur relative. De plus, la méthode converge en moins de 50 itérations sur tous les cas testés (IEEE 57, 118, 300 bus).

L'approche développée par [29] repose elle aussi sur une décomposition du SCOPF fondée sur la méthode ADMM. Toutefois, contrairement à [30], cette méthode n'exploite pas de structure analytique particulière sur les matrices de susceptance ou les opérateurs de sensibilité. Le problème est reformulé de manière classique en introduisant des variables auxiliaires propres à chaque contingence, permettant une séparation directe des sous-problèmes. Les mises à jour alternées des variables de contrôle et des multiplicateurs de Lagrange suivent le schéma standard de l'ADMM, sans adaptation spécifique au contexte électrique.

En parallèle, les auteurs combinent ce mécanisme avec une décomposition de Benders, dans laquelle la génération des coupes de faisabilité repose sur une linéarisation locale du sous-problème post-contingence. Cette approche hybride améliore la qualité des coupes dans des contextes non convexes, mais ne propose pas de mécanisme explicite de reconstruction des matrices PTDF post-contingences. En ce sens, l'algorithme de [29] reste plus général mais moins spécialisé structurellement que celui proposé par [30], notamment en ce qui concerne l'efficacité du calcul distribué et la précision des conditions de sécurité.

D'autres développements récents, proposent des stratégies de résolutions distribuées du DCOPF exploitant la structure topologique du réseau pour construire des sous-problèmes pouvant être résolus localement. Dans [42], l'algorithme repose sur une coordination via ADMM entre régions adjacentes, assurée par des mises à jour duales sur les variables d'interface. Cette approche permet une convergence rapide sans compromettre la décentralisation computationnelle.

Dans [43], les auteurs s'intéressent au SCOPF multipériode post-contingence avec charge sujette à de l'incertitude. La méthode consiste à encadrer l'ensemble des réalisations possibles par deux problèmes déterministes extrêmes, dont la résolution parallèle par ADMM fournit des bornes sûres sur les décisions optimales. La structure naturelle du SCOPF, fondée sur une séparation entre état nominal et états post-contingence, est exploitée pour construire une décomposition à grande échelle, particulièrement efficace sur des réseaux standards (IEEE 30, 57 et 118 nœuds).

Ces contributions illustrent la capacité d'ADMM à traiter efficacement des formulations du DCOPF et du SCOPF à forte dimensionnalité, y compris en présence d'incertitudes, en s'appuyant sur des décompositions naturelles et des mécanismes de coordination à faible coût.

### 2.1.3 Autres méthodes de parallélisation

Au-delà des approches fondées sur la décomposition de Benders ou la méthode ADMM, plusieurs contributions ont exploré des stratégies de parallélisation distinctes pour résoudre des instances de grande dimension du problème de SCED ou du SCOPF. Ces méthodes s'appuient sur des structures algorithmiques ou physiques particulières du réseau électrique, et mobilisent des techniques adaptées à une architecture de calculs parallèle.

Certaines approches visent à paralléliser directement des méthodes classiques d'optimisation continue. C'est le cas de [44], où les auteurs développent une version massivement parallèle de la méthode de points intérieurs primale-duale, appliquée au DCOPF. L'accent est mis sur la résolution distribuée des systèmes linéaires creux issus du calcul de la direction de descente de Newton, ce qui permet de maintenir une convergence rapide tout en exploitant les ressources d'un environnement multi-cœurs. Cette stratégie est particulièrement efficace pour les réseaux de grande taille, où le solveur devient un goulot d'étranglement.

D'autres travaux exploitent la structure géographique du réseau pour proposer des décompositions coordonnées. Dans [45], SCED est formulé de manière conjointe sur les réseaux de transport et de distribution. Une stratégie de décomposition hétérogène y est proposée, permettant aux opérateurs de transport et de distribution d'optimiser leurs décisions localement tout en assurant la cohérence des variables aux nœuds de couplage. Cette approche repose sur une coordination explicite des prix nodaux et des injections aux interfaces, et préfigure les interactions à venir entre opérateurs dans un contexte de décentralisation croissante.

Une troisième catégorie de méthodes tire parti de la topologie du réseau pour structurer la parallélisation. Dans [46], les auteurs s'appuient sur une décomposition zonale guidée par les écoulements physiques pour améliorer l'agrégation des zones de marché, en distinguant les

écoulements intra-zones et inter-zones. Cette démarche permet une optimisation conjointe de la localisation des frontières de zone et des signaux de congestion. Par ailleurs, [47] propose une méthode de décomposition rapide pour le SCOPF, fondée sur l'activation sélective de contraintes. Les non-respects de limite potentiels sont identifiés via des règles de gestion des contraintes, ce qui permet d'adapter dynamiquement les bornes du problème de base en intégrant de manière incrémentale les effets des contingences, sans devoir expliciter chaque scénario dans la formulation initiale. Cette stratégie itérative, compatible avec les architectures parallèles, permet de traiter des réseaux de plus de 10,000 nœuds tout en maintenant des temps de calcul compatibles avec une application temps réel.

Ces contributions illustrent la diversité des stratégies de parallélisation envisageables pour les problèmes de type SCED ou SCOPF. En s'affranchissant des cadres classiques de décomposition duale, elles ouvrent la voie à d'autres méthodes adaptées à la complexité structurelle des réseaux modernes et aux exigences de calcul distribuées.

## 2.2 Méthodes d'apprentissage supervisé pour la prédiction de contraintes actives

Les techniques de décomposition présentées précédemment, exploitent des structures spécifiques du problème pour distribuer la charge de calcul sur plusieurs sous-problèmes partiellement découpés. Cependant, leur performance computationnelle peut se dégrader significativement lorsque le nombre de scénarios de contingence est élevé ou lorsque le réseau électrique est fortement congestionné [48]. Dans ces situations, ces méthodes requièrent souvent un grand nombre d'itérations pour atteindre la convergence, ce qui limite leur applicabilité dans un contexte opérationnel contraint par le temps. Dans ce contexte, l'essor récent des techniques d'apprentissage automatique offre une alternative prometteuse pour transférer la complexité vers une phase d'entraînement hors ligne, en capitalisant sur les redondances structurelles présentes dans les solutions optimales des problèmes SCED et SCOPF.

Une synthèse approfondie de la littérature [49] fait apparaître deux paradigmes distincts mais complémentaires : la prédiction des variables optimales d'une part, et la prédiction des contraintes actives d'autre part. Ces deux familles de méthodes visent à circonscrire, par apprentissage supervisé, l'espace de recherche dans lequel le solveur doit opérer, en exploitant les corrélations entre les conditions d'opération du réseau électrique et la structure des solutions optimales.

**Prédiction des variables optimales.** Le premier axe, qualifié d'apprentissage bout en bout (*end-to-end learning*, E2E) [49], consiste à entraîner un modèle de régression supervisée qui, à partir d'un vecteur de conditions d'entrée, comme le profil de charge, prédit directement

certaines variables de décision du SCOPF, i.e., la puissance générée, les tensions ou les angles de phase. Dans cette approche, illustrée notamment par DeepOPF [50] ou par les architectures SDAE [51], le modèle apprend une approximation fonctionnelle de la solution optimale. Si les performances de prédiction sont souvent élevées (des taux d'erreur relatifs inférieurs à 1 % sont rapportés dans certains cas [48]), la faisabilité des solutions reste généralement non garantie, ce qui impose des procédures de post-traitement ou de projection pour rétablir le respect des contraintes physiques.

**Prédiction des contraintes actives.** Le second axe, dit d'apprendre-pour-optimiser (*learning-to-optimize*, L2O) [49], repose sur une classification supervisée du statut des contraintes d'inégalité. Il s'agit de prédire, pour une configuration donnée, l'ensemble des contraintes susceptibles d'être actives aux optima, typiquement les contraintes thermiques sur les lignes de transport. Ce paradigme s'appuie sur l'observation bien documentée que, dans un grand réseau, seule une fraction négligeable des contraintes de contingence est active aux optima [52, 53]. Plusieurs travaux ont exploré des classificateurs tels que les réseaux neuronaux profonds [51, 52] ou les modèles de plus proches voisins [48], afin d'apprendre cette cartographie. Ces méthodes permettent de simplifier considérablement le problème initial en ne conservant qu'un sous-ensemble prédict de contraintes, réduisant ainsi la dimension et le temps de résolution du SCOPF. Leur efficacité repose cependant sur la précision de la prédiction : les erreurs de faux négatifs peuvent compromettre la faisabilité du problème réduit, d'où l'introduction de critères de minimisation asymétrique [53] ou de vérifications incrémentales de faisabilité.

Les sections qui suivent présentent de manière structurée ces deux familles de méthodes. L'analyse porte notamment sur les structures exploitées, les performances observées, et les compromis entre rapidité, faisabilité et qualité des solutions. Une attention particulière est portée aux cas à contraintes multiples, notamment dans les formulations de type SCOPF avec contingences  $N-1$  et  $N-k$ .

### 2.2.1 Apprentissage régressif de contraintes actives

L'approche régressive consiste à approximer directement une solution optimale du problème, en apprenant une relation fonctionnelle entre les conditions d'exploitation du réseau électrique et les variables de décision. Cette stratégie repose sur l'idée que, dans un grand nombre de cas d'usage du SCOPF, la solution dépend de manière régulière et stable des conditions opérationnelles du réseau, notamment des profils de charge et des paramètres topologiques ou physiques.

### 2.2.1.1 Réseaux neuronaux pour la régression directe des optima

Dans [50], les auteurs introduisent DeepOPF, une architecture fondée sur un réseau de neurones profond entraîné pour approximer directement la solution du SCOPF. Plus précisément, le modèle vise à apprendre une cartographie fonctionnelle

$$\mathbf{d} \mapsto \hat{\mathbf{p}} \in \mathbb{R}^{|\mathcal{G}|},$$

où  $\mathbf{d} \in \mathbb{R}^{|\mathcal{N}|}$  désigne le vecteur des charges nodales, et  $\hat{\mathbf{p}}$  l'estimation des puissances actives injectées par les générateurs, correspondant à une solution optimale ou proche de la valeur optimale du SCOPF. L'approche repose exclusivement sur des entrées d'injection nodale, sans recourir à des variables topologiques, aux paramètres de ligne, ni à des scénarios de contingence explicitement modélisés. Ce choix vise à maximiser la générnicité du modèle tout en conservant une faible complexité d'entrée.

L'architecture du réseau de neurones comprend plusieurs couches entièrement connectées avec fonctions d'activation de type ReLU, calibrées pour capturer la relation complexe entre les charges et les injections optimales. Le modèle est entraîné en minimisant une fonction de coût de type quadratique, mesurant l'écart entre la sortie prédite  $\hat{\mathbf{p}}$  et la solution de référence  $\mathbf{p}^*$  obtenue par résolution exacte du SCOPF.

Cependant, une limitation structurelle de cette approche réside dans l'absence de garantie sur la faisabilité des solutions prédites. En effet, les vecteurs  $\hat{\mathbf{p}}$  générés par le réseau peuvent ne pas respecter les contraintes de sécurité sur les lignes de transport, notamment les bornes thermiques  $\bar{s}_{ij}$  ou les équilibres nodaux. Pour pallier à cette limite, les auteurs de [50] proposent une procédure de projection postérieure, fondée sur un LP de correction :

$$\min_{\mathbf{p} \in \mathcal{F}(\mathbf{d})} \|\mathbf{p} - \hat{\mathbf{p}}\|_1, \quad (2.27)$$

où  $\mathcal{F}(\mathbf{d})$  désigne l'ensemble des solutions admissibles au SCOPF pour la charge  $\mathbf{d}$ , et la norme  $\ell_1$  est choisie pour favoriser la parcimonie des ajustements correctifs. Cette étape permet de projeter la prédiction sur la variété faisable, au prix d'une légère augmentation du temps de calcul.

Les expériences réalisées sur les cas tests IEEE-30, IEEE-118 et PEGASE-1354 montrent que l'approche DeepOPF permet de réduire drastiquement le temps de calcul par rapport à une résolution directe du SCOPF, tout en conservant une qualité de solution satisfaisante. L'écart relatif par rapport au coût optimal reste inférieur à 0,2 % dans la majorité des cas, avec un taux de réussite élevé après projection. Cette méthode illustre donc la faisabilité d'un

apprentissage direct des variables optimales dans un contexte réaliste, tout en soulignant l'importance d'un mécanisme explicite de rétablissement de la faisabilité.

### 2.2.1.2 Garanties de faisabilité et mécanismes de réparation intégrés

Si les approches purement régressives, telles que celle introduite dans [50], offrent des prédictions rapides des variables optimales du SCOPF, elles souffrent d'une limitation fondamentale : en l'absence de structure explicite du problème, rien ne garantit que la prédiction soit faisable. Les vecteurs de puissance prédits peuvent ne pas respecter les contraintes d'équilibre nodal, de limites thermiques  $\bar{s}_{ij}$  sur les lignes de transport  $(i, j) \in \mathcal{L}$ , ou encore les contraintes de sécurité post-contingence. La question de faisabilité est particulièrement critique dans les réseaux congestionnés ou soumis à de multiples scénarios de contingence [48].

Pour remédier à cette difficulté, plusieurs contributions récentes ont proposé des mécanismes explicites de réparation [54] ou de projection des prédictions dans l'espace admissible [55, 56], intégrés soit comme étape postérieure [54], soit comme composante différentiable de l'apprentissage [55, 56]. Ces approches visent à assurer que la solution finale satisfasse les contraintes de sécurité tout en préservant un faible écart à la prédiction initiale.

Dans [54], les auteurs proposent une méthode hybride combinant apprentissage profond et optimisation robuste. Leur chaîne de traitement (*pipeline*) repose sur un réseau de neurones  $\phi_\theta(\mathbf{d})$  prédisant un vecteur de puissance injectée  $\hat{\mathbf{p}} = \phi_\theta(\mathbf{d})$  à partir du profil de charge nodale  $\mathbf{d} \in \mathbb{R}^{|\mathcal{N}|}$ . Pour rétablir la faisabilité, ils introduisent une étape de correction fondée sur un algorithme de rétablissement de la faisabilité par génération de colonnes et de contraintes (*feasibility recovery CCGA*), défini comme suit :

$$\begin{aligned} \min_{\mathbf{p} \in \mathcal{F}(\mathbf{d})} \quad & \|\mathbf{p} - \hat{\mathbf{p}}\|_2^2 \\ \text{s.c.} \quad & \mathbf{A}\mathbf{p} \leq \mathbf{b} \end{aligned} \tag{2.28}$$

où  $\mathcal{F}(\mathbf{d})$  désigne l'ensemble des solutions réalisables du SCOPF pour la charge  $\mathbf{d}$ , et  $\mathbf{A}\mathbf{p} \leq \mathbf{b}$  regroupe l'ensemble des contraintes physiques du DCOPF. Cette projection, résolue par un solveur rapide et démarrée à chaud, garantit la faisabilité de la solution finale, avec un écart d'optimalité inférieur à 0,1 % sur des réseaux allant jusqu'à 1888 noeuds.

Dans [55], les auteurs développent une approche dite apprentissage faisable bout en bout (*end-to-end feasible learning*), dans laquelle la projection est directement intégrée dans le pipeline différentiable d'apprentissage. Le prédicteur  $\phi_\theta(\mathbf{d})$  est couplé à une couche de projection  $\Pi_{\mathcal{F}}$ ,

implémentée comme une opération différentiable définie par :

$$\Pi_{\mathcal{F}}(\hat{\mathbf{p}}) := \arg \min_{\mathbf{p} \in \mathcal{F}(\mathbf{d})} \|\mathbf{p} - \hat{\mathbf{p}}\|_2^2. \quad (2.29)$$

L'entraînement s'effectue alors sur la composition  $\mathbf{p}_{\text{proj}} = \Pi_{\mathcal{F}}(\phi_{\boldsymbol{\theta}}(\mathbf{d}))$ , permettant de rétro-propager l'erreur jusqu'aux paramètres  $\boldsymbol{\theta}$ , tout en assurant que la sortie finale satisfait les contraintes du SCOPF. Cette structure permet une généralisation robuste sur des réseaux de grande taille (PEGASE-1354, NESTA-162) et s'intègre de manière transparente dans un environnement opérationnel.

Enfin, [56] propose une approche interprétable fondée sur des arbres de décision affines, dans lesquels chaque feuille  $k$  correspond à une région polytopic de l'espace des charges  $\mathbf{d} \in \mathbb{R}^{|\mathcal{N}|}$ . Plus précisément, chaque feuille est associée à :

- un système d'inégalités linéaires  $\mathbf{A}_k \mathbf{d} \leq \mathbf{b}_k$  définissant la région de validité de la règle locale (c'est-à-dire l'ensemble des charges qui mènent à cette feuille) ;
- une prédiction affine  $\hat{\mathbf{p}} = \mathbf{B}_k \mathbf{d} + \mathbf{c}_k$ , représentant l'estimation des injections optimales dans cette région.

Pour chaque région  $k$ , les auteurs définissent une procédure analytique de projection dans l'espace des solutions admissibles, qui exploite la linéarité des contraintes du DC-OPF afin de corriger la prédiction tout en assurant sa faisabilité.

Le modèle est entraîné en minimisant une fonction de coût composée de deux termes : un terme d'erreur quadratique sur la variable projetée, et un terme de pénalisation sur les limites de contraintes thermiques sur les lignes de transport :

$$\mathcal{L}(\boldsymbol{\theta}) = \|\mathbf{p}_{\text{proj}} - \mathbf{p}^*\|_2^2 + \lambda \max_{(i,j) \in \mathcal{L}} [p_{ij}^2 - (\bar{s}_{ij})^2]_+, \quad (2.30)$$

où :

- $\mathbf{p}^*$  est la valeur optimale du SCOPF obtenue par résolution exacte ;
- $\mathbf{p}_{\text{proj}}$  est la solution prédite par le modèle, puis projetée analytiquement dans l'espace faisable ;
- $p_{ij}$  désigne la puissance active écoulée sur la ligne de transport  $(i, j) \in \mathcal{L}$  associée à la prédiction ;
- $\bar{s}_{ij}$  est la limite thermique de cette ligne ;
- $[\cdot]_+ = \max(\cdot, 0)$  indique que seules les non-respects sont pénalisés ;
- $\lambda > 0$  est un coefficient de pondération contrôlant le compromis entre précision prédictive et faisabilité.

Ce cadre permet de produire des prédictions interprétables, construites à partir d'une partition

polyédrique explicite de l'espace des charges, tout en intégrant un mécanisme de régularisation visant à réduire les violations de contraintes thermiques. Bien que la faisabilité ne soit pas garantie de manière formelle, les résultats expérimentaux indiquent une satisfaction effective des contraintes sur l'ensemble des échantillons testés, après projection dans l'espace admissible. En outre, cette approche autorise une analyse fine des variables critiques responsables des dépassements de limites d'écoulement de puissance.

Ces trois contributions illustrent différentes stratégies pour garantir la robustesse et la validité physique des prédictions apprises dans un contexte SCOPF. Elles constituent un progrès important vers l'intégration fiable de l'apprentissage automatique dans les outils opérationnels de gestion des réseaux électriques.

### 2.2.1.3 Architectures hybrides classification-régression pour la prédiction structurée

L'approche purement régressive, examinée dans les sections précédentes, consiste à estimer directement les variables de décision optimales (typiquement les puissances actives injectées  $p_i^*$ ) en fonction de l'état du réseau électrique. Toutefois, cette stratégie tend à négliger certaines structures discrètes fondamentales du problème, notamment la prévisibilité de certaines bornes de saturation sur les générateurs, observée empiriquement dans les solutions du SCED. Une telle structure suggère la possibilité de restreindre la régression aux seules variables dont la valeur optimale ne correspond pas à une borne active, simplifiant ainsi significativement la tâche d'apprentissage.

Dans cette perspective, [57] propose une architecture hybride articulée autour de deux composantes successives : un classificateur  $\mathfrak{C}_{\mathbf{w}_1} : \mathbb{R}^d \rightarrow \{0, 1\}^{2|\mathcal{G}|}$  permettant d'identifier les générateurs atteignant leurs bornes de production active (inférieure ou supérieure), suivi d'un régresseur  $\mathfrak{R}_{\mathbf{w}_2} : \mathbb{R}^{d+2|\mathcal{G}|} \rightarrow \mathbb{R}^{|\mathcal{G}|}$ , chargé de prédire les valeurs optimales  $p_i$  uniquement pour les générateurs dits non contraints. La fonction de coût associée à l'étape de classification repose sur l'entropie croisée :

$$\mathbf{w}_1^* \in \arg \min_{\mathbf{w}_1} \sum_{i=1}^N \mathfrak{L}_c(\mathbf{y}_i^c, \mathfrak{C}_{\mathbf{w}_1}(\mathbf{x}_i)), \quad (2.31)$$

où  $\mathfrak{L}_c$  désigne la perte logistique binaire, et  $\mathbf{y}_i^c \in \{0, 1\}^{2|\mathcal{G}|}$  encode l'état de saturation des générateurs. L'étape de régression s'écrit quant à elle :

$$\mathbf{w}_2^* \in \arg \min_{\mathbf{w}_2} \sum_{i=1}^N \mathfrak{L}_r \left( \mathbf{y}_i^r, \mathcal{E} \left( \mathfrak{R}_{\mathbf{w}_2}(\mathbf{x}_i), \mathfrak{C}_{\mathbf{w}_1^*}(\mathbf{x}_i) \right) \right), \quad (2.32)$$

où  $\mathfrak{L}_r$  est une erreur absolue moyenne (MAE) et  $\mathcal{E}$  est un opérateur d'agrégation qui impose la cohérence entre les prédictions de la régression et le masque fourni par la classification.

L'architecture repose sur un bloc dit d'interventions chirurgicales latentes (*latent surgical interventions*, LSI), qui permet d'adapter dynamiquement le réseau de neurones à la structure induite par les prédictions de saturation. Ce mécanisme agit comme une opération d'activation conditionnelle (*gating*), permettant d'annuler dynamiquement certaines sorties du régresseur en fonction des prédictions du classificateur. Cette structuration améliore significativement la précision de l'apprentissage, notamment pour les générateurs de petite ou moyenne taille, et permet une inférence en temps réel de plusieurs ordres de grandeur plus rapide que la résolution explicite du SCED [57].

### 2.2.2 Méthodes de classification pour l'identification de contraintes actives

La section précédente a montré que la prédiction des variables optimales par régression peut, dans certains cas, bénéficier d'une structuration préalable de l'espace de sortie, comme en témoigne l'approche hiérarchique de [57]. Dans cette perspective, l'identification directe des contraintes susceptibles de saturer aux minima offre une alternative prometteuse. Ce changement de paradigme repose sur une observation structurelle majeure : seule une fraction réduite des contraintes d'un problème de type SCED ou SCOPF est effectivement active aux optima. Cette propriété a été formalisée par [23] qui démontrent que même dans des réseaux de grande taille, le nombre de lignes simultanément saturées reste très limité. Par exemple, dans le réseau ERCOT, parmi plus de 10 000 lignes de transport, jamais plus de 14 contraintes n'ont été actives simultanément durant toute l'année 2013, et ce, seulement lors de deux journées exceptionnelles. Cette rareté des contraintes réellement actives justifie le recours à des modèles de classification supervisée, capables de prédire à partir des conditions d'exploitation les contraintes à inclure dans la formulation réduite du problème, tout en garantissant l'exactitude ou la sécurité des solutions obtenues.

Ceci donne lieu à une nouvelle famille de méthodes, centrées sur la classification supervisée de l'activation des contraintes. Ces approches consistent à apprendre, à partir d'un vecteur de conditions du réseau électrique (typiquement, la charge nodale  $\mathbf{P}_d$ ), une prédiction du statut binaire de chaque contrainte d'inégalité (active ou inactive), souvent sous la forme d'un vecteur booléen. Le problème devient alors celui de la sélection structurée d'un sous-ensemble pertinent de contraintes, suffisant pour caractériser la solution optimale du SCED ou du SCOPF dans une formulation réduite.

Les sections qui suivent présentent les fondements de cette approche, ses applications à l'écoulement de puissance linéarisé, et les mécanismes d'intégration de ces prédictions dans

une résolution fiable, même en présence d'erreurs de classification.

### 2.2.2.1 Classification supervisée par réseaux neuronaux profonds

Un premier groupe de méthodes de classification repose sur l'utilisation de réseaux de neurones profonds pour prédire l'activation des contraintes aux optima, à partir des conditions d'exploitation du réseau électrique. Dans cette famille de méthodes, l'objectif n'est plus de prédire directement les variables optimales du problème, mais d'identifier l'ensemble des contraintes susceptibles de saturer aux minima. Ce problème est formalisé comme une tâche de classification binaire pour chaque contrainte, à partir d'un vecteur de conditions d'exploitation du réseau électrique, tel que la charge nodale  $\mathbf{P}_d \in \mathbb{R}^{|\mathcal{N}|}$ . Soit un ensemble  $\mathcal{K}$  de contraintes candidates. L'apprentissage consiste à approximer une fonction de décision

$$\phi_{\theta}: \mathbb{R}^{|\mathcal{N}|} \longrightarrow \{0, 1\}^{|\mathcal{K}|},$$

où chaque sortie  $\phi_{\theta}^{(k)}(\mathbf{P}_d)$  indique si la contrainte  $k \in \mathcal{K}$  est active dans la solution optimale associée à  $\mathbf{P}_d$ .

Dans [51], les auteurs introduisent une approche fondée sur un autoencodeur profond débruitant (*stacked denoising autoencoder*), entraîné pour capturer les représentations non linéaires de la charge nodale et prédire efficacement les contraintes actives. Le modèle prend en entrée les charges normalisées sur chaque nœud et produit, en sortie, un vecteur de scores associés aux contraintes de sécurité. Pour chaque contrainte de ligne  $(i, j) \in \mathcal{L}$ , une variable binaire  $y_{ij} \in \{0, 1\}$  indique son activation aux optima. La fonction de perte adoptée est la somme des entropies croisées binaires sur l'ensemble des lignes de transport :

$$\mathfrak{L}(\theta) = - \sum_{(i,j) \in \mathcal{L}} [y_{ij} \log \hat{y}_{ij} + (1 - y_{ij}) \log(1 - \hat{y}_{ij})],$$

où :

- $y_{ij} \in \{0, 1\}$  est la variable cible indiquant si la contrainte sur la ligne  $(i, j) \in \mathcal{L}$  est active à l'optimalité du SCOPF (valeur 1 si la contrainte est saturée, 0 sinon) ;
- $\hat{y}_{ij} \in (0, 1)$  est la probabilité prédite par le modèle pour l'activation de la contrainte sur la ligne  $(i, j)$  ;
- $\theta$  désigne les paramètres du modèle appris par descente de gradient.

Le vecteur de sortie est ensuite utilisé pour sélectionner un sous-ensemble de contraintes à inclure dans une résolution restreinte du problème. Le nombre de contraintes sélectionnées peut être fixé à un seuil ou déterminé adaptativement à partir d'un seuil probabiliste. L'approche permet de diviser par un facteur supérieur à 10 le temps de résolution sur des réseaux IEEE-118

et PEGASE-1354, tout en maintenant un écart d'optimalité empirique inférieur à 0,2%.

L'article [58] généralise cette approche en s'affranchissant de la condition  $N=1$ , et en proposant un réseau de neurones entièrement connecté entraîné pour prédire, pour chaque contingence  $c \in \mathcal{C}$ , un vecteur de scores  $\mathbf{s}_c \in [0, 1]^{\mathcal{L}}$  indiquant la probabilité que la contrainte associée à chaque ligne soit active dans le scénario post-contingence. Le problème est alors reformulé comme une tâche de classification multilabel, chaque entrée  $(\mathbf{d}, c)$  donnant lieu à une prédiction structurée sur l'espace des contraintes. Le modèle est entraîné sur une base de données de paires  $(\mathbf{d}, c) \rightarrow \mathbf{y}_c$ , où  $\mathbf{y}_c$  encode le statut binaire d'activité des contraintes dans le scénario  $c$ . Une attention particulière est portée à la calibration des seuils de décision afin de minimiser les erreurs de type II (fausse inactivation), susceptibles de compromettre la sécurité du réseau. L'étude démontre de manière expérimentale que l'approche permet de réduire drastiquement la taille des programmes SCOPF à résoudre, tout en assurant un taux empirique de non-respect de limite inférieur à 0,5% sur l'ensemble des scénarios.

Les méthodes de classification pour l'identification de contraintes actives reposent sur l'hypothèse que les contraintes actives présentent une structure stable ou prévisible, et que les modèles de réseau neuronal peuvent apprendre cette structure à partir de données historiques. En exploitant la redondance structurelle présente dans les ensembles actifs, elles permettent une réduction importante de la taille du problème d'optimisation, ouvrant la voie à une inférence rapide en contexte temps réel.

Dans ce cadre, une autre approche consiste à considérer directement l'ensemble actif global comme une entité discrète, à prédire comme une classe dans un espace en classes multiples structuré.

### 2.2.2.2 Classification multiples des ensembles actifs optimaux par réseaux neuronaux

Dans cette perspective, chaque profil de charge  $\mathbf{d} \in \mathbb{R}^{|\mathcal{N}|}$  est associé à un unique ensemble actif optimal  $\mathcal{A}_k \subset \{1, \dots, m\}$ , où  $m$  désigne le nombre total de contraintes d'inégalité. L'objectif est d'apprendre une fonction de classification supervisée  $\mathfrak{C} : \mathbb{R}^{|\mathcal{N}|} \rightarrow \{1, \dots, K\}$  identifiant l'indice de l'ensemble actif  $\mathcal{A}_k$  parmi un dictionnaire  $\{\mathcal{A}_1, \dots, \mathcal{A}_K\}$  construit à partir de données historiques.

Dans [52], les auteurs montrent empiriquement que, malgré la croissance exponentielle du nombre théorique d'ensembles actifs possibles, seuls quelques-uns sont effectivement actifs dans la pratique pour un profil d'incertitude donné. L'incertitude  $\boldsymbol{\omega} \in \mathbb{R}^{|\mathcal{N}|}$  modélise une déviation aléatoire de la charge nodale par rapport à un vecteur de charge nominale  $\boldsymbol{\mu}$ , selon

la relation  $\mathbf{d} = \boldsymbol{\mu} + \boldsymbol{\omega}$ , avec  $\boldsymbol{\omega} \sim \mathcal{N}(0, \boldsymbol{\sigma})$ , où  $\boldsymbol{\sigma}$  est l'écart-type de la dérivation.

À partir de cette observation, les auteurs proposent une stratégie en deux étapes : (i) collecter, via échantillonnage, les ensembles actifs pertinents  $\mathcal{A}_k$  associés à des réalisations aléatoires de l'incertitude  $\boldsymbol{\omega}$ , et (ii) entraîner un réseau neuronal profond  $f_\theta: \boldsymbol{\omega} \mapsto \hat{k}$  prédisant l'indice  $\hat{k}$  de l'ensemble actif à partir de  $\boldsymbol{\omega}$ . La solution optimale s'obtient alors par l'évaluation explicite de l'expression analytique

$$\mathbf{p}^*(\boldsymbol{\omega}) = \mathbf{B}_{\mathcal{A}_{\hat{k}}}^\dagger \begin{bmatrix} \mathbf{b}_{\mathcal{A}_{\hat{k}}} + \mathbf{C}_{\mathcal{A}_{\hat{k}}} \boldsymbol{\omega} \\ \mathbf{e}^\top (\mathbf{d} - \boldsymbol{\mu} - \boldsymbol{\omega}) \end{bmatrix},$$

où :

- $\mathcal{A}_{\hat{k}} \subset \mathcal{K}$  est l'ensemble des contraintes actives prédictives pour la réalisation  $\boldsymbol{\omega}$  ;
- $\mathbf{B}_{\mathcal{A}_{\hat{k}}} \in \mathbb{R}^{(|\mathcal{G}|+1) \times |\mathcal{G}|}$  est une matrice formée par les lignes de contraintes actives (issues de  $\mathbf{Ap} \leq \mathbf{b} + \mathbf{C}\boldsymbol{\omega}$ ) augmentées de la contrainte d'équilibre  $\mathbf{e}^\top \mathbf{p} = \mathbf{e}^\top \mathbf{d}$  ;
- $\mathbf{b}_{\mathcal{A}_{\hat{k}}} + \mathbf{C}_{\mathcal{A}_{\hat{k}}} \boldsymbol{\omega}$  est le vecteur second membre des contraintes actives après réalisation de l'incertitude ;
- $\mathbf{e} \in \mathbb{R}^{|\mathcal{G}|}$  est le vecteur tout-unité imposant la somme des injections égale à la charge totale.

Ce procédé garantit une estimation exacte de la solution lorsque  $\hat{k}$  correspond à l'ensemble actif réel, et repose sur la linéarité des contraintes du DC-OPF. Il permet ainsi de contourner entièrement l'étape d'optimisation, à condition que la classification soit correcte.

Dans une contribution complémentaire, [59] proposent un cadre plus général, applicable à une large classe de programmes convexes. Ils démontrent que les ensembles actifs optimaux définissent une partition de l'espace des données d'entrée, chaque région correspondant à un polytope défini par les contraintes saturées. En exploitant cette structure, ils démontrent que la classification en classes multiples des ensembles actifs peut être vue comme une approximation de la solution optimale par une fonction par morceaux linéaire, chaque morceau étant défini par un ensemble actif. L'intérêt de cette approche est double : d'une part, elle permet une inférence rapide des solutions optimales, d'autre part, elle fournit une base pour des stratégies hybrides mêlant apprentissage statistique et techniques déterministes.

Enfin, les auteurs montrent que cette stratégie de classification peut être renforcée en utilisant des prédictions de rang  $K$ , où le classificateur renvoie un ensemble restreint de  $K$  ensembles actifs candidats. Une recherche par évaluation explicite dans cette liste permet de combiner rapidité et robustesse, en évitant les effets négatifs d'une mauvaise prédiction unique. Cette stratégie est appelée politique d'ensemble top- $K$  (*top-k ensemble policy*) dans [52], et présente une très bonne performance même pour des systèmes de grande taille.

Ces approches positionnent la classification en classes multiples des ensembles actifs comme une méthode centrale pour l'accélération des problèmes d'optimisation en contexte opérationnel incertain, tout en exploitant la structure combinatoire du problème. Toutefois, elles ne tirent parti que de manière implicite de la structure géométrique sous-jacente à la solution optimale. Dans le cas d'un SCOPF formulé comme un programme linéaire ou quadratique convexe, la solution optimale est une fonction affine par morceaux des paramètres d'entrée, tels que le vecteur de charge. Chaque morceau correspond à un ensemble actif donné, et définit une région de faisabilité convexe, généralement un polytope fermé, dans l'espace des paramètres. Ces régions recouvrent l'espace faisable de manière conjointe, avec des intérieurs disjoints mais des frontières possiblement partagées [21]. Cette structure de partition polytopic est fondamentale pour justifier l'usage de classificateurs géométriques.

### 2.2.2.3 Méthodes géométriques et classificateurs à séparation polytopic

Dans le cas du SCOPF convexe, l'espace des conditions d'exploitation du réseau induit une structure géométrique remarquable : les classes d'ensembles actifs optimaux correspondent à des régions convexes, délimitées par des hyperplans issus des contraintes d'inégalité saturées. Cette structure résulte directement des conditions KKT associées à des programmes quadratiques convexes paramétrés, dans lesquels la solution optimale reste affine par morceaux en fonction du vecteur de paramètres  $\mathbf{w}$ , et où chaque région affine correspond à un ensemble actif constant [21]. Cette classe de problèmes s'écrit sous la forme :

$$\min_{\mathbf{x}} \quad \frac{1}{2} \mathbf{x}^\top \mathbf{H} \mathbf{x} + \mathbf{g}^\top \mathbf{x} \quad (2.33a)$$

$$\text{s.c.} \quad \mathbf{C} \mathbf{x} + \mathbf{D} \mathbf{w} = \mathbf{0} \quad (2.33b)$$

$$\mathbf{A} \mathbf{x} + \mathbf{B} \mathbf{w} \leq \mathbf{0}, \quad (2.33c)$$

où  $\mathbf{w}$  est un vecteur de paramètres (ici, les injections nodales nettes,  $\mathbf{p}_d$ ), et les matrices  $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$  encodent les contraintes physiques du réseau. Comme démontré dans [60], l'ensemble des points  $\mathbf{w} \in \mathbb{R}^{|\mathcal{W}|}$  conduisant à un même ensemble actif optimal forme un polytope convexe. Cette structure partitionne l'espace des conditions d'exploitation en régions polytopicques disjointes, chacune associée à un ensemble actif donné. La séparation de ces régions repose sur le résultat fondamental suivant :

**Théorème 1** (Hyperplan séparateur [61]). *Soient  $\mathcal{D} \subset \mathbb{R}^n$  et  $\mathcal{E} \subset \mathbb{R}^n$  deux ensembles convexes, non vides et disjoints, i.e.,  $\mathcal{D} \cap \mathcal{E} = \emptyset$ . Alors, il existe un vecteur  $\mathbf{a} \in \mathbb{R}^n$ ,  $\mathbf{a} \neq \mathbf{0}$ ,*

et un scalaire  $b \in \mathbb{R}$  tels que :

$$\mathbf{a}^\top \mathbf{x} \leq b \quad \text{pour tout } \mathbf{x} \in D, \quad \text{et} \quad \mathbf{a}^\top \mathbf{x} \geq b \quad \text{pour tout } \mathbf{x} \in \mathcal{E}.$$

L'hyperplan  $\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^\top \mathbf{x} = b\}$  est alors dit séparateur de  $\mathcal{D}$  et  $\mathcal{E}$ .

Ainsi, deux ensembles actifs distincts induisent des régions disjointes, séparables par un hyperplan affine. Ce résultat, garanti par le théorème 1, justifie l'utilisation de classificateurs linéaires pour discriminer les ensembles actifs.

Plutôt que de recourir à des architectures neuronales apprenant implicitement la frontière entre classes, les auteurs proposent d'utiliser des classificateurs linéaires construits explicitement à partir des propriétés géométriques du problème, en particulier l'analyse discriminante linéaire (*linear discriminant analysis*, LDA). Leur approche repose sur une résolution en deux étapes :

1. prédire l'ensemble actif optimal comme une classe discrète  $\mathcal{A}_k$ , via un classificateur supervisé en classes multiples  $\psi_\theta(\mathbf{w}) \in \{\mathcal{A}_1, \dots, \mathcal{A}_K\}$ ;
2. calculer la solution optimale  $\mathbf{x}^*$  en résolvant les équations de KKT associées à l'ensemble  $\mathcal{A}_k$  prédit, ce qui revient à une simple évaluation linéaire  $\mathbf{x}^* = \mathbf{L}_k \mathbf{w}$ , où  $\mathbf{L}_k \in \mathbb{R}^{n \times d}$  est la matrice d'application associée à l'ensemble actif  $\mathcal{A}_k$ , construite en inversant le système d'égalités de KKT restreint aux contraintes actives, et dépendant uniquement des coefficients du problème original.

L'étude comparative menée dans [60] montre que l'analyse discriminante linéaire présente une excellente adéquation à cette tâche. Contrairement aux réseaux de neurones ou aux arbres de décision, la LDA ne nécessite aucun réglage d'hyperparamètres et fournit directement des probabilités *a posteriori* interprétables. Ces probabilités permettent de sélectionner les  $k$  ensembles actifs les plus probables, d'évaluer les solutions correspondantes, et de retenir celle qui satisfait les contraintes tout en minimisant la fonction objectif.

Les résultats obtenus sur le réseau IEEE-162-DTC avec 50 % de production renouvelable montrent que LDA atteint une précision empirique de 97,8 % en prédiction du bon ensemble actif parmi les 15 meilleurs candidats, tout en surpassant les autres méthodes sur le critère du temps d'inférence [60]. Cette efficacité computationnelle s'explique par le fait que l'entraînement du LDA se réduit à l'estimation des moyennes et covariances de chaque classe, sans optimisation itérative. En exploitant explicitement les propriétés géométriques du DCOPF, l'approche proposée offre ainsi une alternative frugale et robuste aux prédicteurs neuronaux plus complexes.

## 2.3 Méthodes de réduction de contraintes dans la formulation SCOPF

Les sections précédentes ont présenté diverses méthodes d'apprentissage supervisé visant à prédire directement l'ensemble actif d'un problème SCOPF, soit en apprenant une règle de décision binaire pour chaque contrainte de manière indépendante, soit en identifiant globalement l'ensemble actif au moyen d'une classification en classes multiples. Ces approches s'appuient sur la régularité empirique des ensembles actifs, exploitée par des modèles d'apprentissage entraînés sur des historiques de profils de charge.

Cependant, cette stratégie soulève plusieurs limitations fondamentales. D'une part, la prédiction d'ensembles actifs souffre d'un manque de garanties formelles : les contraintes prédictées comme actives peuvent ne pas conduire à une solution réalisable, ou au contraire négliger des contraintes critiques, compromettant la sûreté du réseau. D'autre part, ces modèles doivent en général apprendre une correspondance complexe entre conditions d'exploitation  $\mathbf{p}_d \in \mathbb{R}^{|\mathcal{N}|}$  et configurations optimales, dans un espace combinatoire extrêmement vaste.

À l'inverse, les méthodes dites de réduction de contraintes (*constraint screening*) ne cherchent pas à prédire les optima, mais à reformuler le problème SCOPF en conservant uniquement les contraintes pertinentes. Cette famille d'approches repose sur un principe structurel : dans la majorité des cas, seule une faible proportion des contraintes, de sécurité ou de capacité, est effectivement active aux optima du SCED [23], et même utile pour définir l'ensemble de faisabilité du problème. Cette observation est confirmée empiriquement à large échelle, comme dans [12], où il est démontré que plus de 95 % des contraintes de sécurité peuvent être écartées sans affecter la solution optimale, y compris sur des réseaux de dimension réelle tels que le système européen ENTSO-E.

Ces techniques visent donc à identifier a priori un sous-ensemble minimal de contraintes à conserver, soit à l'aide de tests structurels, soit via des modèles prédictifs apprenant à estimer l'utilité de chaque contrainte en fonction des conditions du réseau, dans le but de limiter l'empreinte mémoire et le coût de résolution. Elles permettent de simplifier considérablement la formulation SCOPF, réduisant à la fois les temps de calcul et la charge mémoire, tout en facilitant l'implémentation de méthodes de décomposition ou d'optimisation parallèle.

### 2.3.1 Fondements de la détection de redondance

Sur le plan théorique, la réduction de contraintes revient à identifier, dans un système de contraintes d'inégalités affines  $\mathbf{A}_0\mathbf{x} \leq \mathbf{b}_0$ , les contraintes qui sont redondantes, c'est-à-dire déduites des autres. Soit  $\mathbf{a}_i\mathbf{x} \leq t$  une contrainte d'inégalité candidate à la suppression, correspondant à une ligne particulière du système  $\mathbf{A}_0\mathbf{x} \leq \mathbf{b}_0$ , où  $\mathbf{a}_i \in \mathbb{R}^n$  est une rangée de la

matrice  $\mathbf{A}_0$ , et  $t \in \mathbb{R}$  l'entrée correspondante dans le vecteur second membre  $\mathbf{b}_0$ . Considérons le sous-système  $\mathbf{Ax} \leq \mathbf{b}$ , formé par l'ensemble des autres lignes de  $\mathbf{A}_0$ , à l'exception de  $\mathbf{a}_i$ . La question est de savoir si  $\mathbf{a}_i \mathbf{x} \leq t$  est logiquement impliquée par  $\mathbf{Ax} \leq \mathbf{b}$ , autrement dit, si elle est satisfaite par tout  $\mathbf{x} \in \mathbb{R}^n$  tel que  $\mathbf{Ax} \leq \mathbf{b}$ .

Ce test de redondance se formule naturellement comme un programme linéaire (*linear programming test*, LP-test) [62] :

$$\max_{\mathbf{x}} \quad \mathbf{a}_i \mathbf{x} \tag{2.34a}$$

$$\text{s.c.} \quad \mathbf{Ax} \leq \mathbf{b}, \tag{2.34b}$$

$$\mathbf{a}_i \mathbf{x} \leq t + 1. \tag{2.34c}$$

Si la valeur optimale  $f^*$  de (2.34) est telle que  $f^* \leq t$ , alors la contrainte d'inégalité  $\mathbf{a}_i \mathbf{x} \leq t$  est nécessairement satisfaite sur l'ensemble  $\{\mathbf{x} \mid \mathbf{Ax} \leq \mathbf{b}\}$ , et peut être éliminée sans modifier l'ensemble des solutions admissibles. Dans le cas contraire, elle est indispensable pour décrire correctement le polytope défini par  $\mathbf{A}_0 \mathbf{x} \leq \mathbf{b}_0$ .

En répétant cette opération pour chaque ligne de  $\mathbf{A}_0$ , nous obtenons un sous-système minimal (non nécessairement unique) équivalent, exempt de contraintes redondantes. Ce processus de filtrage est fondamental dans les méthodes modernes de réduction polytopique, et sert de base à de nombreuses stratégies d'accélération dans les formulations du SCOPF, comme le proposent les travaux de [63] et [64].

Toutefois, sa complexité devient prohibitive pour les systèmes comportant un très grand nombre de contraintes. Des améliorations fondées sur la méthode de balayage directionnel (*rayshoot*) introduite par Clarkson [62] permettent d'accélérer ce processus en restreignant les tests de redondance à un sous-ensemble réduit de contraintes déjà identifiées comme nécessaires, plutôt que de comparer chaque contrainte candidate à l'ensemble complet du système.

Dans un cadre plus opérationnel, [65] démontre que l'élimination préalable de contraintes redondantes permet d'améliorer les performances des solveurs d'optimisation mixte, notamment dans les problèmes de couverture d'ensembles (*set covering*) ou de localisation-tournées (*location-routing*), sans altérer l'espace des solutions admissibles.

Ces fondements justifient le développement de méthodes spécifiques au cas des réseaux électriques, où la topologie du réseau et la structure du problème permettent une exploitation fine de la redondance, notamment via les matrices PTDF et LODF. Les sections suivantes présentent les contributions récentes dans ce domaine, avec un accent particulier sur les

méthodes combinant tests géométriques, heuristiques structurelles, et apprentissage supervisé pour l'identification rapide des contraintes non critiques.

### 2.3.2 Méthodes déterministes de réduction de contraintes

Dans le contexte du SCOPF, les méthodes déterministes de réduction de contraintes exploitent la structure physique du réseau pour identifier des contraintes d'inégalités redondantes ou structurellement inactives, sans recourir à l'apprentissage statistique. Ces approches s'appuient sur des critères analytiques pour construire des sous-ensembles de contraintes dits dominants. Cette logique de couverture implicite repose sur l'observation que, pour un ensemble donné de scénarios de contingence, un nombre réduit de contraintes suffit souvent à garantir la sécurité du réseau dans toutes les situations pertinentes. Les travaux présentés dans cette section illustrent différentes déclinaisons de ce principe, selon qu'il s'agisse de tests géométriques, de formulations exactes ou de conditions suffisantes explicites.

Les méthodes présentées ici partagent une même intuition fondatrice : dans les formulations SCOPF ou SCED à grande échelle, une fraction significative des contraintes de sécurité n'est jamais active, ou bien se trouve redondante au regard d'un sous-ensemble plus restreint de contraintes dites dominantes. Cette redondance peut résulter de la topologie du réseau, des corrélations structurelles entre scénarios de contingence, ou encore de propriétés convexes de l'espace admissible des injections nodales [12, 66, 67]. Nous cherchons alors à identifier, avant même l'optimisation, un sous-ensemble minimal de contraintes suffisant pour garantir la sécurité du réseau, ce qui permet une réduction drastique de la taille du problème.

Dans [66], les auteurs introduisent la notion de contingence dominante (*umbrella contingency*). L'objectif est de réduire le nombre de contingences à inclure dans la formulation du SCOPF, tout en garantissant que la solution obtenue reste identique à celle du problème complet. Une contingence  $c \in \mathcal{C}$  est dite dominante sur un sous-ensemble  $\mathcal{C}' \subset \mathcal{C} \setminus \{c\}$  si la satisfaction des contraintes de sécurité associées à  $c$  implique automatiquement la faisabilité de la solution vis-à-vis de toutes les contingences  $c' \in \mathcal{C}'$ .

Pour identifier les contingences pertinentes, l'article propose un critère fondé sur les multiplicateurs de Lagrange associés aux contraintes d'équilibre post-contingence. Plus précisément, pour chaque  $c \in \mathcal{C}$ , on considère le vecteur  $\boldsymbol{\mu}^{(c)}$  des multiplicateurs de Lagrange associés aux contraintes d'équilibre nodal dans le régime post-contingence  $c$ . Une contingence est retenue si :

$$\|\boldsymbol{\mu}^{(c)}\|_p \geq \varepsilon,$$

où  $\varepsilon > 0$  est un seuil fixé à l'avance, et  $\|\cdot\|_p$  désigne une norme  $\ell_p$ . Ce critère repose sur une

analyse de sensibilité, selon laquelle la variation marginale de la fonction objectif par rapport à la sévérité du scénario  $c$  est proportionnelle à  $\boldsymbol{\mu}^{(c)}$ , i.e.,

$$\frac{\partial \mathcal{L}}{\partial \alpha^{(c)}} = \boldsymbol{\mu}^{(c)},$$

où  $\mathcal{L}$  est la fonction de Lagrangien du problème, et  $\alpha^{(c)}$  un paramètre pondérant l'intensité de la contingence  $c$ . Ainsi, seules les contingences dont l'impact marginal est significatif (en norme) sont conservées, les autres étant éliminées sans compromettre la sécurité de la solution finale.

Dans cet article, les auteurs valident cette approche sur un réseau test et montrent empiriquement que le nombre de contingences dominantes est en général très faible, stable par rapport à certaines variations de la charge, et suffisant pour garantir une couverture adéquate de la sécurité. Cette méthode offre donc un moyen rigoureux et efficient de filtrer les contingences à faible impact, avec des applications directes à la planification hors ligne des marchés d'électricité [66].

Cette démarche est approfondie dans deux articles successifs des mêmes auteurs, Ardakani et Bouffard, qui proposent une méthode de sélection déterministe, fondée sur une définition rigoureuse des contraintes parapluie (*umbrella constraints*), i.e., l'ensemble minimal de contraintes nécessaires à la description de l'enveloppe convexe réalisable du problème, dans [68] et [69].

Le premier article [68] formalise cette intuition sous forme d'un problème d'optimisation appelé découverte des contraintes parapluie (*umbrella constraint discovery*, UCD), qui vise à déterminer l'ensemble des contraintes dont la suppression modifierait l'enveloppe convexe réalisable. Soit  $j$  l'indice d'une contrainte d'inégalité candidate du système initial  $\mathbf{Ax} \leq \mathbf{b}$ , avec  $\mathbf{a}_j^\top$  la  $j$ -ième ligne de  $\mathbf{A}$  et  $b_j$  l'entrée correspondante de  $\mathbf{b}$ . La méthode UCD repose sur la résolution, pour chaque  $j \in \{1, \dots, \mathcal{J}\}$  où  $\mathcal{J}$  désigne le nombre total de contraintes initiales, du programme linéaire suivant :

$$\min_{\mathbf{w}_j, s_j} \quad s_j \tag{2.35}$$

$$\text{s.c.} \quad \mathbf{a}_{j'}^\top \mathbf{w}_j \leq b_{j'}, \quad \forall j' \in \{1, \dots, \mathcal{J}\} \setminus \{j\}, \tag{2.36}$$

$$\mathbf{a}_j^\top \mathbf{w}_j + s_j \geq b_j, \tag{2.37}$$

où  $\mathbf{a}_j^\top \mathbf{w}_j = b_j$  représente l'hyperplan de la contrainte  $j$ , et  $s_j \geq 0$  est une variable d'écart. Si  $s_j = 0$ , la contrainte est dite parapluie (i.e., son retrait change l'ensemble réalisable), sinon elle est redondante. Ce test est appliqué à toutes les contraintes du problème, ce qui nécessite la résolution de  $J$  programmes linéaires de taille  $\mathcal{O}(J)$ , rendant la méthode coûteuse

mais exhaustive. Les auteurs proposent alors une stratégie de décomposition récursive pour exploiter la structure modulaire du SCOPF et paralléliser les tests.

Dans l'article suivant [69], les auteurs introduisent une amélioration à travers la méthode UCD partiel (*partial UCD*, P-UCD), fondée sur une partition heuristique des contraintes en deux sous-ensembles :  $\mathcal{J}_u$ , l'ensemble des contraintes potentiellement parapluie, et  $\mathcal{J}_n$ , l'ensemble des contraintes potentiellement non-parapluie). L'idée est d'identifier rapidement, via un test partiel, les contraintes de  $\mathcal{J}_n$  qui sont redondantes par rapport aux contraintes de  $\mathcal{J}_u$ , à l'aide du programme suivant :

$$\min_{\mathbf{w}, \mathbf{s}} \quad \sum_{j \in \mathcal{J}_n} s_j \quad (2.38)$$

$$\text{s.c.} \quad \mathbf{a}_{j'}^\top \mathbf{w} \leq b_{j'}, \quad \forall j' \in \mathcal{J}_u, \quad (2.39)$$

$$\mathbf{a}_j^\top \mathbf{w} + s_j \geq b_j, \quad \forall j \in \mathcal{J}_n, \quad (2.40)$$

ce qui permet de concentrer les comparaisons sur un sous-ensemble supposé structurant. Cette approche, bien qu'approximative, conserve une propriété importante, selon laquelle toute contrainte identifiée comme redondante l'est de manière certaine, ce qui garantit que la région réalisable du problème initial n'est jamais altérée.

Les résultats numériques sur les réseaux IEEE RTS et IEEE 118-noeuds montrent une réduction notable de la taille du problème (jusqu'à un facteur 100 sur le temps de résolution du UCD), en combinant P-UCD et décomposition hiérarchique. Ces travaux jettent ainsi les bases d'un filtrage structurel des contraintes basé sur leur pouvoir de couverture géométrique, ouvrant la voie à une exploitation systématique de la redondance implicite dans les formulations sécuritaires.

Une perspective complémentaire à cette approche par couverture implicite est développée dans [70], qui propose une méthode rigoureuse de réduction a priori des contraintes de sécurité pour les problèmes de SCUC. Plutôt que de s'appuyer sur une caractérisation analytique des contraintes parapluies, les auteurs considèrent directement la région de sécurité, définie comme l'ensemble des configurations d'injection  $\mathbf{x} \in \mathbb{R}^{|\mathcal{N}|}$  satisfaisant à la fois les équations DC du réseau électrique et l'ensemble complet des contraintes de sécurité, exprimées sous la forme :

$$\mathbf{l}^{(c)} \leq \Phi^{(c)} \mathbf{x} \leq \mathbf{u}^{(c)}, \quad \forall c \in \mathcal{C},$$

où  $\Phi^{(c)} \in \mathbb{R}^{|\mathcal{L}| \times |\mathcal{N}|}$  est la matrice  $\Phi$  sous la contingence  $c$ , et  $\mathbf{l}^{(c)}, \mathbf{u}^{(c)} \in \mathbb{R}^{|\mathcal{L}|}$  désignent respectivement les bornes inférieure et supérieure admissibles pour les écoulements de puissance sur les lignes de transport.

La méthode repose sur une procédure de bornes fiables (*reliable bounds*) combinée à de l'arithmétique d'intervalles, permettant d'éliminer toute contrainte dont la plage admissible est disjointe d'une sur-approximation convexe de la région de sécurité. Plus précisément, considérons le polytope de sécurité défini par :

$$\mathcal{P} := \left\{ \mathbf{x} \in \mathbb{R}^{|\mathcal{N}|} \mid \mathbf{Ax} = \mathbf{b}, \quad \mathbf{l}^{(c)} \leq \Phi^{(c)} \mathbf{x} \leq \mathbf{u}^{(c)}, \quad \forall c \in \mathcal{C} \right\}, \quad (2.41)$$

où  $\mathbf{A} \in \mathbb{R}^{m \times |\mathcal{N}|}$  encode les contraintes d'équilibre de puissance, et chaque contingence  $c$  contribue à une famille de contraintes linéaires encodées par  $\Phi^{(c)}$ ,  $\mathbf{l}^{(c)}$  et  $\mathbf{u}^{(c)}$ . Le cœur de la méthode consiste à construire un hypercube englobant la projection de  $\mathcal{P}$  sur chaque direction canonique  $\mathbf{e}_i \in \mathbb{R}^{|\mathcal{N}|}$ , afin d'évaluer si une contrainte est incompatible avec toute réalisation admissible de l'injection, et donc éligible à l'élimination. Pour chaque composante  $i \in \{1, 2, \dots, |\mathcal{N}|\}$ , les bornes fiables sont obtenues à partir des programmes linéaires suivants :

$$\underline{x}_i^{\text{rel}} = \min_{\mathbf{x} \in \mathbb{R}^{|\mathcal{N}|}} x_i \quad (2.42a)$$

$$\text{s.c.} \quad \mathbf{Ax} = \mathbf{b}, \quad (2.42b)$$

$$\mathbf{l}^{(c)} \leq \Phi^{(c)} \mathbf{x} \leq \mathbf{u}^{(c)}, \quad \forall c \in \mathcal{C}, \quad (2.42c)$$

et

$$\bar{x}_i^{\text{rel}} = \max_{\mathbf{x} \in \mathbb{R}^{|\mathcal{N}|}} x_i \quad (2.43a)$$

$$\text{s.c.} \quad \mathbf{Ax} = \mathbf{b}, \quad (2.43b)$$

$$\mathbf{l}^{(c)} \leq \Phi^{(c)} \mathbf{x} \leq \mathbf{u}^{(c)}, \quad \forall c \in \mathcal{C}. \quad (2.43c)$$

Ces deux problèmes définissent les composantes des vecteurs de bornes fiables  $\underline{\mathbf{x}}^{\text{rel}}, \bar{\mathbf{x}}^{\text{rel}} \in \mathbb{R}^{|\mathcal{N}|}$ . L'enveloppe rectangulaire associée est donnée par l'hypercube :

$$\mathcal{B}^{\text{rel}} := \left\{ \mathbf{x} \in \mathbb{R}^{|\mathcal{N}|} \mid \underline{x}_i^{\text{rel}} \leq x_i \leq \bar{x}_i^{\text{rel}}, \quad \forall i \in \{1, 2, \dots, |\mathcal{N}|\} \right\}.$$

Le filtrage s'effectue ensuite en testant, pour chaque contrainte  $j$  associée à une contingence  $c \in \mathcal{C}$ , si l'hyperplan définie par  $\Phi_j^{(c)\top} \mathbf{x} \leq \mathbf{u}_j^{(c)}$  (ou  $\geq \mathbf{l}_j^{(c)}$ ) est automatiquement satisfaite pour tout  $\mathbf{x} \in \mathcal{B}^{\text{rel}}$ . Si tel est le cas, la contrainte est dite inactive par construction et peut être supprimée du modèle. Ce test est non itératif et parallélisable, ce qui en fait une stratégie de réduction de contraintes particulièrement adaptée aux formulations de très grande dimension.

L'efficacité de l'approche est démontrée sur des réseaux réalistes de grande taille, dont ERCOT, pour lequel les 26 681 334 contraintes d'inégalités linéaires initiales sont réduites à seulement

467 contraintes nécessaires après filtrage et raffinement complet, tout en conservant l'intégralité de la région admissible du problème [70].

L'approche introduite par [12] propose une méthode systématique de réduction de contraintes pour le problème de SCOPF, en exploitant à la fois la structure redondante du polytope de faisabilité et l'impact différencié des contingences sur les lignes de transport, dans un cadre multipériode. Cette méthode constitue l'une des rares contributions de réduction déterministe efficace appliquée aux formulations multipériodes du SCOPF.

Le point de départ de cette méthode repose sur la définition de l'ensemble de faisabilité sous forme polytopique. Soit  $\mathcal{I} \subset \mathbb{N}$  l'ensemble des indices correspondant aux contraintes thermiques sur les lignes de transport, associées à l'ensemble des contingences  $c \in \mathcal{C} \cup \{0\}$ . Nous définissons alors :

$$\mathcal{F}(\Phi, \bar{s}, \mathcal{I}) := \left\{ \mathbf{x} \in \mathbb{R}^{|\mathcal{N}|} \mid \Phi_i \mathbf{x} \leq \bar{s}_i, \forall i \in \mathcal{I} \right\},$$

où  $\Phi \in \mathbb{R}^{(|\mathcal{C}|+1)|\mathcal{L}| \times |\mathcal{N}|}$  désigne les coefficients des matrices  $\Phi$  associés aux contraintes de sécurité associées à l'ensemble des scénarios de contingence  $c \in \mathcal{C}$  ainsi qu'au cas de base  $c = 0$ . Cette matrice induit un système potentiellement surdéterminé d'inégalités linéaires, dont chaque bloc de lignes encode les écoulements sous une contingence donnée. Une contrainte d'indice  $k$  est dite essentielle si sa suppression modifie l'ensemble  $\mathcal{F}$ , autrement dit, si  $k \in \mathcal{I}$  est tel que  $\mathcal{F}(\Phi, \bar{s}, \mathcal{I} \setminus \{k\}) \neq \mathcal{F}(\Phi, \bar{s}, \mathcal{I})$ .

Pour détecter cette propriété, les auteurs proposent le test suivant :  $k$  est non-redondant si la solution optimale  $\mathbf{x}^*$  du LP-Test (2.34) suivant :

$$\max_{\mathbf{x} \in \mathbb{R}^{|\mathcal{N}|}} \Phi_k \mathbf{x} \tag{2.44a}$$

$$\text{s.c. } \Phi_i \mathbf{x} \leq \bar{s}_i, \quad \forall i \in \mathcal{I} \setminus \{k\}, \tag{2.44b}$$

$$\Phi_k \mathbf{x} \leq \bar{s}_k + 1 \tag{2.44c}$$

vérifie  $\Phi_k x^* > \bar{s}_k$ . Ce test, est garanti faisable (puisque  $\mathbf{x} = 0$  est toujours admissible), et permet d'identifier un ensemble minimal  $\mathcal{I}^*$  de contraintes non redondantes.

Afin de réduire la complexité combinatoire liée à l'application systématique du LP-Test, les auteurs introduisent un mécanisme de balayage directionnel (*rayshoot*), formalisé dans l'Algorithme 2. Celui-ci consiste à émettre un rayon à partir d'un point intérieur de l'espace de faisabilité  $\mathcal{F}$ ,  $\mathbf{z}$ , dans la direction  $\mathbf{x}^* - \mathbf{z}$ , et à identifier la première inégalité franchie, i.e, active. Cette contrainte est alors ajoutée à l'ensemble des contraintes essentielles.

L'algorithme global, présenté dans l'Algorithme 1, combine récursivement le test de redon-

dance (2.44) avec les appels successifs à *rayshoot*, pour construire efficacement un ensemble minimal  $\mathcal{I}^*$  de contraintes non redondantes.

---

**Algorithme 1 : REDUNDANCYREMOVAL( $\Phi, \bar{s}, z$ )**


---

**Input :** Un système d'inégalités  $\Phi x \leq \bar{s}$ ; un point strictement intérieur  $z \in \mathbb{R}^n$ .  
**Output :** L'ensemble  $\mathcal{I}$  des indices des contraintes d'inégalités non redondantes.

```

1  $\mathcal{I} \leftarrow \emptyset$ ;                                // Contraintes non redondantes
2  $\mathcal{J} \leftarrow \{1, \dots, M\}$ ;                      // Indices à tester
3 while  $\mathcal{J} \neq \emptyset$  do
4   Sélectionner un indice  $k \in \mathcal{J}$ ;
5    $(p^*, x^*) \leftarrow \text{LP-TEST}(\Phi, \bar{s}, \mathcal{I} \cup \{k\}, k)$ ;
6   if  $p^* > \bar{s}_k$  then
7      $j \leftarrow \text{RAYSHOOT}(\Phi, \bar{s}, z, x^*)$ ;
8      $\mathcal{I} \leftarrow \mathcal{I} \cup \{j\}$ ;
9      $\mathcal{J} \leftarrow \mathcal{J} \setminus \{j\}$ ;
10  else
11     $\mathcal{J} \leftarrow \mathcal{J} \setminus \{k\}$ ;
12 return  $\mathcal{I}$ 
```

---



---

**Algorithme 2 : RAYSHOOT( $\Phi, \bar{s}, z, x^*$ )**


---

**Input :** Système  $\Phi x \leq \bar{s}$ ; point intérieur  $z$ ; point  $x^*$  sur ou hors de  $\mathcal{F}$ .  
**Output :** L'indice  $i$  de la première contrainte franchie par le rayon.

```

1  $\epsilon \leftarrow \epsilon_0$ ;
2  $\mathcal{H} \leftarrow \emptyset$ ; // Initialisation de l'ensemble des contraintes non-respectées
3  $r \leftarrow \frac{x^* - z}{\|x^* - z\|_2}$ ;                                // Direction normalisée
4 repeat
5    $z \leftarrow z + \epsilon r$ ;                                // Avancée du rayon
6    $\mathcal{H} \leftarrow \{i \mid \Phi_i z > \bar{s}_i\}$ ;                // Contraintes franchies
7   if  $|\mathcal{H}| > 1$  then
8      $z \leftarrow z - \epsilon r$ ;                                // Retour
9      $\epsilon \leftarrow \epsilon/10$ ;                                // Réduction du pas
10 until  $|\mathcal{H}| = 1$ ;
11 return  $i \in \mathcal{H}$ 
```

---

Dans un second temps, les auteurs exploitent les bornes techniques sur les injections nodales pour affiner l'ensemble  $\mathcal{I}^*$  obtenu. Soit  $\mathcal{T} \subset \mathbb{N}$  l'ensemble des instants  $t$  de l'horizon temporel considéré. Pour chaque nœud  $n \in \mathcal{N}$ , on définit alors :

$$\check{x}_n = \min_{t \in \mathcal{T}} (p_{d,t,n}) + \min_{t \in \mathcal{T}} (\bar{p}_{n,t}),$$

$$\hat{x}_n = \max_{t \in \mathcal{T}} (\bar{p}_{n,t}),$$

où  $p_{d,t,n}$  désigne la charge au noeud  $n$  à l'instant  $t$ , et  $\bar{p}_{n,t}$  la puissance maximale disponible des générateurs connectés à ce même noeud à l'instant  $t$ . Ces bornes  $\check{x}_n$  et  $\hat{x}_n$  encadrent l'ensemble des injections nodales possibles sur l'horizon, et permettent de restreindre l'analyse aux configurations effectivement réalisables.

Pour préserver la symétrie du polytope de faisabilité (centré en  $\mathbf{x} = \mathbf{0}$ ), nous définissons ensuite une borne symétrique sur chaque injection nodale :

$$\bar{x}_n = \max \{|\check{x}_n|, |\hat{x}_n|\}. \quad (2.45)$$

L'analyse de redondance est alors restreinte à la région admissible :

$$\mathcal{X} = \left\{ \mathbf{x} \in \mathbb{R}^{|\mathcal{N}|} \mid -\bar{x}_n \leq x_n \leq \bar{x}_n, \forall n \in \mathcal{N} \right\}.$$

ce qui permet de retirer des contraintes jugées essentielles dans le cas général, mais inactives dans l'espace effectivement réalisable.

Enfin, un mécanisme de filtrage par impact maximal permet de rejeter les contingences dont l'effet sur une contrainte est jugé négligeable. Soit  $p_{ij}^{(t,c)}$  l'écoulement de puissance active sur la ligne  $(i,j) \in \mathcal{L}$  à l'instant  $t$ , sous la contingence  $c \in \mathcal{C}$ . Les auteurs établissent une borne a priori fondée sur les LODF :

$$\left| p_{ij}^{(t,c)} - p_{ij}^{(t,0)} \right| \leq \left| \Lambda_{(ij),o} p_o^{(t,0)} \right| \leq \left| \Lambda_{(ij),o} \right| \bar{s}_o, \quad (2.46)$$

où  $o \in \mathcal{L}$  désigne la ligne en défaut (*outage*), et  $\bar{s}_o$  sa limite thermique.

Si cette variation estimée reste inférieure à un seuil de sécurité  $\eta > 0$ , alors la contrainte associée est ignorée dans le processus de sélection :

$$\left| \Lambda_{(ij),o} \right| \bar{s}_o < \eta. \quad (2.47)$$

Ce filtrage réduit le nombre de contraintes testées, sans compromettre la couverture des cas critiques à condition que  $\eta$  soit choisi de manière conservatrice.

L'approche hybride, combinant le test de redondance (2.44), le *rayshoot* de l'Algorithme 2, les bornes symétriques (2.45), et le filtrage par impact maximal fondé sur les LODF (2.47), présentée dans [12], permet de réduire la taille du problème SCOPF tout en préservant l'optimalité. Sur le cas test IEEE 118 noeuds, les auteurs rapportent une réduction initiale

de 87 % du nombre de contraintes grâce au filtrage par impact avec une marge de sécurité  $\eta = 5 \%$ , ce qui diminue le temps de résolution de 81 %. En appliquant ensuite la procédure de réduction de contraintes, une réduction supplémentaire de 41 % est obtenue, puis jusqu'à 98 % au total lorsque le test de redondance conditionnelle est activé. La solution optimale du SCOPF est conservée, à l'exception d'une légère variation (environ 3 %) lorsque la marge  $\eta$  est interprétée comme une réduction effective de la capacité des lignes.

De plus, au-delà des cas statiques, la méthode s'étend naturellement aux problèmes multipériodes. Les auteurs l'appliquent à un cas pratique de marché nodal multi-horaire pour le réseau allemand, incluant des contraintes temporelles sur le stockage d'énergie. Sur ce problème de grande taille, le SCOPF complet ne peut être résolu directement, faute de mémoire. La procédure de réduction permet de rendre le problème computationnellement traitable à grande échelle. En moins de 200 minutes, elle élimine près de 99,7 % des contraintes initiales, tout en préservant la faisabilité globale. Une fois cette phase terminée, le SCOPF  $N-1$  peut être résolu en quelques secondes, y compris dans un contexte multipériode horaire avec 24 pas de temps. Dans cette configuration, le nombre total de paires ligne-contingence critiques (*critical branch-contingency-outage pairs*) actives sur l'horizon temporel complet reste inférieur à 30, contre plusieurs centaines de milliers dans la formulation initiale non réduite.

Cette méthode illustre la puissance des techniques structurelles pour isoler un sous-ensemble minimal de contraintes actives, avec une performance qui s'améliore lorsque le réseau est redondant. L'efficacité computationnelle, la généralisation multipériode, et la garantie de sécurité apportent une réponse pertinente aux limitations des méthodes précédentes.

Toutefois, l'approche déterministe atteint ses limites dès lors que la structure du réseau varie significativement dans le temps, que les contraintes dépendent de paramètres incertains ou que les interactions non linéaires rendent inopérant tout filtrage a priori. Par ailleurs, ces méthodes considèrent implicitement que l'ensemble des contraintes du modèle doit être satisfait strictement, ce qui contredit le fonctionnement réel des marchés opérationnels, où les contraintes d'écoulement de puissance sont généralement modélisées comme souples à l'aide de pénalités dans la fonction objectif [7, 55]. De plus, les approches déterministes ne tirent parti d'aucune information issue des solutions passées, alors même que les opérateurs disposent de vastes historiques d'optimisation réutilisables. C'est dans ce contexte que s'inscrivent les approches fondées sur l'apprentissage supervisé, qui visent à apprendre, à partir de données précédemment observées, des règles efficaces de réduction de contraintes.

### 2.3.3 Réduction de contraintes par apprentissage machine

Les méthodes de réduction de contraintes par apprentissage infèrent dynamiquement une hiérarchie d'importance ou de probabilité d'activation pour chaque contrainte, ajustée aux conditions spécifiques du système. Les sections qui suivent présentent ces stratégies d'apprentissage hiérarchique, organisées selon le type de structure exploitée et le degré d'intégration dans la formulation réduite.

#### 2.3.3.1 Réduction progressive des contraintes par classement ou score de priorité

L'article [20] présente une approche hybride combinant apprentissage profond et filtrage itératif afin d'identifier, en fonction des conditions d'exploitation, un sous-ensemble pertinent de ces contraintes à inclure dans la résolution. L'objectif est de construire une version tronquée du problème de répartition économique dynamique (*dynamic economic dispatch*, D-ED), i.e., une répartition économique sur un horizon multipériode en considérant les contraintes de rampes, contenant uniquement les contraintes jugées nécessaires ou potentiellement critiques, de manière à réduire les coûts de résolution tout en garantissant l'optimalité du plan de production. La formulation complète du problème D-ED (2.48) s'écrit :

$$\min_{\mathbf{p}} \quad \sum_t \sum_u \gamma_{u,t} p_{u,t} \quad (2.48)$$

$$\text{s.c.} \quad \sum_u p_{u,t} = \sum_n d_{n,t} \quad \forall t \quad (2.49)$$

$$\underline{p}_u \leq p_{u,t} \leq \bar{p}_u \quad \forall u, t \quad (2.50)$$

$$p_{u,t} - p_{u,t-1} \leq \bar{r}_u \quad \forall u, t \quad (2.51)$$

$$p_{u,t-1} - p_{u,t} \leq \underline{r}_u \quad \forall u, t \quad (2.52)$$

$$-\bar{s}_{ij} \leq \Phi_{ij}(\mathbf{p}_t^{\text{inj}} - \mathbf{d}_t) \leq \bar{s}_{ij} \quad \forall (i, j), t \quad (2.53)$$

où  $\mathbf{p}$  désigne le vecteur de génération,  $\gamma_{u,t}$  le coût marginal de génération pour l'unité  $u$  à l'instant  $t$ ,  $\bar{r}_u$  et  $\underline{r}_u$  les limites de rampes montantes et descendantes,  $\bar{s}_{ij}$  la limite thermique sur la ligne  $(i, j)$ ,  $\mathbf{p}_t^{\text{inj}}$  les injections nodales nettes, et  $\mathbf{d}_t$  la charge au pas de temps  $t$ .

L'approche proposée repose sur l'apprentissage supervisé du statut de ces contraintes, à partir d'un ensemble d'instances de simulation où les contraintes actives et pseudo-actives sont identifiées par une boucle d'optimisation itérative. Le terme pseudo-active désigne ici les contraintes qui, bien que non saturées dans une solution optimale donnée, doivent être conservées pour garantir l'unicité ou la faisabilité de cette solution dans le cas de générateurs

aux coûts identiques. L'apprentissage est ensuite utilisé pour prédire, à partir des données connues à l'instant de la pré-répartition économique.

Pour encoder chaque scénario d'exploitation, les auteurs construisent une image tridimensionnelle dite charge-coût-topologie (*demand-cost-topology*, DCT), dans laquelle chaque dimension représente une matrice distincte :

- la charge nodale, codée dans une matrice  $\mathbf{D} \in \mathbb{R}^{|\mathcal{N}| \times T}$ , où chaque colonne contient les valeurs de charge pour une période donnée ;
- les coefficients de coût de génération, étendus à l'ensemble des nœuds sous la forme d'une matrice  $\mathbf{\Gamma} \in \mathbb{R}^{|\mathcal{N}| \times T}$ , les entrées nulles correspondant aux nœuds sans générateur ;
- la diagonale de la matrice d'admittance nodale  $\mathbf{Y}_{\text{bus}}$ , codée dans une matrice  $\mathbf{Y} \in \mathbb{R}^{|\mathcal{N}| \times T}$ .

Ces trois matrices sont concaténées pour former un tenseur  $\mathbf{X} \in \mathbb{R}^{|\mathcal{N}| \times T \times 3}$ , dans lequel chaque « canal » correspond à une information physique distincte. Ce tenseur est interprété comme une image tridimensionnelle en fausses couleurs, au sens où les canaux rouge, vert et bleu ne représentent pas des composantes spectrales visuelles, mais des attributs structurés du réseau électrique, i.e., charge, coût et topologie, utilisés pour exploiter la capacité des réseaux convolutifs à extraire des motifs spatiaux corrélés.

Ce tenseur sert d'entrée à un réseau de neurones convolutif pré-entraîné (architecture *EfficientNet-B7*), adapté par transfert d'apprentissage à une tâche de classification binaire. Trois classificateurs distincts sont ainsi construits pour prédire l'activation des contraintes : un pour les écoulements de puissance sur les lignes de transport, un pour les rampes montantes, et un pour les rampes descendantes.

Afin de corriger les éventuelles erreurs de classification, une boucle de validation est intégrée après la phase de prédiction. Lorsque la résolution du modèle tronqué révèle l'activation de contraintes absentes dans l'ensemble initial, celles-ci sont ajoutées, et le problème est résolu à nouveau. Ce processus est répété jusqu'à ce qu'aucune nouvelle contrainte active ou pseudo-active ne soit identifiée. Cette procédure garantit que l'ensemble final de contraintes retenues permet de reconstituer une solution exacte, i.e., identique à celle obtenue à partir du problème complet.

Les résultats empiriques montrent des gains substantiels en mémoire et en temps de résolution, notamment pour les grands réseaux. Sur un cas réel à 6515 nœuds, le temps de résolution est divisé par un facteur 100, et l'utilisation mémoire est réduite d'un facteur supérieur à 200. L'intégralité de l'espace de faisabilité est conservée, et l'écart sur la fonction objectif est négligeable (inférieur à  $10^{-7}$ ). L'approche proposée surpassé également les méthodes de génération itérative classiques, en réduisant significativement le nombre d'itérations nécessaires

à la convergence.

Ce travail illustre la puissance des méthodes d'apprentissage pour orchestrer de manière efficace la sélection des contraintes pertinentes dans des modèles de grande dimension, tout en intégrant des garanties de faisabilité via des boucles correctives. Il ouvre des perspectives pour l'intégration de techniques d'analyse d'images à base de réseaux convolutifs dans les chaînes de traitement des modèles SCED multipériodes à grande échelle.

### 2.3.3.2 Hiérarchies implicites guidées par structure réseau ou topologie

Dans un contexte où la fiabilité et la résilience du réseau électrique devient une exigence centrale face à l'augmentation des aléas climatiques et à la croissance des charges distribuées, le problème du SCOPF avec contingences multiples  $N-k$  devient incontournable. Les méthodes conventionnelles présentées à la section 2.1 atteignent rapidement leurs limites lorsque le nombre de scénarios de contingence croît. En particulier, leur complexité combinatoire rend leur exécution difficilement compatible avec des applications nécessitant une résolution répétée à intervalles rapprochés, comme dans les environnements temps réel ou les procédures de pré-répartition. L'article [19] propose une alternative fondée sur une approche dite guidée par les contraintes, dans laquelle la structure du réseau, encodée via les LODF, est utilisée pour piloter un apprentissage profond visant à approximer efficacement la solution du SCOPF avec contingences multiples. L'approche proposée repose sur une architecture à trois étapes successives : une prédiction initiale des puissances actives injectées par les générateurs, une correction par projection pour satisfaire les contraintes physiques à l'état de base, et une évaluation des dépassements de limite post-contingence pour l'ensemble des scénarios  $N-k$ .

La première étape consiste à prédire, à partir d'un profil de charge  $\mathbf{p}_d \in \mathbb{R}^{|\mathcal{N}|}$ , un vecteur de génération  $\hat{\mathbf{p}}_g \in \mathbb{R}^{|\mathcal{G}|}$  respectant les bornes de chaque générateur. Le réseau de neurones  $\Phi_{\mathbf{W}}$  paramétré par des poids  $\mathbf{W}$  produit un vecteur de coefficients de mise à l'échelle  $\boldsymbol{\alpha}_g \in [0, 1]^{|\mathcal{G}|}$ , permettant d'encoder explicitement les contraintes de boîte :

$$\boldsymbol{\alpha}_g = \Phi_{\mathbf{W}}(\mathbf{p}_d) \quad (2.54)$$

$$\hat{\mathbf{p}}_g = \boldsymbol{\alpha}_g \circ (\bar{\mathbf{p}}_g - \underline{\mathbf{p}}_g) + \underline{\mathbf{p}}_g \quad (2.55)$$

où  $\bar{\mathbf{p}}_g, \underline{\mathbf{p}}_g \in \mathbb{R}^{|\mathcal{G}|}$  désignent respectivement les puissances actives maximales et minimales admissibles. Le produit  $\circ$  est à comprendre comme un produit composante par composante.

Cette prédiction, bien que conforme aux limites de génération, ne garantit ni l'équilibre des injections, ni le respect des limites thermiques sur les lignes. Pour restaurer la faisabilité

vis-à-vis du modèle DC, une projection est introduite. Elle ajuste la prédiction  $\hat{\mathbf{p}}_g$  vers un point réalisable  $\mathbf{p}_g \in \mathbb{R}^{|\mathcal{G}|}$ , solution d'un DCOPF classique :

$$\min_{\mathbf{p}_g, \boldsymbol{\theta}} \|\mathbf{p}_g - \hat{\mathbf{p}}_g\|_2^2 \quad (2.56a)$$

$$\text{s.c. } \mathbf{B}_{\text{bus}} \boldsymbol{\theta} = \mathbf{p}_g - \mathbf{p}_d \quad (2.56b)$$

$$-\bar{s}_{ij} \leq b_{ij}(\theta_i - \theta_j) \leq \bar{s}_{ij} \quad \forall (i, j) \in \mathcal{L} \quad (2.56c)$$

$$\underline{p}_{gi} \leq p_{Gi} \leq \bar{p}_{gi} \quad \forall i \in \mathcal{G} \quad (2.56d)$$

où :

- $\boldsymbol{\theta} \in \mathbb{R}^{|\mathcal{N}|}$  désigne le vecteur des angles de phase ;
- $\mathbf{B}_{\text{bus}} \in \mathbb{R}^{|\mathcal{N}| \times |\mathcal{N}|}$  est la matrice de susceptance nodale ;
- $b_{ij}$  est la susceptance de la ligne  $(i, j) \in \mathcal{L}$  ;
- $\bar{s}_{ij} \in \mathbb{R}_+$  est la limite thermique de la ligne  $(i, j)$ .

Cette étape est nécessaire pour corriger les déséquilibres ou non-respects des limites introduits par la prédiction. Le problème (2.56) est implémentée comme une couche différentiable via CVXPYLayer [71], permettant une rétropropagation du gradient à travers la projection.

Une fois la solution projetée  $\mathbf{p}_g$  obtenue, les écoulements post-contingence sont évalués sur l'ensemble des scénarios  $c \in \mathcal{C}$  correspondant aux combinaisons de  $k$  lignes mises en défaut (analyse  $N-k$ ). Pour chaque scénario  $c$ , les écoulements  $\mathbf{p}^{(c)} \in \mathbb{R}^{|\mathcal{L}|}$  sont évalués efficacement à l'aide de la matrice de sensibilité  $\Lambda^{(c)}$  :

$$\mathbf{p}^{(c)} = \mathbf{p}^0 + \Lambda^{(c)} \mathbf{p}^0, \quad (2.57)$$

où  $\mathbf{p}^0$  désigne les écoulements à l'état de base. La structure creuse des matrices  $\Lambda^{(c)}$  permet une évaluation rapide et vectorisée sur un grand nombre de scénarios.

Le modèle est entraîné en minimisant une fonction de perte pénalisant explicitement les violations de contraintes à l'état nominal et post-contingence, ainsi que les déséquilibres globaux de production :

$$\begin{aligned} \mathcal{L} = & \lambda_3 \mathbf{c}_g^\top \mathbf{p}_g + \lambda_0 \left\| \left[ |\mathbf{p}^0| - \bar{\mathbf{s}} \right]_+ \right\|_1 \\ & + \lambda_1 \left\| \left[ |\mathbf{p}^{(c)}| - \bar{\mathbf{s}} \right]_+ \right\|_1 + \lambda_2 \left| \mathbf{1}^\top \hat{\mathbf{p}}_g - \mathbf{1}^\top \mathbf{p}_d \right| \end{aligned} \quad (2.58)$$

où :

- $\mathbf{c}_g \in \mathbb{R}^{|\mathcal{G}|}$  désigne le vecteur des coûts marginaux des générateurs ;
- $\bar{\mathbf{s}} \in \mathbb{R}^{|\mathcal{L}|}$  est le vecteur des limites thermiques des lignes ;
- $\lambda_0, \lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}_+$  sont des coefficients de pondération des différentes composantes de la perte.

Le premier terme dans (2.58), pondéré par  $\lambda_3$ , correspond au coût total de production associé au vecteur  $\mathbf{p}_g$  après restauration de la faisabilité. Le second, pondéré par  $\lambda_0$ , pénalise les dépassements des limites thermiques à l'état nominal. Le troisième, associé à  $\lambda_1$ , cumule les violations post-contingence pour un scénario donné  $c \in \mathcal{C}$ . Enfin, le terme pondéré par  $\lambda_2$  quantifie le déséquilibre global entre la génération prévue initialement et la charge totale, avant projection.

Grâce à cette formulation dite guidée par les contraintes, l'apprentissage est réalisé sans données étiquetées : le modèle apprend directement à respecter les contraintes opérationnelles du SCOPF, à l'aide de la rétropropagation à travers une structure différentiable. Le graphe computationnel est allégé en ne conservant, pour la rétropropagation, que les scénarios  $c$  présentant un non-respect des limites, ce qui permet une accélération significative du processus sur les grands réseaux.

L'approche est ensuite étendue à un cadre probabiliste dans lequel les contingences sont pondérées par leur probabilité d'occurrence, estimée via une copule gaussienne [72] modélisant les corrélations spatiales (ex. : événements météorologiques extrêmes). La fonction de perte est adaptée en conséquence pour minimiser les dépassements de limite attendus en post-contingence.

Les résultats empiriques sur les réseaux IEEE à 39 et 118 noeuds indiquent une réduction significative du temps de calcul (jusqu'à un facteur 165), ainsi qu'une diminution de la mémoire requise par un facteur 100. Par ailleurs, le modèle conserve de bonnes performances face à des événements rares de type haut impact et probabilité faible (*high-impact low-probability*, HILP), comme simulé dans un scénario de tremblement de terre, en maintenant des niveaux de fiabilité supérieurs aux approches classiques tout en limitant l'augmentation du coût de répartition.

Ce travail illustre un usage très pertinent des structures topologiques du réseau, encodées dans la matrice LODF, pour piloter un apprentissage supervisé capable de capturer implicitement une hiérarchie des contraintes pertinentes. Il ouvre des perspectives prometteuses pour le traitement efficace des formulations SCOPF à grande échelle, incluant des contingences multiples avec prise en compte du risque.

### 2.3.3.3 Réduction paramétrée ou pilotée par représentation latente

Dans la continuité des approches structurelles guidées par les contraintes physiques du réseau, une nouvelle classe de méthodes propose de construire une représentation latente du réseau électrique via des réseaux de neurones à graphes hiérarchiques. L'article [73] s'inscrit dans cette perspective, en introduisant une stratégie de réduction conjointe des variables et des contraintes fondée sur un modèle adaptatif à deux étages. L'objectif est de construire, à partir d'un profil de charge  $\mathbf{p}_d \in \mathbb{R}^{|\mathcal{N}|}$ , une approximation réduite du SCED, conservant les caractéristiques critiques de la solution optimale.

Le cadre de référence est un SCED résolu initialement dans sa formulation complète, incluant l'ensemble des contraintes de génération et des contraintes thermiques sur les lignes. Cette formulation est ensuite comparée à trois variantes réduites :

- ROPFL : modèle réduit aux lignes critiques uniquement (*reduced OPF with line constraints*) ;
- ROPFG : modèle réduit aux générateurs non saturés uniquement (*reduced OPF with generator constraints*) ;
- ROPFLG : modèle réduit combiné, à la fois sur les lignes et les générateurs.

La construction de ces modèles repose sur deux modules d'apprentissage séquentiels :

1. un premier réseau de neurones à graphes prédit, à partir de la charge nodale, l'ensemble des lignes  $\mathcal{A} \subset \mathcal{L}$  susceptibles d'atteindre leur limite thermique, c'est-à-dire les lignes critiques pour lesquelles la contrainte  $|p_{ij}| \leq \bar{s}_{ij}$  risque d'être active ;
2. un second réseau de neurones à graphes, conditionné sur les lignes prédictes comme critiques, prédit les générateurs  $i \in \mathcal{G}$  qui atteindront leur limite supérieure  $\bar{p}_i$ .

Ces prédictions permettent de reformuler une version réduite du problème, dans laquelle :

- seules les contraintes d'écoulement de puissance (2.56c) sur les lignes  $(i, j) \in \mathcal{A}$  sont conservées ;
- les variables  $p_i$  des générateurs prédicts comme saturés sont supprimées du problème et remplacées par leur valeur fixée  $\bar{p}_i$ , agissant comme paramètres constants.

Une caractéristique centrale de cette approche réside dans le mécanisme dit de duplication virtuelle des générateurs, conçu pour augmenter la capacité expressive des réseaux de neurones à graphes (*graph neural networks*, GNN) appliqués aux modèles de SCED. Concrètement, chaque générateur  $i \in \mathcal{G}$  est associé à un nœud virtuel  $i'$  ajouté au graphe, relié par une arête au nœud physique auquel il est connecté électriquement. Cette opération transforme le graphe initial de taille  $|\mathcal{N}|$  en un graphe étendu de taille  $|\mathcal{N}| + |\mathcal{G}|$ , dans lequel les nœuds physiques (charge, injection, topologie) et les nœuds générateurs sont traités distinctement. Ce graphe

élargi sert de support à la propagation de messages (*message passing*) dans les architectures GNN, qui agrègent l'information locale à partir des voisins connectés. Grâce à la séparation explicite des générateurs, le GNN peut encoder de manière ciblée des attributs propres à chaque unité de production, tels que le coût marginal  $\gamma_i$ , les rampes de montée et de descente, ou les réserves assignées, sans les confondre avec les caractéristiques des noeuds du réseau. Cette spécialisation structurelle améliore la capacité du GNN à extraire des représentations latentes différencierées, facilitant ainsi la prédiction de la criticité des contraintes associées. La matrice d'adjacence du graphe passe alors d'une taille  $|\mathcal{N}|^2$  à  $(|\mathcal{N}| + |\mathcal{G}|)^2$ , impliquant un surcoût marginal en calcul, mais une augmentation significative de la granularité de l'encodage topologique et fonctionnel. Le GNN opère ensuite sur ce graphe enrichi, en apprenant à projeter chaque noeud (physique ou virtuel) dans un espace latent, à partir duquel la criticité des lignes ou des générateurs est prédite via des couches de sortie spécialisées.

L'apprentissage est effectué hors ligne sur un ensemble de 10 000 scénarios de charge générés aléatoirement autour d'un profil de référence. Chaque scénario est résolu avec le SCED, et les lignes sont étiquetées comme critiques si leur chargement excède un seuil fixé (ex. 70 % de  $\bar{s}_{ij}$ ). La seconde phase d'apprentissage utilise les sorties du premier GNN comme entrée pour guider la prédiction des générateurs saturés.

Les deux modules sont entraînés en minimisant une fonction de classification binaire, utilisant une entropie croisée pondérée pour corriger les déséquilibres de classe. Soit  $\hat{y}_\ell \in [0, 1]$  la probabilité prédite d'activation pour la contrainte  $\ell \in \mathcal{L}$ , et  $y_\ell \in \{0, 1\}$  l'étiquette issue du SCED. La fonction de perte est donnée par :

$$\mathfrak{L}_{\text{bin}} = - \sum_{\ell} \beta_+ y_\ell \log(\hat{y}_\ell) + \beta_- (1 - y_\ell) \log(1 - \hat{y}_\ell)$$

où  $\beta_+$  et  $\beta_-$  sont des coefficients de pondération inverses des fréquences de classes.

En phase opérationnelle, le processus est le suivant : à partir d'un nouveau profil  $\mathbf{p}_d$ , le premier GNN prédit les lignes critiques, le second identifie les générateurs susceptibles d'être contraints à leur limite, et une formulation réduite du problème ROPFLG est construite à la volée en activant uniquement les contraintes prédites comme pertinentes, avant d'être résolue numériquement. Ceci permet d'accélérer la résolution sans compromettre la qualité de la solution.

Les expériences menées sur un réseau test à 73 noeuds montrent que la méthode complète ROPFLG permet une réduction moyenne de 32 % du temps de résolution par rapport au modèle complet, avec un écart relatif empirique sur le coût inférieur à 0,1 %. Le taux d'erreur total sur la prédiction des lignes critiques est inférieur à 1,2 %, les erreurs de type I et II

restant marginales.

Ce travail illustre la possibilité de projeter une formulation SCED vers un espace latent réduit, en exploitant conjointement la topologie du réseau, les caractéristiques des noeuds, et la hiérarchie entre contraintes. Le mécanisme de duplication virtuelle augmente la résolution du GNN sur les attributs spécifiques aux générateurs, tandis que la structure hiérarchique permet une prédiction multi-étage contrôlée. L'approche ouvre la voie à des implémentations temps réel du SCED ou du SCOPF, notamment pour des réseaux à grande échelle.

## 2.4 Synthèse critique et lien avec la méthode proposée

Les approches analysées dans la littérature témoignent d'efforts pour diminuer le temps de résolution des problèmes de SCED, en particulier dans les cas à scénario de contingences multiples. Trois grandes familles se distinguent : d'une part les méthodes de décomposition structurelle et de parallélisation, exploitant le découplage entre le régime nominal et les états post-contingence, les symétries topologiques, ou la séparabilité temporelle du problème, et d'autre part les techniques d'apprentissage machine pour prédire les contraintes actives à l'optimalité, et finalement les méthodes de réduction de contraintes, basées sur des tests structurels ou sur des prédictions guidées par apprentissage automatique.

Les méthodes de décomposition de type Benders, largement représentées dans la littérature [25, 26], reposent sur la séparation entre les variables de contrôle partagées et les sous-problèmes post-contingence. Cette structure autorise une parallélisation des scénarios et une progression incrémentale via l'ajout de coupes, mais suppose l'exactitude des contraintes de contingence et l'absence de tolérance aux dépassements. Certaines extensions, comme l'intégration explicite de dispositifs FACTS [25], ou la modélisation hybride linéaire et non linéaire selon les scénarios [27], visent à mieux refléter la réalité physique, mais complexifient le schéma de résolution et demeurent incompatibles avec des formulations à contraintes souples.

Les approches de type ADMM [30], en répartissant le calcul entre scénarios tout en assurant la cohérence globale via des multiplicateurs de Lagrange, offrent une voie de mise à l'échelle pour le traitement parallèle du SCOPF. Toutefois, leur efficacité dépend fortement de l'architecture de calcul distribuée disponible, et elles n'exploitent pas explicitement la redondance des contraintes pour alléger la dimension du problème global.

Plus récemment, des méthodes de mise à jour analytique, telles que la mise à jour efficiente de l'inverse par le lemme de Sherman–Morrison–Woodbury [35], ont été proposées pour accélérer la résolution incrémentale du SCED lors de l'ajout progressif de scénarios de contingence. Bien que performantes, ces stratégies ne relèvent pas à proprement parler de la réduction structurelle

de contraintes, et ne sont pas directement transposables aux formulations pénalisées.

Toutes ces approches, bien que performantes, s'appuient sur l'hypothèse implicite que, dans la formulation linéaire du SCED, l'ensemble des contraintes post-contingence est strictement nécessaire à la caractérisation du polytope réalisable. Or, cette hypothèse est contredite empiriquement dans plusieurs études [12, 66, 68], qui montrent que la majorité des contraintes sont structurellement redondantes. En effet, dans un réseau électrique à grande échelle, la majorité des scénarios de contingence induisent des écarts négligeables sur les écoulements de puissance, et leurs contraintes associées ne sont jamais saturées aux optima. Cependant, ces travaux s'inscrivent dans un cadre où les contraintes sont supposées strictes, sans prise en compte de la possibilité de dépassement introduite dans les formulations pénalisées du SCED, réellement utilisée par les marchés. Cette hypothèse limite leur applicabilité dans les contextes de marché réels, où la faisabilité est assurée non par satisfaction stricte, mais par pénalisation progressive des dépassements de contraintes dans la fonction objectif.

La quasi-totalité des méthodes existantes reposent sur une modélisation à contraintes dures, incompatible avec la formulation pénalisée du SCED utilisée dans les environnements de marché réels [55]. Dans cette dernière, les dépassements des contraintes thermiques post-contingence sont autorisés moyennant une pénalisation explicite dans la fonction objectif (1.20a), assurant ainsi que le problème reste réalisable pour toute configuration de charge et d'état du réseau, sans zone d'infaisabilité, même en présence de conflits entre exigences de sécurité et optimum économique. Nous adoptons ce même cadre de contraintes souples que celui mis en œuvre par des opérateurs tels que MISO et ERCOT, tout en recherchant une exactitude de la solution dans le processus de réduction, de manière à préserver la cohérence entre faisabilité opérationnelle et optimalité économique.

Plusieurs contributions récentes ont exploré la réduction de contraintes dans le SCED à l'aide de méthodes d'apprentissage automatique, mais la majorité d'entre elles s'appuie sur des prédictions binaires isolées, sans cohérence globale ni mécanisme formel de correction. L'approche de [20] classe chaque contrainte comme active ou inactive à partir d'une représentation image du réseau, via un réseau convolutif entraîné en supervision. Cette méthode permet une réduction efficace dans un contexte multi-période, mais repose sur une classification contrainte par contrainte sans structure hiérarchique, ce qui empêche toute garantie sur la précision globale de la solution obtenue. Dans le même esprit, l'approche non supervisée de [19], centrée sur des critères probabilistes de sécurité, permet une estimation du sous-ensemble actif, mais ne prévoit aucun mécanisme de repli en cas de sous-estimation et ne propose pas de contrôle explicite sur le niveau de conservatisme retenu. L'article [73], quant à lui, adopte une stratégie hiérarchique fondée sur des GNN pour identifier à la fois les contraintes d'écoulement

de puissance actives et les générateurs saturés, mais la réduction induite s'appuie sur une reconstruction manuelle du problème, dont la faisabilité peut être remise en cause par des erreurs locales.

Face aux limites identifiées dans les approches traditionnelles, les travaux présentés dans ce mémoire introduisent une nouvelle méthode fondée sur deux innovations conceptuelles :

- L'introduction explicite d'une hiérarchie de contraintes, sélectionnée pour assurer le maintien de la solution optimale du modèle, et ;
- la prédition dynamique du niveau hiérarchique pertinent à l'aide d'un modèle d'apprentissage supervisé.

Cette approche prolonge les principes de sélection géométrique développés dans [12], au cadre du SCED à contraintes souples, en les adaptant à la formulation pénalisée (1.20f).

La méthode s'appuie en amont sur une procédure itérative de suppression des contraintes redondantes, appliquée à l'ensemble des limites d'écoulement de puissance optimal. Cette démarche conduit à l'élaboration progressive d'une hiérarchie ordonnée de sous-ensembles de contraintes, chacun représentant une relaxation décroissante de l'ensemble admissible du problème original. À chaque niveau est associé un modèle réduit, plus léger sur le plan computationnel, et dont la solution est, pour un niveau optimal, exactement égale à celle du problème complet. Le problème devient ainsi une tâche de classification cumulative hiérarchique, où nous cherchons, à partir d'un profil de charge nodale  $\mathbf{p}_d \in \mathbb{R}^{|\mathcal{N}|}$ , à prédire le plus petit niveau hiérarchique assurant l'exactitude de la solution réduite, tout en contrôlant la tolérance au risque de sous-estimation par un seuil de décision ajustable. Contrairement aux approches par classification binaire contrainte-par-contrainte [20, 51, 53], qui ne capturent ni les interdépendances structurelles ni les propriétés d'imbrication des régions réalisables, notre stratégie exploite la structure hiérarchique pour garantir la cohérence de la prédition, en assurant que la validité d'un niveau implique celle de tous les niveaux inférieurs. Par la suite, une correction incrémentale est possible en cas d'erreur afin de garantir la faisabilité de la solution obtenue par notre méthode.

Par sa capacité à articuler sélection géométrique, robustesse computationnelle et adaptabilité par apprentissage, cette méthode constitue une contribution nouvelle au traitement du SCED avec pénalité. Elle offre un compromis entre précision et rapidité, aligné avec les réalités industrielles de planification sous incertitude, et fournit une base rigoureuse pour la réduction adaptative de contraintes dans des formulations massivement dimensionnées.

## CHAPITRE 3 MÉTHODE DE RÉDUCTION DE CONTRAINTE GUIDÉE PAR L'APPRENTISSAGE SUPERVISÉ

Les méthodes de réduction de contraintes présentées à la section 2.3.2 permettent de diminuer significativement la complexité des problèmes de SCED et de SCOPF, tout en préservant l'exactitude des solutions obtenues. Toutefois, ces approches reposent sur l'hypothèse que le problème est formulé avec des contraintes strictes, dont la faisabilité doit être strictement respectée. Or, dans le contexte opérationnel des marchés de l'électricité, tel que décrit à la section 1.2.4, les formulations pratiques s'appuient majoritairement sur des contraintes souples, pénalisées dans la fonction objectif [6, 7]. Cette spécificité limite l'applicabilité directe des techniques classiques de détection de redondance structurelle.

Le présent chapitre introduit une méthode de réduction de contraintes spécifiquement adaptée à la formulation pénalisée du SCED sous contraintes de sécurité  $N-1$  et  $N-k$ . Cette méthode constitue une généralisation de l'approche proposée par l'article [12], étendue ici aux formulations à contraintes souples, et enrichie par l'intégration d'un modèle d'apprentissage supervisé. Celui-ci permet de prédire, pour un profil de charge donné, le niveau minimal de contraintes à activer pour garantir l'exactitude de la solution, tout en limitant le nombre de contraintes effectivement injectées dans le solveur.

Dans un premier temps la section 3.1 présente la méthode dans sa globalité, en détaillant ses principales composantes structurelles et ses objectifs. La section 3.2 introduit formellement les outils mathématiques mobilisés, notamment l'algorithme de construction hiérarchique des contraintes et le modèle de classification supervisée fondé sur un algorithme d'arbres à gradients légers, connu sous le nom *Light Gradient Boosted Machine* (LightGBM). La chaîne complète de traitement, incluant la génération de données, l'apprentissage et le déploiement, est ensuite exposée dans la section 3.3 avant de conclure sur les propriétés de l'approche développée.

### 3.1 Présentation générale de l'approche

La formulation pénalisée du SCED (1.20), introduite à la section 1.2.4, permet de garantir la faisabilité du problème pour toute injection  $\mathbf{p} \in \mathbb{R}^{|\mathcal{N}|}$  respectant les contraintes sur les limites d'injections (1.20c), en autorisant un dépassement contrôlé des contraintes thermiques post-contingence via l'introduction explicite de variables de pénalité  $\xi_{ij}^{(c)}$ . Cette propriété de faisabilité généralisée constitue un avantage opérationnel majeur dans les contextes de

planification de marché. Toutefois, elle induit une croissance rapide de la taille du problème d'optimisation, accentuée par la densité des matrices  $\Phi^{(c)}$  associées à chaque scénario de contingence  $c \in \mathcal{C}$ .

Comme discuté à la section 1.3, cette croissance dimensionnelle engendre une complexité computationnelle importante et problématique dans les contextes de résolution fréquente et multipériode rencontrés dans les marchés de l'électricité. Plus fondamentalement, la densité structurelle des contraintes affecte défavorablement la performance des solveurs, en altérant la structure creuse des systèmes KKT exploités dans les méthodes de points intérieurs, et en allongeant les temps de présolution [22].

Dans ce contexte, la méthode proposée réduit le nombre de contraintes considérées lors de la résolution du SCED, tout en préservant l'exactitude de la solution, et ce, dans le contexte des contraintes souples telles qu'utilisées par le modèle (1.18). L'objectif n'est donc pas de sécuriser la résolution, mais d'identifier, pour chaque instance, le plus petit sous-ensemble de contraintes post-contingence à inclure explicitement afin d'améliorer la résolution du problème. À cette fin, nous construisons une hiérarchie d'ensembles de contraintes, décrite à la section 3.2.1, à partir des matrices  $\Phi^{(c)}$  associées à l'ensemble des scénarios de contingence  $c \in \mathcal{C}$ . Cette construction, effectuée une seule fois en amont, repose sur l'application récursive d'un algorithme de suppression de redondance.

Soit  $\mathcal{I} \subset \mathbb{N}$  l'ensemble des indices associés aux contraintes de limite d'écoulement de puissance, engendrées par les scénarios de contingence  $\mathcal{C} \cup \{0\}$ , incluant le cas nominal. On introduit une collection d'indices  $k \in \{1, 2, \dots, k_{\max}\}$  correspondant aux niveaux hiérarchiques successifs extraits du système. À chaque itération, un sous-ensemble minimal de contraintes non redondantes est identifié dans le résidu laissé par les niveaux précédents. Il en résulte une collection ordonnée de sous-ensembles disjoints  $\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_{k_{\max}}$  telle que

$$\bigcup_{k=1}^{k_{\max}} \mathcal{I}_k = \mathcal{I},$$

dans laquelle chaque  $\mathcal{I}_k$  regroupe les contraintes structurellement nécessaires au niveau  $k$ , relativement aux contraintes déjà retenues aux niveaux inférieurs. Pour tout entier  $k \in \{1, 2, \dots, k_{\max}\}$ , introduisons la notation cumulative

$$\mathcal{I}_{1:k} := \bigcup_{j=1}^k \mathcal{I}_j,$$

et définissons l'ensemble admissible  $\mathcal{F}_k \subset \mathbb{R}^{|\mathcal{N}|}$  comme l'intersection des demi-espaces induits par les contraintes de  $\mathcal{I}_{1:k}$ . Par construction, ces ensembles vérifient la relation d'imbrication

stricte

$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_{k_{\max}} = \mathcal{F},$$

où  $\mathcal{F}$  désigne la région admissible définie par l'ensemble complet des contraintes thermiques post-contingence, i.e., (1.20e) et (1.20d).

Cette stricte inclusion découle du fait que les ensembles  $\mathcal{I}_k$  sont disjoints et construits après suppression systématique des doublons : aucune contrainte n'est redondante ni dupliquée entre niveaux, ce qui garantit que chaque ajout modifie effectivement le domaine admissible.

Sur cette base, un modèle d'apprentissage supervisé est entraîné pour associer à chaque profil de charge  $\mathbf{p}_d$  un niveau  $k^* \in \{1, \dots, k_{\max}\}$  minimal tel que la solution du SCED restreint aux seules contraintes de  $\mathcal{I}_{1:k^*}$  coïncide exactement avec celle obtenue sur le modèle complet.

La tâche est reformulée comme une suite de classifications binaires indépendantes, tel que, pour chaque niveau  $k$ , un classifieur  $\mathfrak{C}_k : \mathbb{R}^{|\mathcal{N}|} \rightarrow \{0, 1\}$  est entraîné à estimer si les contraintes du niveau  $k$  sont potentiellement nécessaires. En phase d'inférence, les sorties binarisées de ces classificateurs sont agrégées pour déterminer le niveau maximal jugé pertinent, noté  $\hat{k}^*$ , utilisé ensuite pour résoudre un SCED restreint aux contraintes des niveaux 1 à  $\hat{k}^*$ .

Cette résolution partielle permet de réduire optimalement la taille du problème, tant du point de vue du nombre d'inégalités que de la structure des systèmes KKT. Afin de garantir la robustesse de la méthode, un mécanisme de reprise incrémentale est mis en œuvre dès qu'un non respect de contraintes est détectée, le niveau hiérarchique étant alors augmenté et la résolution relancée jusqu'à obtention d'une solution exacte. La nature imbriquée de la hiérarchie assure que ce processus converge en un nombre de pas borné par  $k_{\max}$ , tout en maintenant un compromis efficace entre performance algorithmique et fidélité opérationnelle.

### 3.2 Outils de construction hiérarchique et de prédiction

Cette section introduit les deux composantes structurelles de la méthode. Nous abordons dans un premier temps, la construction hiérarchique des contraintes non redondantes à la section 3.2.1, et dans un second temps, le modèle de prédiction supervisée fondé sur LightGBM, décrit à section 4.3.

#### 3.2.1 Hiérarchie de contraintes et espaces de faisabilité imbriqués

Soit  $\mathcal{N}$  l'ensemble des noeuds du réseau électrique et  $\mathcal{L} \subset \mathcal{N} \times \mathcal{N}$  l'ensemble des lignes du réseau électrique. Soit  $\mathcal{C}$  l'ensemble des scénarios de contingence considérés. Nous associons à chaque scénario  $c \in \mathcal{C} \cup \{0\}$ , incluant le cas nominal  $c = 0$ , une matrice PTDF,  $\Phi^{(c)} \in \mathbb{R}^{|\mathcal{L}| \times |\mathcal{N}|}$ ,

représentant la sensibilité linéaire des écoulements de puissance sur les lignes sous le scénario  $c$ , en fonction des injections nodales  $\mathbf{p} \in \mathbb{R}^{|\mathcal{N}|}$ . Chaque ligne  $\ell \in \mathcal{L}$  est caractérisée par une limite thermique  $\bar{\mathbf{s}}_\ell^{(c)} > 0$ . À chaque ligne  $\ell \in \mathcal{L}$  et chaque scénario  $c \in \mathcal{C} \cup \{0\}$ , nous associons deux contraintes linéaires, issues de la contrainte de boîte suivante :

$$-\bar{\mathbf{s}}_\ell^{(c)} \leq \Phi_\ell^{(c)} \mathbf{p} \leq \bar{\mathbf{s}}_\ell^{(c)},$$

représentant les limites inférieure et supérieure sur l'écoulement de puissance apparente à travers la ligne  $\ell$  sous le scénario  $c$ , qui, dans le cadre de l'approximation linéaire de l'OPF, sont des limites sur les puissances actives. L'ensemble total des contraintes thermiques post-contingence est ainsi indexé par un ensemble global  $\mathcal{I}$ , de cardinalité  $|\mathcal{I}| = 2|\mathcal{L}||\mathcal{C}|$ . Dans le cas particulier du critère de sécurité  $N-1$ , où chaque scénario correspond à l'indisponibilité d'une ligne unique, nous avons  $|\mathcal{C}| = |\mathcal{L}|$ , d'où  $|\mathcal{I}| = 2|\mathcal{L}|(|\mathcal{L}| + 1)$ . Ceci souligne le caractère dense et surdimensionné de la description explicite de la région réalisable.

Nous pouvons alors définir l'ensemble admissible  $\mathcal{F} \subset \mathbb{R}^{|\mathcal{N}|}$  comme l'intersection des demi-espaces définis par les inégalités indexées par  $\mathcal{I}$  :

$$\mathcal{F}(\Phi, \bar{\mathbf{s}}, \mathcal{I}) := \left\{ \mathbf{p} \in \mathbb{R}^{|\mathcal{N}|} \mid \Phi_i \mathbf{p} \leq \bar{\mathbf{s}}_i, \forall i \in \mathcal{I} \right\},$$

où, pour chaque  $i \in \mathcal{I}$ , la ligne  $\Phi_i$  et le seuil  $\bar{\mathbf{s}}_i$  représentent respectivement un hyperplan de contrainte issu d'un scénario de contingence donné et l'intervalle admissible associé.

Dans le cas des formulations à contraintes dures, la notion de redondance structurelle s'appuie sur une propriété géométrique de l'ensemble admissible. Comme nous avons discuté à la section 2.3.2, une contrainte d'indice  $i \in \mathcal{I}$  est dite essentielle si sa suppression modifie l'ensemble réalisable, c'est-à-dire si

$$\mathcal{F}(\Phi, \bar{\mathbf{s}}, \mathcal{I} \setminus \{i\}) \neq \mathcal{F}(\Phi, \bar{\mathbf{s}}, \mathcal{I}).$$

Autrement dit, une contrainte est considérée comme indispensable dès lors qu'elle contribue activement à délimiter la région des injections admissibles.

Cette définition géométrique ne se transpose toutefois pas directement aux problèmes formulés avec des contraintes souples, dans lesquels la faisabilité est assurée par construction. En effet, dans le SCED pénalisé, chaque limite de contrainte peut être dépassée moyennant une pénalité dans la fonction objectif, via une variable de dépassement  $\xi_i^{(c)} \geq 0$ . Ce mécanisme modifie la nature du rôle joué par les contraintes thermiques dans le problème : une contrainte de limite d'écoulement n'influe sur la solution optimale que si elle est dépassée ( $\xi_i^* > 0$ ) ou si le

multiplicateur de Lagrange associé à cette contrainte  $\lambda_i^*$  est strictement positif.

Dans ce contexte, nous généralisons la notion de redondance en nous basant non plus sur l'ensemble  $\mathcal{F}$ , mais sur le comportement de la solution optimale et des multiplicateurs de Lagrange. Soient deux instances (3.1) et (3.5) du SCED pénalisé, associées respectivement à un ensemble complet de contraintes  $\mathcal{I}$  et à un sous-ensemble  $\mathcal{I}' \subset \mathcal{I}$ . Nous notons leurs formulations respectives :

$$\min_{\mathbf{p}, \boldsymbol{\xi}} f(\mathbf{p}) + \sum_{i \in \mathcal{I}} M \xi_i \quad (3.1a)$$

$$\text{s.c.} \quad \sum_{n \in \mathcal{N}} p_n = 0 \quad (3.1b)$$

$$\underline{p}_n \leq p_n \leq \bar{p}_n, \quad \forall n \in \mathcal{N} \quad (3.1c)$$

$$\Phi_i \mathbf{p} \leq \bar{\mathbf{s}}_i + \boldsymbol{\xi}_i, \quad \forall i \in \mathcal{I}, \quad (3.1d)$$

et,

$$\min_{\mathbf{p}, \boldsymbol{\xi}} f(\mathbf{p}) + \sum_{i \in \mathcal{I}'} M \xi_i \quad (3.2a)$$

$$\text{s.c.} \quad \sum_{n \in \mathcal{N}} p_n = 0 \quad (3.2b)$$

$$\underline{p}_n \leq p_n \leq \bar{p}_n, \quad \forall n \in \mathcal{N} \quad (3.2c)$$

$$\Phi_i \mathbf{p} \leq \bar{\mathbf{s}}_i + \boldsymbol{\xi}_i, \quad \forall i \in \mathcal{I}', \quad (3.2d)$$

où les variables  $\xi_i$  modélisent le dépassement admissible des seuils thermiques, avec une pénalisation uniforme d'amplitude  $M > 0$ . Nous dirons qu'une contrainte  $i \in \mathcal{I}$  est structurellement inutile si les ensembles d'optima des problèmes (3.1) et (3.5) coïncident lorsque  $i \notin \mathcal{I}'$ , c'est-à-dire lorsque la contrainte est retirée du modèle. Dans ce cas, la contrainte n'influe ni sur la solution optimale  $\mathbf{p}^*$ , ni sur les dépassements  $\boldsymbol{\xi}^*$ , ni sur la valeur de l'objectif. Une telle situation ne se produit que si, à l'optimum, la contrainte est à la fois respectée ( $\xi_i^* = 0$ ) et dualement inactive ( $\lambda_i^* = 0$ ). Dans le cas contraire, même en l'absence de pénalité explicite, la contrainte peut exercer une influence sur l'orientation du gradient stationnaire via un multiplicateur strictement positif, et sa suppression modifie alors l'équilibre optimal. Ceci est formalisé à la proposition 2.

*Proposition 2* (Conditions suffisantes pour l'équivalence des minima entre problèmes

pénalisés). Considérons le problème pénalisé ;

$$\begin{aligned} \mathfrak{P} : \quad & \min_{\mathbf{x} \in \mathbb{R}^n, \boldsymbol{\xi} \in \mathbb{R}_+^m} \quad f(\mathbf{x}) + \rho \mathbf{1}^\top \boldsymbol{\xi} \\ & \text{s.c.} \quad \mathbf{A}\mathbf{x} \leq \mathbf{b} + \boldsymbol{\xi}, \end{aligned} \tag{3.3}$$

où  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  est convexe et différentiable,  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{R}^m$  et  $\rho > 0$ . Nous notons  $(\mathbf{x}^*, \boldsymbol{\xi}^*) \in \mathbb{R}^n \times \mathbb{R}_+^m$  une solution primal de (3.4), et  $\boldsymbol{\lambda}^* \in \mathbb{R}_+^m$ , un vecteur des solutions duales associées aux contraintes d'inégalité  $\mathbf{a}_i^\top \mathbf{x} \leq b_i + \xi_i$ , où  $\mathbf{a}_i^\top$  désigne la  $i$ <sup>ème</sup> ligne de  $\mathbf{A}$ .

Soit  $i \in \{1, \dots, m\}$ , et soit  $\mathfrak{P}_{-i}$ , le problème obtenu en retirant la  $i$ <sup>ème</sup> contrainte :

$$\begin{aligned} \mathfrak{P}_{-i} : \quad & \min_{\mathbf{x}_{-i} \in \mathbb{R}^n, \tilde{\boldsymbol{\xi}} \in \mathbb{R}_+^m} \quad f(\mathbf{x}_{-i}) + \rho \mathbf{1}^\top \tilde{\boldsymbol{\xi}} \\ & \text{s.c.} \quad \mathbf{a}_j^\top \mathbf{x}_{-i} \leq b_j + \tilde{\xi}_j, \quad \forall j \in \{1, \dots, m\} \setminus \{i\}, \end{aligned} \tag{3.4}$$

dont une solution optimale primal est dénoté  $(\mathbf{x}_{-i}^*, \tilde{\boldsymbol{\xi}}^*)$

Si  $\xi_i^* = 0$  et  $\lambda_i^* = 0$ , alors les minima des deux problèmes coïncident, c'est-à-dire  $\mathfrak{P}^* = \mathfrak{P}_{-i}^*$ .

*Démonstration de la proposition 2.* Soient  $(\mathbf{x}^*, \boldsymbol{\xi}^*) \in \mathbb{R}^n \times \mathbb{R}_+^m$  et  $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \in \mathbb{R}_+^m \times \mathbb{R}_+^m$  une solution primale-duale du problème complet (3.4).

Si  $\xi_i^* = 0$ , alors les problèmes  $\mathfrak{P}$  et  $\mathfrak{P}_{-i}$  s'écrivent respectivement :

$$\begin{aligned} \mathfrak{P}^* : \quad & \min_{\mathbf{x} \in \mathbb{R}^n, \boldsymbol{\xi} \in \mathbb{R}_+^m} \quad f(\mathbf{x}) + \rho \sum_{j \neq i}^m \xi_j \\ & \text{s.c.} \quad \mathbf{a}_j^\top \mathbf{x} \leq b_j + \xi_j, \quad \forall j \in \{1, \dots, m\}, \end{aligned}$$

et,

$$\begin{aligned} \mathfrak{P}_{-i}^* : \quad & \min_{\mathbf{x}_{-i} \in \mathbb{R}^n, \tilde{\boldsymbol{\xi}} \in \mathbb{R}_+^{m-1}} \quad f(\mathbf{x}_{-i}) + \rho \sum_{j \neq i}^m \tilde{\xi}_j \\ & \text{s.c.} \quad \mathbf{a}_j^\top \mathbf{x}_{-i} \leq b_j + \tilde{\xi}_j, \quad \forall j \in \{1, \dots, m\} \setminus \{i\}. \end{aligned}$$

Par construction, le problème  $\mathfrak{P}_{-i}$  constitue une relaxation de  $\mathfrak{P}$ .

Or, comme  $\xi_i^* = 0$  et  $\lambda_i^* = 0$ , la contrainte  $i$  du problème  $\mathfrak{P}$  n'est pas active à un optimum ; autrement dit,  $\mathbf{a}_i^\top \mathbf{x}^* < b_i$ . La contrainte  $i$  peut donc être retirée sans modifier l'ensemble des

solutions optimales, et le problème s'écrit désormais :

$$\begin{aligned} \mathfrak{P}^* : \quad & \min_{\mathbf{x} \in \mathbb{R}^n, \xi \in \mathbb{R}_+^m} \quad f(\mathbf{x}) + \rho \sum_{j \neq i}^m \xi_j \\ & \text{s.c.} \quad \mathbf{a}_j^\top \mathbf{x} \leq b_j + \xi_j, \quad \forall j \in \{1, \dots, m\} \setminus \{i\}, \end{aligned}$$

ce qui est identique à la formulation de  $\mathfrak{P}_{-i}$ . Nous en déduisons que  $\mathfrak{P}^* = \mathfrak{P}_{-i}^*$ .  $\square$

Cette généralisation reflète un changement de perspective : dans une formulation pénalisée, ce n'est plus la position géométrique d'une contrainte vis-à-vis du polytope admissible qui détermine son utilité structurelle, mais son influence effective sur l'optimum du problème. En particulier, une contrainte dont la variable de dépassement  $\xi_i$  est nulle et dont le multiplicateur dual  $\lambda_i$  est nul pour toutes les solutions optimales peut être ignorée sans perte d'exactitude, car elle ne modifie ni la solution ni la valeur de l'objectif. À l'inverse, toute contrainte dépassée ou dualement active contribue à l'équilibre optimal et ne peut être omise sans altérer l'ensemble réalisable, et par conséquent, le comportement du problème.

À partir du système d'inégalités linéaires  $\Phi_i \mathbf{p} \leq \bar{\mathbf{s}}_i$ , indexé par l'ensemble complet de contraintes  $\mathcal{I}$ , nous construisons une hiérarchie d'ensembles de contraintes non redondantes par application récursive de la procédure REDUNDANCYREMOVAL, présentée à l'Algorithme 1 et inspirée de [12]. Soit  $\mathbf{z} \in \text{int } \mathcal{F}(\Phi, \bar{\mathbf{s}})$ , un point intérieur de l'ensemble admissible défini par les contraintes de limite d'écoulement. La procédure de suppression repose sur un point strictement admissible  $\mathbf{z} \in \text{int } \mathcal{F}(\Phi, \bar{\mathbf{s}})$ , garantissant la stabilité géométrique du filtrage. Dans notre contexte, le point  $\mathbf{z}$  peut être choisi égal à  $\mathbf{0}$ , puisque les injections nulles satisfont nécessairement toutes les contraintes d'écoulement de puissance et constituent un point intérieur du polytope admissible, comme le souligne [12].

L'algorithme HIERARCHICALCONSTRAINTEXTRACTION, présenté ci-dessous à l'Algorithme 3, consiste à appliquer REDUNDANCYREMOVAL de manière itérative sur le résidu des contraintes restantes à chaque étape. À chaque itération  $k$ , un sous-ensemble  $\mathcal{I}_k \subset \mathcal{I}$  de contraintes non redondantes, relatives au résidu courant, est extrait et retiré de l'ensemble initial. Ce processus se poursuit jusqu'à épuisement de  $\mathcal{I}$ , révélant ainsi une structure hiérarchique implicite.

Chaque niveau  $\mathcal{I}_k$  peut être interprété comme un ensemble minimal de contraintes pertinentes dans le résidu du problème, relatif aux contraintes non capturées par les niveaux inférieurs. À cette collection d'ensembles de contraintes, nous associons une séquence croissante d'ensembles

---

**Algorithme 3 : HIERARCHICALCONSTRAINTEXTRACTION( $\Phi, \bar{s}, z, \mathcal{I}$ )**


---

**Input :**

- Un système d'inégalités linéaires  $\Phi p \leq \bar{s}$  indexé par  $\mathcal{I}$  ;
- Un point strictement intérieur  $z \in \text{int } \mathcal{F}(\Phi, \bar{s})$  ;
- L'ensemble initial des indices de contraintes  $\mathcal{I}$ .

**Output :** Une séquence hiérarchique d'ensembles  $\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_{k_{\max}}$  telle que  
 $\bigcup_{k=1}^{k_{\max}} \mathcal{I}_k = \mathcal{I}$ .

```

1 Filtrage initial des duplications exactes :
2 foreach  $i \in \mathcal{I}$  do
3   if il n'existe aucun  $j < i$  tel que  $\Phi_i = \Phi_j$  et  $\bar{s}_i = \bar{s}_j$  then
4     Ajouter  $i$  à un nouvel ensemble  $\mathcal{I}^{\text{filtré}}$ ;
5  $\mathcal{I} \leftarrow \mathcal{I}^{\text{filtré}}$                                 // Suppression des duplications exactes
6  $k \leftarrow 1$  ;
7 while  $\mathcal{I} \neq \emptyset$  do
8    $\mathcal{I}_k \leftarrow \text{REDUNDANCYREMOVAL}(\Phi, \bar{s}, z, \mathcal{I})$  ;
9    $\mathcal{I} \leftarrow \mathcal{I} \setminus \mathcal{I}_k$  ;
10   $k \leftarrow k + 1$  ;
11  $k_{\max} \leftarrow k - 1$  ;
12 return  $\{\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_{k_{\max}}\}$ 

```

---

admissibles  $\mathcal{F}_k$ , définis pour tout  $k \in \{1, \dots, k_{\max}\}$  par :

$$\mathcal{F}_k := \left\{ p \in \mathbb{R}^{|\mathcal{N}|} \mid \Phi_i p \leq \bar{s}_i, \forall i \in \bigcup_{j=1}^k \mathcal{I}_j \right\}.$$

Par construction, ces ensembles satisfont la chaîne d'inclusion stricte suivante :

$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots \subset \mathcal{F}_{k_{\max}} \equiv \mathcal{F},$$

où  $\mathcal{F} \subset \mathbb{R}^{|\mathcal{N}|}$  désigne la région admissible complète définie par l'ensemble total des contraintes  $\mathcal{I}$ . L'inclusion stricte entre les ensembles admissibles  $\mathcal{F}_k \subset \mathcal{F}_{k+1}$  résulte notamment d'un filtrage initial intégré aux lignes 2 à 5 de l'algorithme HIERARCHICALCONSTRAINTEXTRACTION (Algorithme 3). Ce filtrage vise à éliminer les duplications exactes parmi les contraintes d'écoulement de puissance active, c'est-à-dire les inégalités redondantes produites par des scénarios de contingence structurellement équivalents. Dans le cas d'un réseau fortement maillé, il est fréquent que certaines contraintes apparaissent plusieurs fois avec des coefficients et des seuils identiques. Ces duplications sont détectées et supprimées avant la première itération de l'algorithme, garantissant ainsi que le premier niveau  $\mathcal{I}_1$  ne contient que des contraintes distinctes, et que chaque niveau ultérieur défini la région admissible.

La construction hiérarchique ainsi obtenue est effectuée une seule fois en amont de notre méthodologie, indépendamment des profils de charge à traiter. Il est par ailleurs possible de renforcer cette procédure en appliquant en amont un critère d'écrêtage fondé sur les LODF [12], moyennant une perte d'exactitude de la solution. Considérons une contrainte limite d'écoulement de puissance associée à la ligne de transport  $\ell \in \mathcal{L}$ , observée sous un scénario de contingence  $c \in \mathcal{C}$ , correspondant à la mise hors service de la ligne  $c$ . Dans ce contexte, la matrice LODF,  $\Lambda$ , encode, pour chaque paire  $(\ell, o)$ , la sensibilité de l'écoulement sur la ligne  $\ell$  induite par la perte de la ligne  $c$ . Une telle contrainte peut être ignorée si la variation induite est négligeable au regard de la limite de puissance  $\bar{s}_\ell$ , selon le critère suivant :

$$|\Lambda_{\ell c}| \cdot \mathbf{p}_c < \eta \cdot \bar{s}_\ell,$$

où  $\mathbf{p}_c$  désigne l'écoulement pré-contingence sur la ligne  $c$ ,  $\bar{s}_\ell$  la limite thermique de la ligne  $\ell$ , et  $\eta > 0$  un seuil fixé a priori, e.g., 0,05 %. Ce test garantit que, même en cas de défaillance de la ligne  $c$ , l'effet sur la ligne  $\ell$  demeure négligeable, ce qui justifie son exclusion dans la hiérarchie. Ce type de filtrage permet d'exclure, sans perte significative d'information, les contraintes liées à des perturbations topologiques dont l'effet sur le réseau est marginal. Dans notre démarche, nous choisissons de ne pas appliquer ce filtrage en amont, afin de préserver l'exactitude structurelle de la hiérarchie construite. Néanmoins, cette approximation constitue une alternative pertinente lorsque la priorité est donnée à l'efficacité de la réduction du nombre de contraintes plutôt qu'à la fidélité de la solution optimale.

Chaque niveau  $k$  de la hiérarchie correspond ainsi à une extension de la région des solutions admissibles  $\mathcal{F}_k$ , l'ajout progressif de contraintes permettant d'encadrer de manière de plus en plus stricte les injections nodales réalisables au regard des limites thermiques. Il devient alors pertinent de rechercher pour chaque profil de charge  $\mathbf{p}_d \in \mathbb{R}^{|\mathcal{N}|}$  le plus petit indice  $k^* \in \{1, 2, \dots, k_{\max}\}$  tel que la solution optimale du SCED restreint aux contraintes  $\mathcal{I}_{1:k^*}$  coïncide avec celle du problème complet. Formellement, nous considérons la version réduite du modèle pénalisé, notée

$$\min_{\mathbf{p}, \boldsymbol{\xi}} f(\mathbf{p}) + \sum_{i \in \mathcal{I}_{1:k^*}} M \xi_i \quad (3.5a)$$

$$\text{s.c.} \quad \sum_{n \in \mathcal{N}} p_n = 0 \quad (3.5b)$$

$$\underline{p}_n \leq p_n \leq \bar{p}_n, \quad \forall n \in \mathcal{N} \quad (3.5c)$$

$$\Phi_i \mathbf{p} \leq \bar{\mathbf{s}}_i + \boldsymbol{\xi}_i, \quad \forall i \in \mathcal{I}_{1:k^*}, \quad (3.5d)$$

et nous nous intéressons au plus petit niveau  $k^*$  tel que la solution optimale  $(\mathbf{p}^*, \boldsymbol{\xi}^*)$  de (3.5)

coïncide avec celle de la formulation complète (3.1). La prédiction de ce niveau  $k^*$ , fondée sur les caractéristiques du profil nodal  $\mathbf{p}_d$ , constitue notre démarche d'apprentissage supervisé présentée dans la section suivante.

### 3.2.2 Modèle de classification supervisée avec LightGBM

Dans le cadre de notre approche, l'objectif est d'associer à chaque profil de charge  $\mathbf{p}_d \in \mathbb{R}^{|\mathcal{N}|}$  un niveau hiérarchique minimal  $k^* \in \{1, \dots, k_{\max}\}$  tel que la résolution du SCED restreint aux seules contraintes des niveaux 1 à  $k^*$  produise une solution optimale identique à celle obtenue à partir de l'ensemble complet. Il s'agit donc d'un problème de classification hiérarchique à structure ordonnée, dans lequel la sortie recherchée n'est pas une classe arbitraire, mais un seuil structurel au sein d'une séquence croissante de relaxations successives du modèle complet. Ce problème se prête naturellement à une formulation cumulative, où nous cherchons à estimer, pour chaque niveau  $k$ , la pertinence structurelle de l'ensemble  $\mathcal{I}_k$  au regard du profil  $\mathbf{p}_d$ . Le niveau  $k^*$  est ensuite déterminé comme le plus petit niveau garantissant l'exactitude du SCED restreint. Le vecteur de charge nodale  $\mathbf{p}_d \in \mathbb{R}^{|\mathcal{N}|}$ , décrivant les charges anticipées à chaque noeud du réseau, constitue l'unique entrée du modèle prédictif. La sortie recherchée est un entier  $k^* \in \{1, 2, \dots, k_{\max}\}$ , correspondant au plus petit niveau hiérarchique pour lequel la résolution du SCED avec l'ensemble cumulé de contraintes  $\mathcal{I}_{1:k^*} := \bigcup_{j=1}^{k^*} \mathcal{I}_j$  restitue exactement la solution optimale du problème complet. Autrement dit,  $k^*$  représente le niveau à partir duquel l'ajout de contraintes supplémentaires n'a plus d'effet sur la variable de décision optimale, ni sur la valeur de l'objectif. Ce problème de classification peut être interprété comme une cartographie

$$\mathbb{R}^{|\mathcal{N}|} \rightarrow \{1, \dots, k_{\max}\},$$

dans laquelle la régularité géométrique des niveaux hiérarchiques, construits de manière récursive et disjoints, permet de structurer l'espace de décision de manière stratifiée. La nature ordonnée du problème se prête alors à un encodage binaire cumulatif, qui permet de transformer chaque étiquette scalaire  $k^*$  en un vecteur de dimension  $k_{\max}$ , tout en conservant l'information hiérarchique. Pour tout profil de charge  $\mathbf{p}_d$ , nous associons le vecteur d'étiquettes binaires cumulatives

$$\mathbf{z}(\mathbf{p}_d) := (z_1, z_2, \dots, z_{k_{\max}})^\top \in \{0, 1\}^{k_{\max}}, \quad (3.6)$$

défini à partir du plus petit niveau  $k^* \in \{1, \dots, k_{\max}\}$  garantissant que le SCED restreint à l'ensemble de contraintes  $\mathcal{I}_{1:k^*}$  reproduit exactement la solution du problème complet. Chaque

composante  $z_k$  est alors donnée pour tout  $k^* \in \{1, \dots, k_{\max}\}$  par

$$z_k := \mathbb{1}_{[k \leq k^*]} = \begin{cases} 1 & \text{si } k \leq k^* \\ 0 & \text{sinon,} \end{cases}$$

où  $\mathbb{1}$  désigne la fonction indicatrice standard, égale à 1 lorsque la condition est satisfaite, et 0 sinon. Ainsi, pour tout  $k < k^*$ , nous avons  $z_k = 1$ , et pour tout  $k > k^*$ ,  $z_k = 0$ .

Cette représentation induit, par construction, une décroissance naturelle des probabilités conditionnelles :

$$P[z_k = 1 | \mathbf{p}_d] \leq P[z_{k-1} = 1 | \mathbf{p}_d],$$

et permet de reconstruire l'étiquette cible initiale à partir des prédictions binaires en définissant :

$$\hat{k}^* := \max\{k \in \{1, \dots, k_{\max}\} \mid \hat{z}_k = 1\},$$

où chaque variable  $\hat{z}_k \in \{0, 1\}$  est obtenue par seuillage d'une probabilité prédictive  $P = P[z_k = 1 | \mathbf{p}_d]$ , selon la règle

$$\hat{z}_k := \begin{cases} 1 & \text{si } P \geq \tau, \\ 0 & \text{sinon,} \end{cases}$$

avec  $\tau \in (0, 1)$  un seuil de décision fixé à l'avance.

Cet encodage préserve explicitement la structure cumulative du problème, tout en permettant une implémentation modulaire via une collection de classificateurs binaires indépendants  $\mathfrak{C}_k$ , chacun entraîné à approximer la loi de Bernoulli associée à  $z_k$ . La prédiction finale est obtenue par agrégation des sorties binaires, selon une règle de maximum structurel. L'apprentissage de cette famille de classificateurs est effectué en utilisant un modèle de LightGBM [74].

Pour exploiter la représentation binaire cumulative introduite ci-dessus, nous construisons une famille ordonnée de modèles de classification supervisée, notée  $\{\mathfrak{C}_k\}_{k=1}^{k_{\max}}$ . Contrairement à une approche directe consistant à prédire une étiquette  $k^*$  unique, nous formulons le problème comme une suite de décisions binaires imbriquées où chaque classificateur  $\mathfrak{C}_k$  est spécifiquement entraîné à estimer la probabilité que le niveau hiérarchique  $k$  soit suffisant pour garantir l'exactitude de la solution du SCED restreint, c'est-à-dire que l'ensemble admissible  $\mathcal{F}_{1:k}$ , défini par les contraintes  $\mathcal{I}_{1:k}$ , contient l'optimum du problème original (3.1). Cette reformulation cumulative permet de structurer l'apprentissage autour d'un encodage plus stable, et de mieux exploiter la nature ordonnée des niveaux hiérarchiques, sans imposer une classification multi-classes explicite ni supposer l'indépendance entre les niveaux. Les modèles  $\mathfrak{C}_k$  sont implémentés à l'aide du modèle LightGBM [74], qui constitue une version hautement optimisée

du paradigme des arbres de décision entraînés par gradient (*gradient boosted decision trees*, GBDT). Cette méthode repose sur la construction récursive d'une suite de fonctions de régression  $f^{(t)} : \mathbb{R}^d \rightarrow \mathbb{R}$ , typiquement des arbres de décision de profondeur fixée, selon la règle additive

$$F^{(t)}(\mathbf{x}) = F^{(t-1)}(\mathbf{x}) + \gamma^{(t)} f^{(t)}(\mathbf{x}),$$

où  $F^{(t)} : \mathbb{R}^d \rightarrow \mathbb{R}$  désigne le modèle agrégé à l'itération  $t$ , et  $\gamma^{(t)} \in \mathbb{R}$  un coefficient d'ajustement optimisé à chaque étape par minimisation du risque empirique. Chaque étape du boosting consiste à approximer le gradient négatif d'une fonction de perte  $\mathfrak{L} : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ , supposée deux fois continûment différentiable par rapport à sa deuxième composante, c'est-à-dire la prédiction du modèle. Notons  $\mathfrak{L}(y, \hat{y})$  la perte associée à une étiquette cible  $y \in \mathbb{R}$  et à une prédiction  $\hat{y} \in \mathbb{R}$ , et considérons la dérivation uniquement par rapport à  $\hat{y}$ . Pour chaque itération  $t$ , le modèle  $f^{(t)} \in \mathcal{H}$  est obtenu comme solution du problème :

$$f^{(t)} \in \arg \min_{f \in \mathcal{H}} \sum_{i=1}^n \left[ g_i f(\mathbf{x}_i) + \frac{1}{2} h_i f^2(\mathbf{x}_i) \right],$$

où  $\mathcal{H}$  désigne l'espace des arbres de régression admissibles, et où

$$g_i := \frac{\partial}{\partial \hat{y}} \mathfrak{L}(y_i, \hat{y}) \Big|_{\hat{y}=F^{(t-1)}(\mathbf{x}_i)}, \quad h_i := \frac{\partial^2}{\partial \hat{y}^2} \mathfrak{L}(y_i, \hat{y}) \Big|_{\hat{y}=F^{(t-1)}(\mathbf{x}_i)}$$

sont respectivement le gradient et la hessienne locaux de la perte, évalués sur l'itération précédente.

L'efficacité computationnelle de LightGBM provient de deux éléments fondamentaux :

- L'échantillonnage unilatéral fondé sur le gradient (*gradient-based one-side sampling*, GOSS) : cette stratégie consiste à sélectionner prioritairement les observations associées à des gradients de perte élevés, qui sont statistiquement plus informatives pour la mise à jour du modèle. Soient  $A$  les indices des  $a\%$  d'observations ayant les plus grands gradients, et  $B$  un échantillon aléatoire parmi les autres. Le gain d'information est alors estimé sur  $A \cup B$ , en répondant  $B$  pour compenser l'échantillonnage, ce qui permet de réduire le nombre d'instances sans compromettre la convergence ;
- Le regroupement exclusif de variables (*exclusive feature bundling*, EFB) : cette technique repose sur la détection d'attributs mutuellement exclusifs, c'est-à-dire des colonnes dupliquées dans lesquelles au plus une entrée non nulle apparaît par ligne, et leur regroupement dans un même espace de représentation à faible dimension. Ce mécanisme exploite la structure creuse des descripteurs pour construire une représentation

compacte, sans perte d'information discriminante, réduisant ainsi la complexité du tri des valeurs candidats.

Ces deux mécanismes permettent de maintenir une complexité algorithmique sous-linéaire en la taille du jeu de données, même lorsque la dimension du vecteur d'entrée  $\mathbf{x} \in \mathbb{R}^{|\mathcal{N}|}$ , représentant dans notre contexte les charges nodales  $\mathbf{p}_d$ , devient élevée. Il en résulte une capacité à s'adapter aux contraintes pratiques des problèmes de décision opératoire, tout en assurant une convergence rapide et une expressivité robuste. C'est à ce titre que LightGBM constitue un choix naturel pour modéliser la famille  $\{\mathfrak{C}_k\}_{k=1}^{k_{\max}}$ , dans laquelle chaque modèle opère indépendamment sur une tâche de classification binaire à données tabulaires. Formellement, chaque modèle  $\mathfrak{C}_k$  approxime la fonction

$$\mathbf{p}_d \longmapsto P[z_k = 1 | \mathbf{p}_d] \quad (3.7)$$

où  $z_k \in \{0, 1\}$  désigne la composante  $k$  du vecteur binaire cumulatif  $\mathbf{z}(\mathbf{p}_d)$ , prenant la valeur 1 si le niveau  $k$  est suffisant, et 0 sinon. L'entraînement s'effectue en minimisant une fonction de perte logistique sur des sous-ensembles d'entraînement simulés, décrits à la section 3.3.1, dans lesquels chaque profil  $\mathbf{p}_d$  est annoté par le plus petit niveau  $k^*$  tel que la solution restreinte au niveau  $\mathcal{I}_{1:k^*}$  coïncide avec celle du SCED complet.

La prédiction globale du niveau hiérarchique repose ensuite sur l'agrégation des sorties binaires issues de la famille  $\{\mathfrak{C}_k\}_{k=1}^{k_{\max}}$ . Chaque classifieur produit une probabilité  $P_k = P[z_k = 1 | \mathbf{p}_d]$ , que nous convertissons en prédiction binaire à l'aide d'un seuil de décision  $\tau \in (0, 1)$ , tel que  $\hat{z}_k = 1$  si et seulement si  $P_k \geq \tau$ . La prédiction finale du niveau de contraintes à considérer s'obtient alors comme le plus grand niveau validé par cette règle, soit

$$\hat{k}^* := \max\{k \in \{1, \dots, k_{\max}\} \mid \hat{z}_k = 1\}.$$

Le paramètre  $\tau$  intervient comme un compromis central entre conservatisme et sélectivité statistique. Un seuil élevé privilégie la précision du classifieur, en restreignant l'inclusion aux seules contraintes dont la nécessité est fortement probable, au prix d'un risque de sous-prédiction plus fort. Inversement, un seuil plus faible élargit la couverture en acceptant un nombre plus important de contraintes potentiellement superflues, mais permet de réduire significativement la probabilité d'omettre une contrainte effectivement nécessaire à l'exactitude de la solution. Dans un cadre de résolution sous contraintes souples, cette stratégie favorise un rappel élevé, aligné avec l'objectif de robustesse structurelle.

Notamment, dans le contexte opérationnel du SCED, les conséquences d'une sous-prédiction sont potentiellement critiques car ignorer une contrainte réellement nécessaire peut conduire à

une solution inexacte, voire à un non respect implicite des limites  $\bar{s}$ . Il est donc préférable, dans notre cadre, de privilégier systématiquement la sur-prédiction conservatrice, quitte à inclure des contraintes redondantes, plutôt que de risquer l'omission d'une contrainte essentielle. En pratique, ce principe se traduit par un choix de seuil strictement inférieur à 0.5, typiquement  $\tau \approx 0.4$ , assurant une couverture hiérarchique complète dans les cas douteux.

Par ailleurs, la structure cumulative de l'encodage implique que toute prédiction positive  $\hat{z}_k = 1$  entraîne automatiquement  $\hat{z}_j = 1$  pour tout  $j < k$ . Ainsi, un seuil conservateur propage naturellement la décision vers les niveaux inférieurs, renforçant le conservatisme global de la prédiction et donc l'inclusion de tous les ensembles hiérarchiques nécessaires. Ce mécanisme garantit que les contraintes les plus fondamentales (niveaux bas) sont généralement considérées, même en cas d'incertitude sur les niveaux supérieurs.

Afin de garantir la robustesse de la procédure vis-à-vis des erreurs de sous-prédiction, un mécanisme de reprise adaptative est mis en œuvre à l'issue de la résolution initiale du SCED réduit (3.5) aux contraintes hiérarchiques  $\mathcal{I}_{1:\hat{k}^*}$ . La solution optimale ainsi obtenue, notée  $\mathbf{p}^*$ , est ensuite soumise à une phase de validation a posteriori portant sur les contraintes restantes, c'est-à-dire celles indexées par

$$\mathcal{I}_{\hat{k}^*+1:k_{\max}} := \bigcup_{k=\hat{k}^*+1}^{k_{\max}} \mathcal{I}_k.$$

Pour chaque contrainte  $i \in \mathcal{I}_{\hat{k}^*+1:k_{\max}}$ , nous vérifions si la limite d'écoulement de puissance est strictement respectée, autrement dit si

$$\Phi_i \mathbf{p}^* < \bar{s}_i. \quad (3.8)$$

Cette opération de validation a posteriori repose uniquement sur le calcul d'un produit scalaire entre la ligne  $i$  de la matrice  $\Phi$  et le vecteur d'injections nodales optimales  $\mathbf{p}^*$ , et peut donc être réalisée de manière efficace. Dans le cas contraire, c'est-à-dire s'il existe un indice  $i$  pour lequel l'inégalité (3.8) n'est pas satisfaite, le niveau hiérarchique est incrémenté et la résolution du SCED est relancée avec l'ensemble des contraintes étendu d'un niveau hiérarchique, soit  $\mathcal{I}_{1:\hat{k}^*+1}$ . Ce mécanisme incrémental garantit, en un nombre fini d'étapes  $k_{\max} - \hat{k}^*$ , la récupération exacte de la solution du SCED complet, dès lors que la condition (3.8) est satisfaite. La reprise permet ainsi d'inclure a posteriori toute contrainte dont la variable de pénalité  $\xi_i$  deviendrait strictement positive, ou dont le multiplicateur de Lagrange associé  $\lambda_i$  serait non nul, conformément à la proposition 2.

Ce mécanisme de reprise, combiné à la structure imbriquée des contraintes hiérarchiques, assure une convergence vers la solution exacte, tout en maintenant un compromis favorable

entre robustesse opérationnelle et efficacité algorithmique.

### 3.3 Chaîne de traitement d'apprentissage supervisé

Cette section décrit en détail les étapes méthodologique mise en œuvre pour l'apprentissage supervisé du niveau hiérarchique requis dans la résolution partielle du SCED. À partir d'une hiérarchie fixe d'ensembles de contraintes  $\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_{k_{\max}}$  construite en amont, notre objectif est d'apprendre une fonction de prédiction prenant en entrée un vecteur de charge nodale  $\mathbf{p}_d \in \mathbb{R}^{|\mathcal{N}|}$ , et renvoyant le plus petit niveau  $k^*$  suffisant pour garantir que la solution du SCED restreint à  $\mathcal{I}_{1:k^*}$  coïncide avec celle du problème complet. Le traitement repose sur la construction automatique d'un jeu de données simulé, l'encodage binaire cumulatif des étiquettes, l'entraînement d'une famille de classificateurs  $\{\mathfrak{C}_k\}_{k=1}^{k_{\max}}$ , la calibration du seuil décisionnel  $\tau$ , et enfin l'exploitation opérationnelle du modèle pour la résolution prédictive. Chaque étape est conçue pour préserver la structure hiérarchique du problème, assurer l'exactitude des solutions, et garantir une robustesse algorithmique face à l'incertitude sur le profil de charge.

#### 3.3.1 Génération du jeu de données

L'objectif de cette phase est de construire un jeu de données supervisé sous la forme d'une collection de couples  $(\mathbf{p}_d, k^*)$ , dans un chaque profil de charge nodale  $\mathbf{p}_d \in \mathbb{R}^{|\mathcal{N}|}$  est associé au plus petit niveau hiérarchique  $k^* \in \{1, \dots, k_{\max}\}$  garantissant que la solution du SCED restreint à l'ensemble  $\mathcal{I}_{1:k^*}$  coïncide avec celle du modèle complet (1.20).

Chaque échantillon est construit de manière indépendante selon le protocole suivant :

1. Résolution du SCED complet (1.20) sur le profil  $\mathbf{p}_d$ , en considérant l'ensemble intégral de contraintes  $\mathcal{I}$ , puis identification du plus petit entier  $k^* \in \{1, \dots, k_{\max}\}$  tel qu'il existe au moins une variable de pénalité  $\xi_i > 1$  ou une variable de Lagrange  $\lambda_i > 1$  associée à une contrainte  $i$  appartenant au niveau  $\mathcal{I}_{k^*}$  ;
2. Pour chaque  $k \in \{1, \dots, k_{\max}\}$ , résolution du SCED restreint aux contraintes  $\mathcal{I}_{1:k^*}$ , et comparaison avec la solution de référence, laquelle reste toujours identique ;
3. Enregistrement structuré de l'échantillon  $(\mathbf{p}_d, k^*)$  dans le jeu de données supervisé.

Afin de garantir une couverture équilibrée de tous les niveaux hiérarchiques dans l'espace des charges, nous avons recours à une stratégie d'exploration forcée fondée sur des mécanismes adaptatifs. L'espace des profils  $\mathbf{p}_d$  étant de dimension élevée, les frontières séparant les régions admissibles successives  $\mathcal{F}_k \subset \mathcal{F}_{k+1}$  ne peuvent être approximées de manière uniforme. L'écart géométrique entre deux régions  $\mathcal{F}_k$  et  $\mathcal{F}_{k+1}$  varie fortement selon les directions explorées : il

peut être étendu dans certaines dimensions mais quasi-nul dans d'autres. Il s'ensuit que la relation entre  $\mathbf{p}_d$  et le niveau minimal requis  $k^*$  n'est ni monotone, ni localement lisse. De faibles perturbations peuvent induire des transitions abruptes entre niveaux hiérarchiques si elles croisent des régions étroites, tandis que des variations substantielles peuvent rester sans effet si elles empruntent des directions topologiquement peu sensibles. Cette absence de régularité provient de la densité des matrices  $\Phi^{(c)}$ , qui induisent un couplage global entre l'ensemble des injections et l'ensemble des contraintes thermiques. Chaque composante de  $\mathbf{p}_d$  peut ainsi influencer tout le spectre des contraintes de sécurité, rendant les frontières entre niveaux  $k^*$  non convexes, non séparables, et difficilement prévisibles. Il en découle l'impossibilité de générer un jeu de données équilibré à partir d'un échantillonnage aléatoire ou isotrope, justifiant pleinement l'emploi d'une exploration guidée.

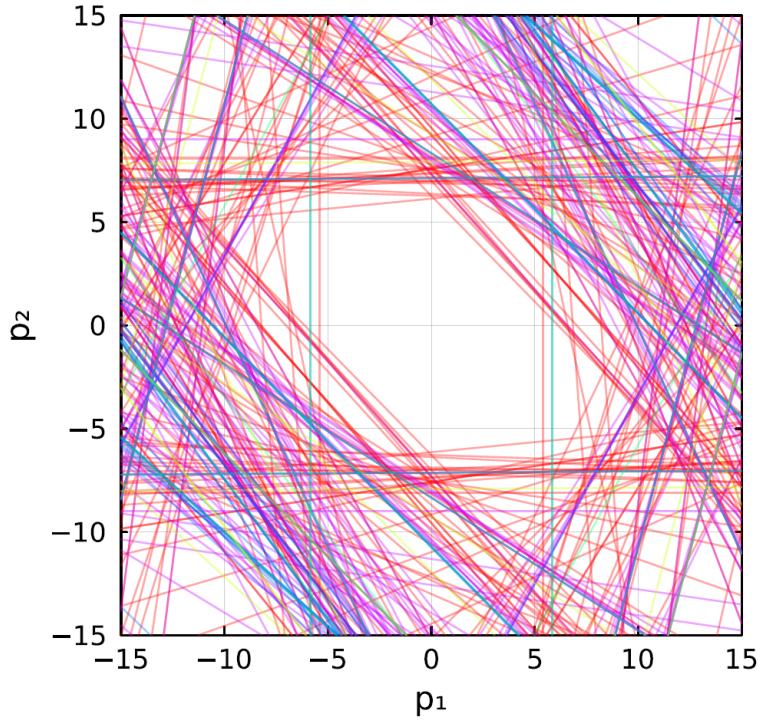


FIGURE 3.1 Projection 2D du polytope admissible  $\mathcal{F}$  sur pour le réseau à 39 bus.

La figure 3.1 illustre un domaine admissible dans un sous-espace à deux dimensions, formé par projection des contraintes d'écoulement de puissance. Elle met en évidence le caractère irrégulier de la frontière de faisabilité : certaines directions de l'espace sont faiblement contraintes, tandis que d'autres sont strictement limitées, parfois de manière contre-intuitive.

Cette géométrie polytopique traduit, dans un cas simplifié, la complexité que l'on retrouve dans l'espace complet des charges nodales. L'inclusion  $\mathcal{F}_k \subset \mathcal{F}_{k+1}$  est garantie par construction, mais l'écart entre niveaux peut être très variable selon la direction explorée. La figure 3.1 permet

ainsi d'illustrer pourquoi les frontières entre niveaux successifs ne peuvent être approchées de façon uniforme, et pourquoi une stratégie de génération aléatoire serait inefficace pour couvrir l'ensemble des niveaux  $k^*$ .

### 3.3.2 Stratégie d'exploration adaptative pour la génération de profils

Afin de générer un jeu d'apprentissage couvrant l'ensemble des niveaux hiérarchiques  $k \in \{1, \dots, k_{\max}\}$ , nous mettons en œuvre une stratégie d'exploration adaptative, reposant sur des mécanismes de génération de vecteurs de charges locaux et globaux, combinés à une politique de couverture dynamique. Cette procédure se décompose en deux phases principales : une phase d'initialisation unique, puis une boucle d'exploration incrémentale.

#### I. Phase d'initialisation

1. Chargement du cas réseau et des données topologiques ;
2. Construction des matrices  $\Phi^{(c)}$  associées aux scénarios  $c \in \mathcal{C} \cup \{0\}$  ;
3. Détermination de la hiérarchie  $\{\mathcal{I}_k\}_{k=1}^{k_{\max}}$  par application de l'algorithme HIERARCHICALCONSTRAINTEXTRACTION (Algorithme 3) ;
4. Définition des quotas par niveau : nombre minimal et maximal d'échantillons souhaités pour chaque  $k$ .

#### II. Boucle d'exploration

À chaque itération, l'algorithme sélectionne un niveau hiérarchique  $k_{\text{cible}}$  dont le quota minimal n'est pas encore atteint, puis génère un profil de charge  $\mathbf{p}_d$  selon l'une des stratégies suivantes :

- Perturbation locale dirigée autour d'un profil existant, en modulant de manière aléatoire ses composantes par un facteur  $\alpha \sim N(1.0, \sigma^2)$ , avec  $\sigma \in [0.01, 1.0]$ , où  $N$  désigne la variable aléatoire gaussienne ;
- Interpolation linéaire entre deux profils de niveaux différents, non-consécutifs, selon la règle :

$$\mathbf{p}_\alpha = (1 - \alpha)\mathbf{p}_{\min} + \alpha\mathbf{p}_{\max}, \quad \alpha \in (0, 1),$$

afin d'explorer les zones de transition entre régions structurellement distinctes ;

- Diversification aléatoire, consistant à sélectionner aléatoirement une stratégie parmi les précédentes lorsque plus de 50% des niveaux sont couverts ;
- Raffinement marginal, dans les zones proches des frontières entre niveaux, en réduisant  $\sigma$  pour générer des perturbations fines.

### III. Évaluation et annotation

Chaque profil candidat est évalué sur le SCED complet. Si le problème est faisable, la solution optimale est utilisée pour identifier le plus haut niveau de contrainte effectivement mobilisé :

$$k^* = \max \{k \in \{1, \dots, k_{\max}\} \mid \exists i \in \mathcal{I}_k \text{ tel que } \xi_i > 0 \text{ et } \lambda_i > 0\},$$

où  $\xi_i$  désigne la variable de pénalité associée à la contrainte  $i$ , et  $\lambda_i$  son multiplicateur de Lagrange. Le profil  $\mathbf{p}_d$  est ensuite validé en résolvant le SCED réduit à  $\mathcal{I}_{1:k^*}$ , et en comparant la solution obtenue à celle du problème complet. Celle-ci s'avère toujours identique.

### IV. Enregistrement et poursuite

Si la solution réduite reproduit la solution complète, le profil  $\mathbf{p}_d$  est enregistré, accompagné de ses caractéristiques optimales  $(\mathbf{p}^*, \boldsymbol{\xi}^*, \boldsymbol{\lambda}^*)$ , du niveau  $k^*$ , ainsi que de métadonnées (stratégie, parent, direction, etc.). La base est mise à jour, et l'algorithme poursuit tant que tous les niveaux hiérarchiques n'ont pas atteint leur quota d'échantillons.

#### 3.3.3 Prétraitement, apprentissage et calibration des modèles

Cette section décrit les étapes de traitement des données, de formulation du problème d'apprentissage supervisé, ainsi que les modalités d'entraînement et de calibration des modèles  $\{\mathfrak{C}_k\}_{k=1}^{k_{\max}}$ .

Dans un premier temps, le jeu de données simulé est converti en une matrice d'entrée structurée  $\mathbf{X} \in \mathbb{R}^{N \times |\mathcal{N}|}$ , où  $N$  est le nombre d'échantillons générés, chaque ligne représentant un profil de charge  $\mathbf{p}_d$ . Les étiquettes entières  $k^* \in \{1, \dots, k_{\max}\}$  associées à chaque exemple sont encodées sous forme binaire cumulative selon la transformation (3.6) de manière à respecter la structure ordonnée des niveaux hiérarchiques. Cette représentation permet d'entraîner indépendamment chaque modèle  $\mathfrak{C}_k$  comme un classifieur binaire sur la cible  $z_k$ .

L'ensemble est ensuite divisé en trois sous-ensembles disjoints (entraînement, validation, évaluation) selon une procédure de stratification multiniveau, assurant une répartition équilibrée de chaque classe  $k^*$  dans chaque split. Cette stratégie préserve la représentativité statistique des niveaux rares tout en garantissant une séparation stricte entre données vues et non vues.

Pour chaque niveau  $k \in \{1, \dots, k_{\max}\}$ , un modèle  $\mathfrak{C}_k$  est entraîné indépendamment pour approximer la fonction (3.7) à l'aide de la méthode LightGBM. L'apprentissage repose sur un sous-ensemble de données étiquetées  $\{(\mathbf{p}_d^{(i)}, z_k^{(i)})\}_{i=1}^N$ , dans lequel chaque profil  $\mathbf{p}_d^{(i)} \in \mathbb{R}^{|\mathcal{N}|}$  est associé à une étiquette binaire  $z_k^{(i)} \in \{0, 1\}$ , indiquant si le niveau  $k$  est requis pour assurer

l'exactitude de la solution du SCED. La fonction de perte utilisée est la perte logarithmique binaire pondérée (*binary log loss*), qui s'écrit, pour un ensemble d'entraînement de taille  $N$  :

$$\mathcal{L}_k = - \sum_{i=1}^N w_1 z_k^{(i)} \log(P_k^{(i)}) + w_0 (1 - z_k^{(i)}) \log(1 - P_k^{(i)}),$$

où  $P_k^{(i)} \in (0, 1)$  désigne la probabilité prédictive par  $\mathfrak{C}_k$  pour l'échantillon  $i$ , et  $(w_0, w_1) \in \mathbb{R}_+^2$  sont des poids de classes utilisés pour corriger le déséquilibre entre étiquettes positives et négatives. Ces poids sont définis comme inverses des fréquences empiriques des classes dans les données d'entraînement :

$$w_0 = \frac{1}{N_0}, \quad w_1 = \frac{1}{N_1},$$

où  $N_0 = \sum_{i=1}^N \mathbb{1}_{[z_k^{(i)}=0]}$  et  $N_1 = \sum_{i=1}^N \mathbb{1}_{[z_k^{(i)}=1]}$ . Cette pondération permet de préserver une contribution équitable de chaque classe à la descente de gradient, même dans les niveaux où  $z_k = 1$  est un événement rare [75].

L'optimisation de  $\mathcal{L}_k$  repose sur une procédure itérative de type *boosting* par gradient, dans laquelle une séquence de modèles faibles, construits sous forme d'arbres de décision, est agrégée pour approximer la fonction cible [74]. À chaque itération, un nouvel arbre est entraîné pour minimiser une approximation quadratique locale de la perte  $\mathcal{L}_k$ , en s'appuyant sur les gradients et ces hessiennes évalués aux prédictions courantes. L'entraînement est effectué via une validation croisée, permettant d'évaluer la stabilité du modèle et de prévenir le sur-apprentissage. Les hyperparamètres du modèle LightGBM sont calibrés de manière indépendante pour chaque niveau  $k$ , via une recherche exhaustive en grille (`search_cv`) sur l'ensemble d'entraînement. Les paramètres optimisés incluent :

- `learning_rate` (taux d'apprentissage),
- `num_leaves` (nombre maximal de feuilles par arbre, contrôlant la non-linéarité),
- `max_depth` (profondeur maximale autorisée),
- `feature_fraction` et `bagging_fraction` (ratios de sous-échantillonnage horizontal et vertical),
- `lambda_11`, `lambda_12` (régularisations  $\ell_1$  et  $\ell_2$ ).

Cette calibration d'hyperparamètres, menée niveau par niveau, permet d'adapter chaque modèle  $\mathfrak{C}_k$  aux propriétés spécifiques de sa tâche, en tenant compte notamment de la rareté relative des étiquettes  $z_k = 1$ , plus marquée pour les niveaux élevés. Les performances sont évaluées à l'aide de métriques de classification (faux positifs, rappel,  $F_1$ ) sur l'ensemble de validation, puis consolidées en phase finale sur le jeu de test.

### 3.3.4 Procédure d'inférence et de résolution

Finalement, la chaîne d'apprentissage supervisée ainsi construite permet d'orchestrer une procédure d'inférence pour la résolution partielle du SCED pénalisé. À partir d'un profil de charge nodale  $\mathbf{p}_d$ , la famille de classificateurs hiérarchiques  $\{\mathfrak{C}_k\}_{k=1}^{k_{\max}}$  produit une séquence de décisions binaires, à partir desquelles est déduite une prédiction de niveau

$$\hat{k}^* := \max_k \{k \in \{1, \dots, k_{\max}\} \mid \mathfrak{C}_k(\mathbf{p}_d) = 1\},$$

interprétée comme le plus petit niveau hiérarchique jugé suffisant pour garantir l'exactitude de la solution du SCED réduit. Le problème d'optimisation est alors résolu en ne considérant explicitement que les contraintes appartenant aux niveaux  $\mathcal{I}_{1:\hat{k}^*}$ , réduisant ainsi la taille du système à résoudre.

Une fois la solution obtenue, une évaluation *a posteriori* est conduite pour vérifier si des contraintes non considérées, i.e., appartenant à  $\mathcal{I} \setminus \mathcal{I}_{1:\hat{k}^*}$ , sont non satisfaites strictement. En cas de non-satisfaction, un mécanisme de reprise incrémentale est déclenché. Le niveau hiérarchique est alors augmenté de façon adaptative, et une nouvelle résolution est effectuée avec les contraintes étendues jusqu'au premier niveau assurant la validité complète. Ce processus garantit la robustesse de la méthodologie, tout en conservant les bénéfices de la réduction lorsque la prédiction est correcte.

## 3.4 Conclusion

La méthode présentée dans ce chapitre s'inscrit dans une perspective de réduction structurée des contraintes thermiques de ligne post-contingence, adaptée à la résolution de SCED pénalisés. En combinant une extraction hiérarchique récursive des inégalités structurellement nécessaires, un modèle de classification supervisée fondé sur une factorisation cumulative binaire, et un mécanisme adaptatif de reprise en phase d'inférence, elle permet de réduire le nombre de contraintes effectivement injectées dans le solveur, tout en garantissant l'exactitude de la solution finale.

Le principal atout de cette approche réside dans sa capacité à anticiper, pour chaque profil de charge, le plus petit sous-ensemble pertinent de contraintes à considérer. Contrairement aux méthodes de réduction classiques, la présente méthode ajuste dynamiquement le niveau de détail requis à la structure du profil observé, assurant un compromis robuste entre efficacité algorithmique et couverture de sécurité.

Sur le plan théorique, cette méthodologie tire bénéfice des concepts de réduction de re-

dondance géométrique, partitions de polytope, et sur l'encodage hiérarchique. Sur le plan computationnel, elle exploite la performance des modèles LightGBM, capables de s'intégrer dans une architecture de résolution temps réel, sans altérer la robustesse globale à l'aide du paramètre de décision,  $\tau$ , offrant la possibilité d'être conservateur dans la prédiction, puis grâce au mécanisme de reprise. La séparation explicite entre phase hors-ligne (extraction de la hiérarchie, entraînement des modèles) et phase en-ligne (prédiction, résolution, reprise) en fait une solution adaptée aux exigences opérationnelles des marchés électriques contemporains.

Le chapitre suivant propose une évaluation expérimentale de cette méthodologie. Il examine, sur plusieurs réseaux tests issus de la littérature, la capacité du modèle à réduire efficacement la taille des instances SCED, tout en maintenant l'exactitude de la solution. Les résultats sont analysés sous l'angle du gain de computationnel, de la qualité de prédiction du niveau hiérarchique  $k^*$ , et de la fréquence effective d'activation du mécanisme de reprise.

## CHAPITRE 4 RÉSULTATS NUMÉRIQUES

Ce chapitre présente une évaluation numérique approfondie de la méthode proposée, en mobilisant des cas tests standards issus de la littérature sur les réseaux électriques, et en reproduisant intégralement le pipeline de réduction, d'apprentissage et d'inférence décrit aux chapitres précédents.

L'ensemble des méthodes ont été implémentées dans le langage de programmation Julia [76], et les expériences ont été menées sur une machine équipée d'un processeur 12<sup>th</sup> Gen Intel(R) Core(TM) i7-12700H cadencé à 2.30 GHz, avec 16 Go de mémoire vive, sous un système d'exploitation 64 bits à architecture *x64*. Aucun recours à l'accélération GPU ou à des clusters de calcul n'a été nécessaire pour les simulations présentées.

Les résultats présentés s'appuient exclusivement sur des profils de charge synthétiques générés à partir de cas tests synthétiques IEEE établis par **Power Grid Lib** [77]. En pratique, la procédure précédente peut s'appliquer à des profils de charge historiques réels, tels que ceux disponibles dans les rapports de type *State Estimator Load Reports* diffusés par ERCOT [78]. Dans le présent travail, nous faisons le choix délibéré d'utiliser des données académiques simulées, issues de cas tests standards, afin de garantir un contrôle précis sur la distribution des charges  $\mathbf{p}_d$  et une couverture effective de l'ensemble des niveaux hiérarchiques  $k^*$ , en utilisant la méthode de génération du jeu de données présentée à la section 3.3.1. Ce choix permet notamment de contourner les limitations statistiques inhérentes aux données empiriques, qui tendent à refléter une dynamique opérationnelle spécifique et concentrée sur une région restreinte de l'espace des injections.

Les sections suivantes décrivent successivement le protocole expérimental adopté, les performances du classifieur sur un réseau IEEE, ainsi que les gains observés en termes de complexité de résolution. Une attention particulière est portée à la capacité du modèle à préserver l'exactitude des solutions SCED tout en réduisant significativement le nombre de contraintes mobilisées.

### 4.1 Parallélisation de la détection de contraintes redondantes

Chaque itération de **HIERARCHICALCONSTRAINTEXTRACTION**, présenté à l'Algorithme 3, repose sur l'appel de la méthode **REDUNDANCYREMOVAL**, présentée à l'Algorithme 1, dont l'objectif est d'extraire un sous-ensemble minimal de contraintes strictement nécessaires parmi les limites d'écoulement de puissance indexées par  $i \in \mathcal{I} \setminus \bigcup_{j=1}^{k-1} \mathcal{I}_j$ , pour construire le niveau

hiérarchique  $\mathcal{I}_k$ . Afin d'assurer une mise à l'échelle effective sur des architectures multicœurs, cet algorithme est implémenté selon une architecture multitraitements par fils d'exécution (*multi-thread*) fondée sur une décomposition dynamique de l'espace des contraintes.

Formellement, à chaque niveau  $k$ , l'ensemble résiduel, défini comme  $\mathcal{R}_k := \mathcal{I} \setminus \bigcup_{s=1}^{k-1} \mathcal{I}_s$ , et est partitionné en segments disjoints  $\{\mathcal{R}_k^{(t)}\}_{t=1}^T$ , chacun étant traité indépendamment sur un fil de traitement parallèle (*thread*)  $t \in \{1, \dots, T\}$ . Pour chaque contrainte candidate  $i \in \mathcal{R}_k^{(t)}$ , le fil local exécute un test de type LP pour déterminer si cette contrainte est structurellement indispensable relativement aux autres contraintes du segment. Le problème considéré s'écrit alors :

$$\max_{\mathbf{p} \in \mathbb{R}^{|\mathcal{N}|}} \Phi_i^\top \mathbf{p} \quad (4.1a)$$

$$\text{s.c.} \quad \Phi_j^\top \mathbf{p} \leq \bar{s}_j, \quad \forall j \in \mathcal{R}_k^{(t)} \setminus \{i\} \quad (4.1b)$$

$$\Phi_i^\top \mathbf{p} \leq \bar{s}_i + 1. \quad (4.1c)$$

Si la solution optimale  $\mathbf{p}^*$  satisfait  $\Phi_i^\top \mathbf{p}^* > \bar{s}_i$ , alors la contrainte  $i$  est déclarée potentiellement redondante et un balayage directionnel (*rayshoot*) est déclenché depuis un point strictement intérieur, typiquement  $\mathbf{z} = \mathbf{0}$ , dans la direction  $\mathbf{p}^*$ . Ce procédé permet d'identifier de manière robuste la première contrainte effectivement franchie. L'ensemble des contraintes jugées essentielles au niveau  $k$  est ensuite agrégé sans contention, en exploitant la séparation mémoire naturelle entre fils de calcul. Cette stratégie itérative exploite à la fois la structure additive des contraintes reliées à la matrice PTDF et la stabilité géométrique de l'enveloppe admissible définie par les niveaux précédents, c'est-à-dire l'ensemble

$$\mathcal{F}_{1:k-1} := \bigcap_{i \in \mathcal{I}_{1:k-1}} \left\{ \mathbf{p} \in \mathbb{R}^{|\mathcal{N}|} \mid \Phi_i \mathbf{p} \leq \bar{s}_i \right\},$$

à partir duquel est déterminé le résidu hiérarchique  $\mathcal{R}_k$ . Cette stabilité du polytope de l'ensemble de faisabilité permet une décomposition parallèle du test de redondance, tout en garantissant l'indépendance des segments disjoints  $\{\mathcal{R}_k^{(t)}\}_{t=1}^T$  traités.

Afin d'améliorer l'allocation des ressources de calcul, la stratégie de parallélisation adopte une logique adaptative fondée sur une réduction progressive du nombre de segments traités en parallèle. Plus précisément, à chaque itération de l'algorithme de filtrage, l'ensemble résiduel  $\mathcal{R}_k$  est partitionné en  $T_k$  sous-ensembles  $\{\mathcal{R}_k^{(t)}\}_{t=1}^{T_k}$ , chacun traité indépendamment sur un fil d'exécution. Le nombre de segments  $T_k$  est initialisé à une valeur proportionnelle au nombre de coeurs physiques disponibles, puis réduit d'un facteur 2 à chaque itération selon la

règle

$$T_k^{(l+1)} := \left\lfloor \frac{T_k^{(l)}}{2} \right\rfloor,$$

où  $l$  désigne l'indice d'itération interne de la procédure de filtrage. Ceci permet de limiter la surcharge des ressources mémoire partagées et de préserver la stabilité thermique et énergétique du système hôte tout en maintenant une efficacité de parallélisme adaptée aux ressources computationnelles. Ce mécanisme de réallocation adaptative, qui s'inscrit dans le paradigme de filtrage progressif (*filter-reduce*), vise à hiérarchiser l'effort de calcul en concentrant progressivement les ressources de traitement sur les contraintes dont le caractère non redondant ne peut être établi qu'au prix d'un test approfondi. Autrement dit, les contraintes dites résistantes sont celles qui ne peuvent être écartées par les tests locaux de redondance, menés sur des segments réduits et parallélisés du problème initial, combinant résolution du *LP-test* et *rayshoot*. Dès les premières itérations, une fraction importante de contraintes manifestement redondantes est éliminée à coût marginal, ce qui réduit significativement la taille du problème à considérer lors des phases ultérieures. Cette stratégie de réduction hiérarchique permet ainsi d'optimiser le temps de convergence tout en préservant l'exactitude structurelle garantie par l'algorithme de détection rigoureuse des redondances.

La phase finale consiste en une passe de validation stricte, au cours de laquelle l'ensemble résiduel des contraintes présélectionnées est réagrégé, puis redistribué entre les *threads* pour une évaluation parallèle. Chaque thread reconstruit localement une instance du modèle d'optimisation et résout un *LP-test* afin de vérifier si la contrainte considérée borne effectivement la région admissible. L'écriture des résultats s'effectue dans des vecteurs booléens indexés, répartis de manière disjointe entre les *threads*, ce qui autorise une synchronisation implicite sans verrou explicite (*lock-free*), garantissant ainsi la sûreté et l'efficience de l'agrégation finale.

Cette architecture, fondée sur une double parallélisation (initiale et finale) et un découpage dynamique des tâches, permet une mise à l'échelle sur des systèmes multicoeurs, tout en garantissant la rigueur de l'élimination. Elle s'avère particulièrement adaptée au traitement des contraintes post-contingence issues des matrices PTDF,  $\Phi^{(c)}$ , dont la densité structurelle justifie une gestion intelligente des ressources computationnelles.

Pour l'ensemble des tâches d'optimisation linéaire et quadratique intervenant dans la détection de redondance ainsi que dans la résolution des problèmes SCED, nous avons retenu le solveur **HIGHS** [79, 80]. Conçu pour les applications à grande échelle, ce solveur *open-source* implémente deux algorithmes fondamentaux : une méthode du simplexe dual compatible avec le traitement parallèle, ainsi qu'une méthode IPM, reposant sur un résolveur linéaire itératif. Ces deux

approches couvrent la totalité des besoins requis par notre méthode. De plus, son intégration native dans l'écosystème Julia, via l'interface `HiGHS.jl`, facilite son incorporation directe au sein des modules de résolution JuMP [81], tout en conservant une transparence complète sur les paramètres internes du solveur.

Enfin, les performances obtenues par `HiGHS` dans les comparaisons récentes avec les solveurs commerciaux établis [82] en font une solution pertinente, combinant efficacité numérique, compatibilité au traitement parallèle, licence ouverte, et reproductibilité scientifique.

## 4.2 Résultats de la construction hiérarchique de contraintes

Les résultats présentés dans le tableau 4.1 permettent d'attester empiriquement de la pertinence de la réduction hiérarchique introduite au chapitre 3. Pour chaque cas testé, l'ensemble complet des contraintes d'écoulement de puissance agrégé à partir de l'ensemble des contingences  $\mathcal{C} \cup \{0\}$  est partitionné en une séquence hiérarchique croissante  $\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_{k_{\max}}$ . Chaque niveau  $\mathcal{I}_k$  regroupe un sous-ensemble non redondant de contraintes jugées structurellement nécessaires, et la résolution du SCED restreint à l'union cumulée  $\mathcal{I}_{1:k^*} := \bigcup_{k=1}^{k^*} \mathcal{I}_k$  est comparée à celle du modèle complet incluant l'ensemble  $\mathcal{I} := \mathcal{I}_{1:k_{\max}}$ .

La structure exacte de chaque hiérarchie testée est présentée ci-dessous :

- `case9` :  $\{35, 39, 12, 8, 2\} \Rightarrow |\mathcal{I}| = 96$ ,
- `case39` :  $\{719, 793, 273, 113, 88\} \Rightarrow |\mathcal{I}| = 1986$ .
- `case118` :  $\{2613, 17671, 22350, 5100, 1491, 2120, 863, 192, 23, 6, 1\} \Rightarrow |\mathcal{I}| = 52424$ .

Les colonnes du tableau 4.1 présentent les mesures clés associées à l'évaluation des modèles complet et réduit. Le temps de résolution du SCED complet, incluant l'ensemble des contraintes  $\mathcal{I}$ , est noté  $T_{\text{full}}$ , tandis que  $T_{\text{red}}$  désigne le temps requis pour résoudre le SCED restreint aux seules contraintes du niveau  $\mathcal{I}_{1:k^*}$ . Les colonnes  $\Delta T$  et  $\Delta T \%$  indiquent respectivement le gain de temps absolu et relatif obtenu grâce à cette réduction. Les quantités  $J_{\text{full}}$  et  $J_{\text{red}}$  correspondent aux valeurs optimales de la fonction objectif dans les formulations complète et réduite. Le nombre total de contraintes d'écoulement initialement considérées est donné par  $|\mathcal{I}|$ , tandis que  $|\mathcal{I}_{1:k^*}|$  indique le nombre de contraintes effectivement retenues dans la version réduite. La différence entre ces deux quantités est représentée par  $\Delta \mathcal{I}$  (absolue) et  $\Delta \mathcal{I} \%$  (relative), traduisant le nombre et la part des contraintes éliminées. Enfin,  $k_{\max}$  désigne la profondeur maximale de la hiérarchie extraite pour le réseau donné, et  $k^*$  indique le niveau hiérarchique jugé suffisant pour garantir une solution optimale identique à celle du problème complet.

TABLEAU 4.1 Comparaison entre SCED complet et formulation hiérarchiquement restreinte

Réseau	$T_{\text{full}}$	$T_{\text{red}}$	$\Delta T$	$\Delta T [\%]$	$J_{\text{full}}$	$J_{\text{red}}$	$ \mathcal{I} $	$ \mathcal{I}_{1:k^*} $	$\Delta \mathcal{I}$	$\Delta \mathcal{I} [\%]$	$k_{\max}$	$k^*$
case9	0,008	0,004	0,003	37,5	5781	5781	96	35	61	63,5	5	1
case39	0,305	0,179	0,126	41,31	215151	215151	1986	1610	1232	18,9	5	2
case118	2,296	1,93	0,366	15,94	3602	3602	52424	49225	3199	6,10	11	5

Dans tous les cas testés, à partir de profiles de charge  $\mathbf{p}_d$  conçus pour cet exemple, la valeur optimale du SCED réduit  $J_{\text{red}}$  coïncide numériquement avec celle du modèle complet  $J_{\text{full}}$ , confirmant que le niveau  $k^*$  retenu suffit à garantir l'exactitude de la solution. Cette équivalence atteste empiriquement de la cohérence géométrique des régions admissibles  $\mathcal{F}_{1:k}$  construites à partir des ensembles hiérarchiques  $\mathcal{I}_{1:k}$ , et valide la capacité du mécanisme de sélection à identifier un sous-ensemble structurellement suffisant de contraintes thermiques.

Au-delà de cette validité structurelle, les gains computationnels observés sont substantiels. Pour le cas **case9**, la hiérarchie atteint une profondeur  $k_{\max} = 5$ , mais un unique niveau  $k^* = 1$  suffit pour contenir l'optimum. Sur les  $|\mathcal{I}| = 96$  contraintes initiales, seules 35 sont effectivement mobilisées, soit une réduction de 63,5 %, qui se traduit par une amélioration du temps de résolution de 0,008 secondes pour le modèle complet à seulement 0,004 secondes pour le modèle réduit, soit un gain relatif de 37,5 %.

Pour le réseau plus complexe **case39**, structuré en 1986 contraintes aussi réparties sur 5 niveaux hiérarchiques, le niveau utile reste minimal ( $k^* = 2$ ), et la réduction atteint 18,9 %, avec seulement 1610 contraintes effectivement utilisées. Le temps de calcul passe alors de 0,305 à 0,179 secondes, soit une amélioration de 41,31 %.

Le réseau **case118**, de taille plus importante, montre que malgré la complexité accrue du modèle, caractérisée par un ensemble initial de 52 424 contraintes réparties sur  $k_{\max} = 11$  niveaux hiérarchiques, la méthode parvient à identifier un niveau réduit  $k^* = 5$  suffisant pour garantir l'exactitude de la solution. Cette sélection hiérarchique permet d'écartier 3199 contraintes structurellement non essentielles, soit une réduction relative de 6,1 %, tout en préservant la totalité de l'information utile à l'optimisation. Ce filtrage partiel se traduit par un gain de temps de l'ordre de 15,9 %, avec une durée de résolution passant de 2,296 à 1,930 secondes. Ce résultat met en évidence deux éléments distincts. D'une part, la hiérarchie conserve sa capacité à identifier les contraintes strictement essentielles, même dans des réseaux plus grands. Cependant, l'efficacité de la hiérarchisation décroît lorsque le niveau utile  $k^*$  coïncide avec un niveau de la hiérarchie regroupant une proportion significative des contraintes. Dans le cas de **case118**, la hiérarchie présente une répartition très hétérogène des contraintes par niveau, avec notamment 17671, 22350 et 5100 contraintes concentrées respectivement aux

niveaux 2, 3 et 4. Ainsi, le fait que le niveau requis  $k^* = 5$  survienne juste après ces niveaux massifs limite fortement le potentiel de réduction : 49225 contraintes sur 52424 restent utiles, soit plus de 93 % du modèle initial. Ce choix de scénario, volontairement construit pour activer un niveau hiérarchique qui implique tout de même un grand nombre de contraintes, permet d'illustrer une faiblesse intrinsèque de la méthode.

Il convient toutefois de noter, pour ces 2 derniers exemples, que si le niveau utile s'était avéré plus bas, e.g.  $k^* = 1$ , alors l'effet de réduction aurait été plus significatif. Notamment, dans le cas du réseau **case118**, le niveau 1 ne comporte que 2613 contraintes, soit moins de 0,05 % de l'ensemble initial.

Dans la suite de cette section, nous procérons à une analyse approfondie des comportements observés sur le cas intermédiaire **case39**, retenu comme instance de référence pour l'étude expérimentale. Ce choix permet de mettre en lumière, sur un réseau moins réaliste mais tractable, les mécanismes et limites pratiques de la hiérarchisation supervisée. L'extension à des réseaux de plus grande taille constitue une perspective d'amélioration future de la méthode pour une mise à l'échelle industrielle. Cette généralisation apparaît d'autant plus envisageable que les données requises sur les profils de charge nodale historiques, et sur les topologies de réseau, matrices PTDF, sont accessibles en pratique.

La figure 4.1 illustre la distribution empirique des gains de temps relatifs obtenus par la résolution du SCED réduit, en fonction du niveau hiérarchique  $k^*$  identifié comme suffisant. L'analyse repose sur un ensemble de 2125 instances résolues à partir du cas réseau **case39**, pour lequel la hiérarchie maximale atteint  $k_{\max} = 5$ . Pour chaque instance, le gain est mesuré par la réduction relative du temps de résolution, défini comme

$$\Delta T(k^*) := \frac{T_{\text{full}} - T_{\text{red}}^{(k^*)}}{T_{\text{full}}} \times 100,$$

où  $T_{\text{full}}$  désigne le temps de résolution du modèle complet, et  $T_{\text{red}}^{(k^*)}$  celui du modèle réduit au niveau  $k^*$ .

Les résultats révèlent une décroissance nette et structurée du gain moyen en fonction de la profondeur hiérarchique, obtenue par évaluation directe du niveau hiérarchique requis, i.e, sans prédiction. Lorsque la solution optimale repose uniquement sur les contraintes du premier niveau ( $k^* = 1$ ), le gain moyen de temps atteint 68,1 %, avec une dispersion modérée ( $\sigma \approx 12,8\%$ ). Ce comportement reflète la réduction de complexité induite par la hiérarchie dans ses niveaux bas, où la majorité des contraintes inutiles sont exclues. À mesure que le niveau hiérarchique  $k^*$  augmente, le nombre de contraintes essentielles croît, réduisant ainsi l'écart entre la formulation réduite et le modèle complet. Les gains moyens chutent ainsi à

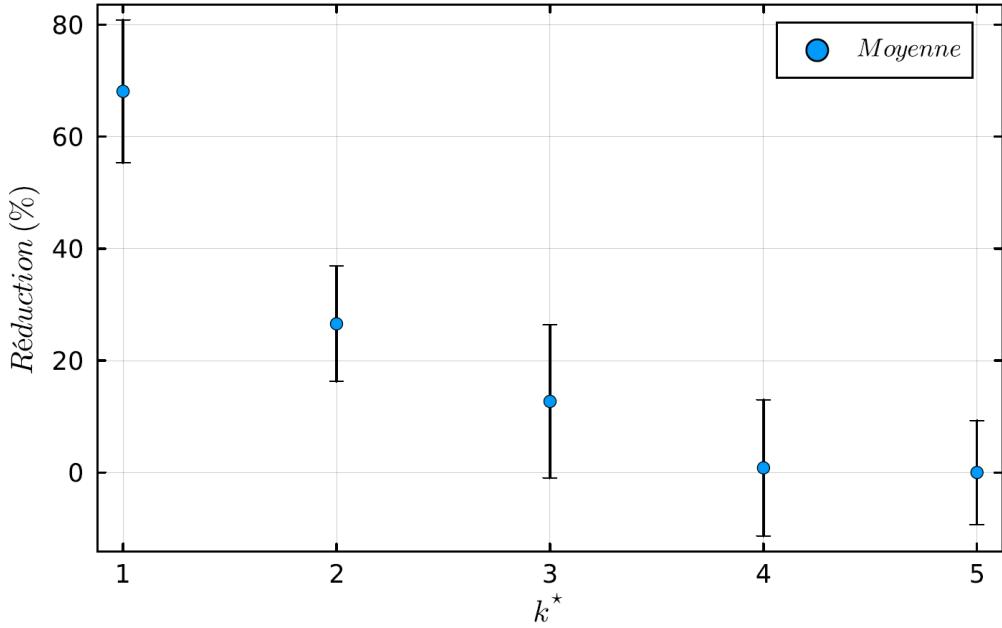


FIGURE 4.1 Réduction moyenne du temps de calcul par niveau hiérarchique  $k^*$ .

26,6 % pour  $k^* = 2$ , puis à 12,7 % pour  $k^* = 3$ , avec une variabilité des écarts-types respectifs croissante de 10,3% et 13,7%. Ce phénomène traduit la sensibilité du solveur à la structure de la matrice des contraintes, notamment à sa densité croissante au fur et à mesure que de nouvelles contraintes sont réintroduites. Les niveaux  $k^* = 4$  et  $k^* = 5$ , proches du cas complet, se distinguent par des gains marginaux voire négatifs. Le niveau  $k^* = 4$  présente un gain moyen de seulement 0,8 % avec un écart-type significatif ( $\sigma \approx 12,2\%$ ), tandis que  $k^* = 5$ , correspondant au modèle intégral, n'induit par construction aucun gain moyen, mais affiche néanmoins un écart-type non nul ( $\sigma \approx 9,3\%$ ). Ces écarts témoignent de cas où la résolution du modèle réduit s'avère plus lente que celle du modèle complet. Cette contre-performance met en évidence l'instabilité numérique du solveur dans les configurations proches de la saturation, e.g., lorsque le nombre de contraintes  $|\mathcal{I}|$  est de 1986, où la densité et la condition numérique des matrices  $\Phi_i$  peuvent altérer la vitesse de convergence de l'algorithme d'optimisation, malgré un nombre légèrement réduit de contraintes. Ce constat souligne que, dans les niveaux élevés, l'allègement structurel ne suffit pas à garantir un gain de performance, et que les bénéfices de la réduction hiérarchique sont d'autant plus marqués que  $k^*$  est faible. Il justifie en pratique l'intérêt d'une prédiction précise de  $k^*$ , afin d'éviter la considération inutilement des niveaux où les effets combinés du solveur et de la structure matricielle peuvent inverser le gain espéré.

Ces résultats empiriques démontrent la capacité de la structure hiérarchique à instaurer une gradation effective du compromis entre précision et complexité, en réduisant significativement

le temps de calcul pour les instances simples, résolubles avec un nombre restreint de contraintes. Ils mettent également en évidence la possibilité de substituer de manière robuste au SCED complet une formulation réduite, sans perte de précision, dès lors que le niveau hiérarchique utile  $k^*$  est correctement identifié. La stratégie de prédiction développée à la section 4.3 vise précisément à exploiter cette variabilité. En apprenant à associer à chaque profil nodal  $\mathbf{p}_d$  un niveau  $k^*$  adapté, nous obtenons un levier computationnel pour ajuster dynamiquement le périmètre de résolution. Nous analysons dans la section suivante 4.3 les performances de ce modèle d'apprentissage supervisé.

### 4.3 Résultats du modèle de prédiction de niveau hiérarchique de contrainte

Les résultats présentés dans cette section visent à évaluer les performances du modèle d'apprentissage supervisé introduit à la section 4.3, chargé de prédire le niveau hiérarchique  $k^*$  minimal garantissant l'exactitude de la solution du SCED réduit. L'ensemble de données utilisé pour l'entraînement et la validation provient d'un processus d'exploration adaptative appliqué au cas test `case39`, selon les modalités décrites au chapitre précédent.

La distribution des étiquettes  $k^*$  obtenues dans l'échantillon présente une forte concentration autour des classes intermédiaires, avec une majorité des profils associés au niveau  $k = 3$  (52,89 %) et une plus faible représentation du niveau extrême  $k = 5$ . En particulier, la classe  $k = 5$ , correspondant à l'activation de l'ensemble complet des contraintes, ne contient que 7 échantillons sur un total de plus de 21 000, soit moins de 0,05 %. Cette rareté, conséquence directe de la difficulté à générer de tels profils par exploration, empêche toute modélisation robuste pour cette classe.

Par conséquent, seuls les niveaux  $k \in \{1, 2, 3, 4\}$ , représentant chacun au moins 5 % des données, ont été conservés dans la phase de modélisation. De plus, le niveau  $k = 1$  étant inclus par construction dans tous les profils admissibles, sa présence ne résulte jamais d'une décision isolée, mais découle de la non-activation des niveaux supérieurs. Il n'est donc pas pertinent de lui associer un classifieur dédié, et aucun modèle  $\mathfrak{C}_1$  n'a été entraîné. Les résultats qui suivent évaluent ainsi la capacité des modèles restants à discriminer efficacement les niveaux hiérarchiques  $k \in \{2, 3, 4\}$ , tout en respectant la structure cumulative de la tâche.

L'ensemble de données a été partitionné en deux sous-ensembles disjoints, selon une séparation 80%/20% entre les phases d'apprentissage et d'évaluation. Les modèles  $\{\mathfrak{C}_k\}_{k=1}^{k_{\max}}$  ont été entraînés indépendamment sur les données d'apprentissage, avec une validation croisée à 5 plis permettant d'optimiser les hyperparamètres du modèle LightGBM.

Les résultats du tableau 4.2 attestent de la précision de la procédure de prédiction, avec un

taux d'assignation exacte supérieur à 94,2 %, et une erreur absolue moyenne de seulement 0,061 niveau hiérarchique. Cette performance confirme que le modèle parvient à estimer efficacement le niveau minimal  $\hat{k}^*$  suffisant pour garantir l'exactitude de la solution. La distribution des erreurs est également instructive : les sous-prédictions ne représentent que 3,5 % des cas, tandis que les sur-prédictions atteignent 2,25 %. Nous observons une quasi-symétrie des erreurs moyennes ( $\approx 1,06$  niveaux) quel que soit le sens de l'erreur.

TABLEAU 4.2 Évaluation globale du modèle de prédiction du niveau hiérarchique  $k^*$

Indicateur	Valeur	Unité
Précision exacte	94.28	%
Taux de sous-prédictions	3.48	%
Taux de sur-prédictions	2.25	%
Erreur absolue moyenne	0.061	(niveau)
Erreur moyenne de sur-prédiction	1.063	(niveau)
Erreur moyenne de sous-prédiction	1.061	(niveau)

L'analyse par niveau dans le tableau 4.3 révèle une performance différenciée selon la fréquence des classes. Le niveau  $k^* = 1$ , qui représente environ un quart des données, est prédit correctement dans 99,6 % des cas. Même les niveaux intermédiaires ( $k = 2, 3, 4$ ) affichent des précisions supérieures à 89 %, avec des erreurs strictement bornées à un niveau hiérarchique adjacent, ce qui confirme le bon fonctionnement du codage cumulatif et de l'architecture modulaire de prédiction.

TABLEAU 4.3 Performance du modèle par niveau hiérarchique réel  $k^*$

Niveau	Support (%)	Précision	Sous (%)	Sur (%)	Err. abs	Err. ↓	Err. ↑
1	22.34	99.58	0.00	0.42	0.010	0.000	2.250
2	11.42	89.86	0.00	10.14	0.104	0.000	1.020
3	52.91	93.30	4.83	1.88	0.070	1.065	1.000
4	13.33	93.09	6.91	0.00	0.073	1.051	0.000

D'un point de vue opérationnel, les erreurs de prédiction n'ont pas toutes la même implication. Une sous-prédition du niveau hiérarchique  $k^*$ , c'est-à-dire un cas où le modèle propose un niveau trop bas, peut entraîner l'omission de contraintes critiques et donc compromettre

l'exactitude du problème. Ce type d'erreur active un mécanisme de reprise, présenté à la section , impliquant la résolution incrémentale d'un ou plusieurs problèmes successifs, ce qui alourdit le coût de calcul. À l'inverse, une sur-prédiction ne compromet jamais la validité de la solution puisqu'elle consiste à inclure davantage de contraintes que nécessaire, ce qui peut allonger légèrement le temps de résolution, sans affecter la qualité du résultat.

Les résultats présentés au tableau 4.2 illustrent que la fréquence des sous-prédictions reste contenue, avec un taux de 3,48 % de sous-prédiction, et une erreur absolue moyenne de 1,061 niveaux, comparable à celle observée pour les sur-prédictions (1,063). Cette symétrie indique que les erreurs demeurent bornées à plus ou moins deux niveaux maximum et n'entraînent pas de dérive systématique. Néanmoins, du fait de leur impact structurel sur le processus de résolution, les sous-prédictions doivent être minimisées. Ce constat justifie, dans l'apprentissage, le choix d'un seuil de décision  $\tau$  légèrement conservateur, privilégiant une marge de sécurité sur les classes supérieures afin de limiter les reprises coûteuses.

La figure 4.2 illustre la courbe de convergence de la perte d'entropie croisée binaire (*log cross-entropy loss*) pour les trois modèles de classification  $\mathfrak{C}_k$ , respectivement associés aux niveaux hiérarchiques  $k = 2, 3, 4$ . Chaque modèle de LightGBM est entraîné en minimisant une fonction de perte logistique sur les échantillons binaires associés à la composante  $z_k \in \{0, 1\}$  du vecteur cumulatif. Dans les trois cas, les courbes de convergence présentent une diminution rapide de la perte d'entropie croisée binaire lors des premières itérations, suivie d'un palier relativement stable. Ce comportement traduit une phase d'apprentissage initial efficace, suivie d'une stabilisation progressive. Afin de prévenir tout phénomène de sur-apprentissage lié à une remontée tardive de la perte, un critère d'arrêt anticipé a été appliqué à la 10<sup>e</sup> itération. Cette régularisation précoce permet de préserver un compromis satisfaisant entre biais et variance, en limitant le risque d'ajustement excessif aux fluctuations spécifiques de l'échantillon d'apprentissage.

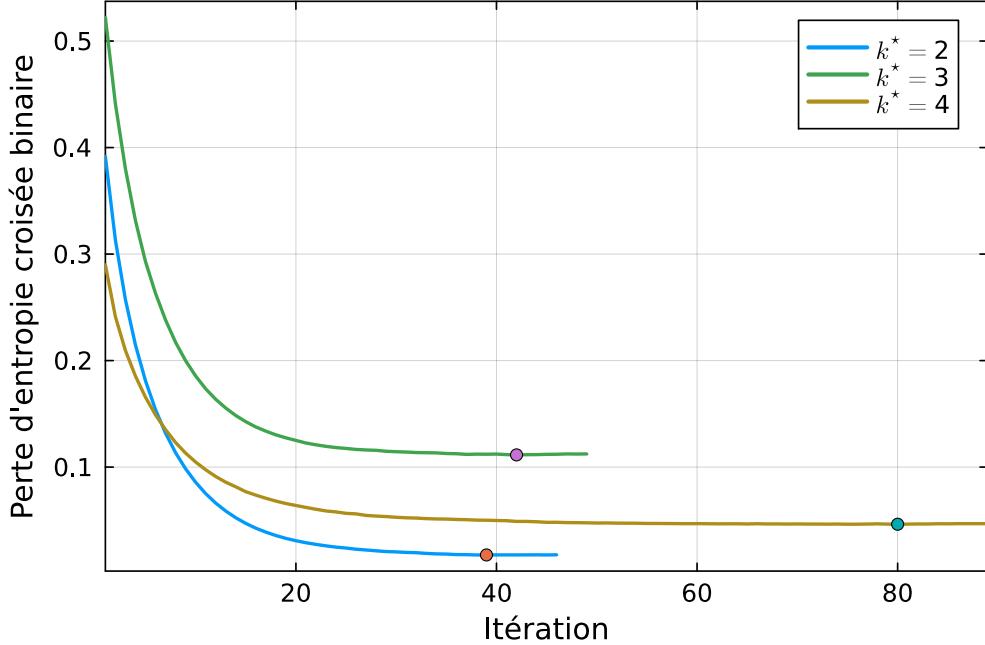


FIGURE 4.2 Courbes de convergence par niveau avec arrêt anticipé (*early stopping*)

Nous observons une décroissance rapide de la perte pour le modèle  $\mathfrak{C}_2$ , qui atteint une stabilité autour de 0,03 dès la 10<sup>e</sup> itération, sans remontée significative par la suite. Cette dynamique traduit une convergence efficace et une capacité discriminante élevée. Ce comportement est attendu, le niveau  $k = 2$  étant à la fois bien représenté dans les données (11,42 % des échantillons) et associé à une frontière de décision relativement marquée entre les classes.

Le modèle  $\mathfrak{C}_3$ , majoritaire en termes de support (52,9 % des observations), converge initialement vers une perte minimale proche de 0,10, avant de présenter une légère remontée au-delà de la 10<sup>e</sup> itération. Ce comportement suggère un début de surapprentissage, probablement induit par la plus grande diversité des profils de charge associés à ce niveau. L'application d'un arrêt anticipé permet ici de prévenir la dérive et de stabiliser la performance généralisable.

Enfin, le modèle  $\mathfrak{C}_4$ , associé au niveau hiérarchique  $k = 4$ , présente des résultats de convergence intermédiaires. La perte décroît rapidement durant les premières itérations, avant d'atteindre un plateau aux alentours de 0,08. Cette tendance met en évidence une capacité d'ajustement correcte du classifieur, tout en soulignant l'intérêt d'un encadrement régulier du nombre d'itérations pour éviter toute instabilité asymptotique.

Dans l'ensemble, ces résultats illustrent la capacité du cadre proposé à prédire avec une bonne fiabilité le niveau hiérarchique optimal  $k^*$  à partir d'un profil de charge nodale. L'encodage binaire cumulatif confère au système de classification une structure modulaire et interprétable,

en permettant d'isoler les contributions décisionnelles de chaque classifieur  $\mathfrak{C}_k$ .

Afin de mieux comprendre les mécanismes internes à la décision des classificateurs hiérarchiques  $\mathfrak{C}_k$ , nous avons conduit une analyse d'explications additives de type Shapley (*SHapley Additive exPlanations*, SHAP) sur chaque modèle [83]. Cette méthode permet de décomposer le score logit associé à un profil de charge donné  $\mathbf{p}_d$  en une somme d'effets additifs spécifiques à chaque variable d'entrée (ici, les noeuds du réseau) :

$$\text{logit}(p_k(\mathbf{p}_d)) = \beta_k + \sum_{j=1}^p \gamma_j^{(k)}(\mathbf{p}_d),$$

où  $\beta_k \in \mathbb{R}$  est la valeur de base du modèle (score moyen en absence d'information), et  $\gamma_j^{(k)}(\mathbf{p}_d) \in \mathbb{R}$  représente la contribution marginale de la charge au noeud  $j$  à la prédiction du niveau  $k$ . Le signe de  $\gamma_j^{(k)}$  indique si la variable fait tendre la décision vers l'activation de  $z_k = 1$  (niveau atteint) ou au contraire l'inhibe, tandis que son amplitude reflète l'intensité de cet effet. L'agrégation des valeurs absolues  $|\gamma_j^{(k)}|$  sur l'ensemble de l'échantillon (*mean absolute SHAP*) permet de produire un classement opérationnel des noeuds les plus influents, fournissant une lecture directe des facteurs structurants la décision à chaque niveau.

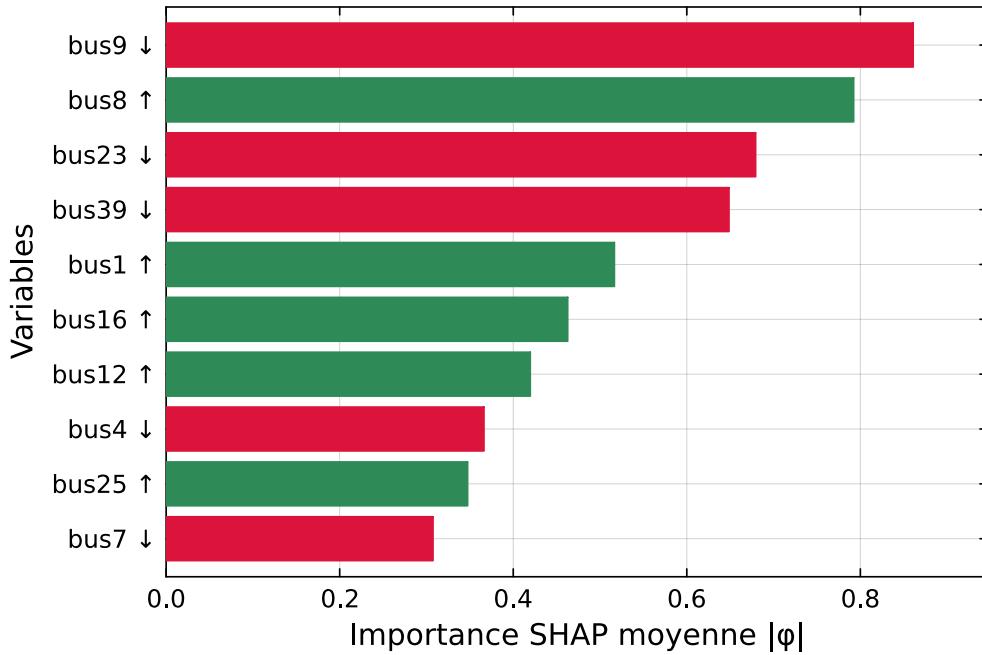


FIGURE 4.3 Niveau 2 — Valeurs SHAP par variable. Longueur des barres = importance moyenne  $|\phi|$ ; flèche ( $\uparrow$  en vert et  $\downarrow$  en rouge) = signe moyen de la contribution. Top 10 des variables classées par importance décroissante.

La figure 4.3 illustre la décomposition des contributions au niveau  $k = 2$ . Les variables

dominantes sont **bus9** (0.8609,  $\downarrow$ ), **bus8** (0.7923,  $\uparrow$ ), **bus23** (0.6793,  $\downarrow$ ) et **bus39** (0.6486,  $\downarrow$ ). La structure de contributions observée est mixte : certaines charges telles que **bus8**, **bus1**, **bus16**, **bus12** ou **bus25** augmentent la probabilité d'activation du niveau 2, tandis que d'autres (**bus9**, **bus23**, **bus39**, **bus4**, **bus7**) exercent une pression inhibitrice. Cette dissymétrie indique que, localement, éviter certaines hausses de charge (notamment **bus9** et **bus23**) est aussi décisif que de renforcer les leviers positifs comme **bus8** ou **bus1**.

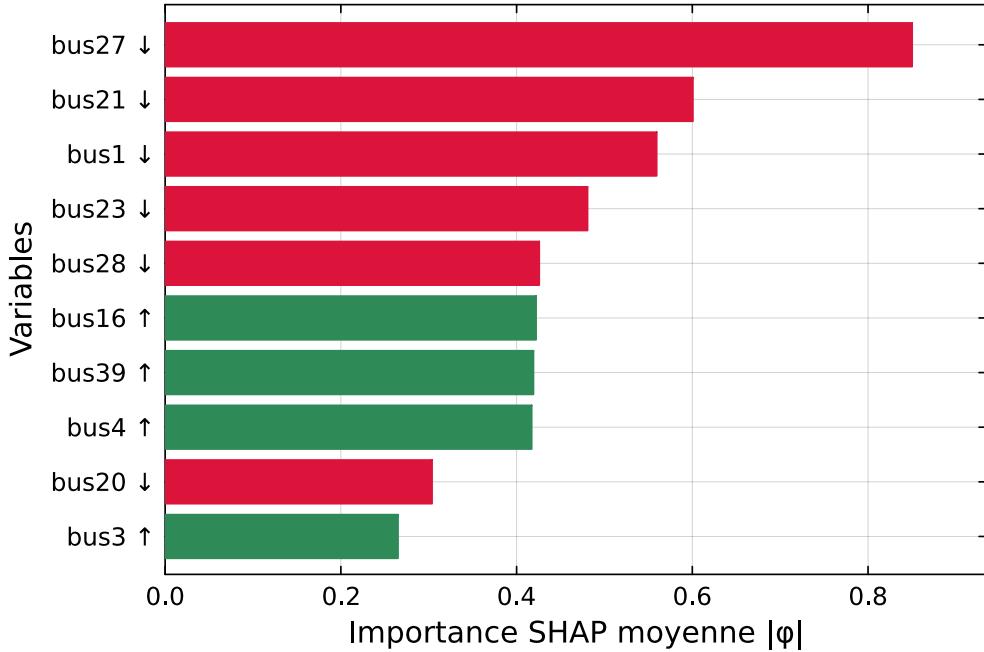


FIGURE 4.4 Niveau 3 — Valeurs SHAP par variable. Longueur des barres = importance moyenne  $|\phi|$ ; flèche ( $\uparrow$  en vert et  $\downarrow$  en rouge) = signe moyen de la contribution. Top 10 des variables classées par importance décroissante.

Au niveau  $k = 3$ , représenté en figure 4.4, les variables les plus influentes sont **bus27** (0.8502,  $\downarrow$ ), **bus21** (0.6007,  $\downarrow$ ), **bus1** (0.5595,  $\downarrow$ ) et **bus23** (0.4808,  $\downarrow$ ). Le profil général est plus négatif que précédemment : la majorité des variables les plus influentes contribuent à bloquer l'activation de ce niveau. Notamment, **bus1**, qui avait un effet positif pour  $k = 2$ , bascule ici en influence négative. D'autres variables comme **bus16**, **bus39** et **bus4** conservent un effet d'activation. Ces inversions de signe signalent que certaines variations de charge deviennent pénalisantes à mesure que le niveau hiérarchique devient plus contraint.

La figure 4.5, relative au niveau  $k = 4$ , révèle une concentration marquée autour de **bus24** (0.8512,  $\uparrow$ ), suivie de **bus39** (0.5835,  $\uparrow$ ), **bus1** (0.3641,  $\uparrow$ ) et **bus31** (0.3464,  $\uparrow$ ). Contrairement au niveau précédent, la structure de contributions observée est ici majoritairement positif, indiquant que plusieurs variables (ex. **bus24**, **bus39**, **bus1**, **bus31**, **bus4**) tendent à favoriser

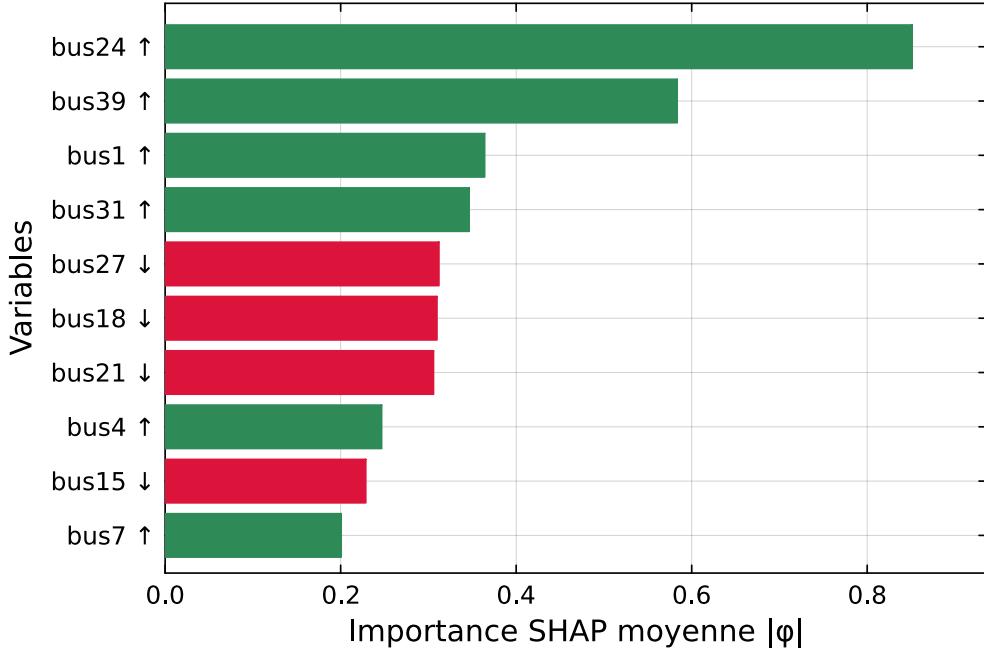


FIGURE 4.5 Niveau 4 — Valeurs SHAP par variable. Longueur des barres = importance moyenne  $|\phi|$ ; flèche ( $\uparrow$  en vert et  $\downarrow$  en rouge) = signe moyen de la contribution. Top 10 des variables classées par importance décroissante.

l'atteinte de  $z_4 = 1$ , tandis que d'autres (bus27, bus18, bus21, bus15) la restreignent. Nous notons une nouvelle inversion du signe de bus1, qui revient en contribution positive, ainsi qu'un renforcement du rôle de bus39, présent à tous les niveaux.

Une lecture croisée des trois figures révèle plusieurs propriétés structurelles du modèle. D'abord, certaines variables apparaissent de manière récurrente parmi les plus influentes, mais leur signe moyen varie selon le niveau, ce qui met en évidence la spécificité locale des décisions hiérarchiques. Ensuite, la concentration de l'importance explicative sur un nombre restreint de nœuds (ex. bus24 pour  $k = 4$ , bus27 pour  $k = 3$ ) facilite l'identification de leviers dominants dans l'activation des niveaux. Enfin, l'évolution du signe des contributions à travers la hiérarchie est cohérente avec un durcissement progressif des contraintes : à mesure que l'on progresse vers des niveaux plus élevés, certaines hausses de charge deviennent progressivement plus pénalisantes, voire bloquantes, tandis que d'autres conservent ou renforcent leur effet facilitateur.

Il convient toutefois de rappeler que les valeurs SHAP sont issues d'une agrégation d'effets locaux sur l'ensemble de l'échantillon : les flèches directionnelles indiquent des tendances moyennes, et certaines inversions de signe peuvent résulter de corrélations structurelles entre variables ou d'interactions non linéaires. Malgré ces limitations, les figures 4.3 à 4.5 offrent

un diagnostic robuste, interprétable et directement exploitable pour identifier les nœuds qui influencent le plus la décision d'activation de chaque niveau  $z_k$ .

Cette démarche d'analyse ne se limite pas à un diagnostic global du modèle : elle permet de caractériser, pour chaque niveau  $k$ , les variables qui favorisent ou inhibent l'activation de  $z_k$ , révélant ainsi des motifs décisionnels spécifiques à chaque profondeur hiérarchique. Ces effets différenciés, cohérents avec la structure ordonnée du problème, renforcent la validité interne du classifieur. Sur le plan pratique, ils constituent un support opérationnel pour l'explicabilité des prédictions et l'ajustement stratégique de la politique de décision. En identifiant les cas où la classification pourrait être rendue plus conservatrice ou plus permissive, l'analyse SHAP contribue à affiner les marges de sécurité sans compromettre la performance globale du système.

Par ailleurs, la nature hiérarchique de cet encodage autorise l'introduction d'un seuil de décision  $\tau$ , explicitement contrôlable, afin de calibrer le compromis entre conservatisme, surestimation du niveau requis, et risque de sous-prédition. Pour garantir la validité de la solution finale en aval de la prédiction, un test est systématiquement effectué pour déterminer l'activation du mécanisme de reprise est nécessaire.

Nous décrivons à présent l'intégration de ce processus dans une procédure complète de résolution adaptative du SCED réduit.

#### 4.4 Procédure complète de résolution guidée par la prédiction de niveau hiérarchique

La prédiction repose sur l'encodage binaire cumulatif défini à la section 4.3, dans lequel chaque composante  $z_k \in \{0, 1\}$  est estimée par un modèle de classification  $\mathfrak{C}_k$ , produisant une probabilité  $P_k = P(z_k = 1 | \mathbf{p}_d)$ . Le seuil de décision  $\tau \in (0, 1)$  intervient pour convertir ces probabilités en prédictions binaires, selon la règle  $\hat{z}_k = 1$  si  $P_k \geq \tau$ , 0 sinon. Le niveau hiérarchique final utilisé pour la résolution du SCED est ensuite défini par  $\hat{k}^* := \max\{ k \in \{1, \dots, k_{\max}\} \mid \hat{z}_k = 1 \}$ . Le choix de  $\tau$  contrôle directement le compromis entre conservatisme et efficacité algorithmique : un seuil faible favorise les sur-prédictions, garantissant l'inclusion des contraintes critiques mais réduisant les gains, tandis qu'un seuil élevé augmente le risque de sous-prédition, et donc de reprise.

Nous procédons ainsi à une calibration empirique de  $\tau$  sur un ensemble de validation, en analysant son impact sur l'exactitude initiale de la prédiction  $\hat{k}^*$ , le taux de reprises déclenchées, et les gains de temps observés. L'objectif est d'identifier une valeur optimale de  $\tau$  maximisant l'efficacité de la méthode tout en maintenant la robustesse de la solution.

La figure 4.6 illustre l'effet du seuil de décision  $\tau$  sur le taux d'exactitude initiale de la prédiction du niveau hiérarchique  $\hat{k}^*$ , mesuré avant l'activation du mécanisme de reprise.

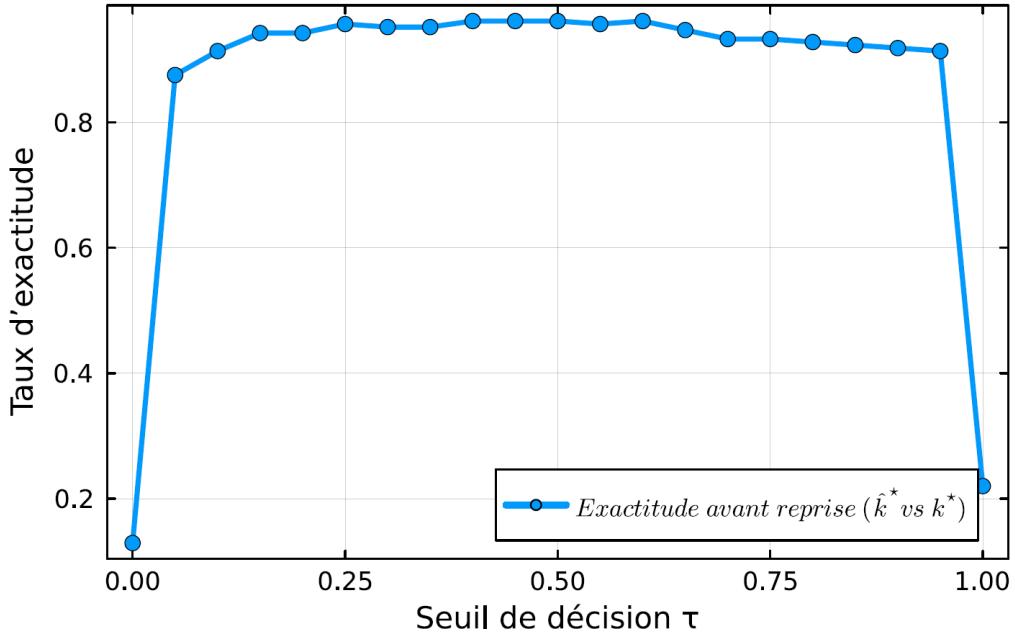


FIGURE 4.6 Taux d'exactitude de la prédiction initiale  $\hat{k}^*$  en fonction du seuil de décision  $\tau$ , avant activation du mécanisme de reprise.

Nous observons une zone plate de haute performance pour des seuils intermédiaires  $\tau \in [0.15, 0.85]$ , avec un maximum autour de 0.5, où l'exactitude dépasse 95 %. À l'extrême gauche, une valeur quasi-nulle de  $\tau$  entraîne une sur-prédiction systématique, et tous les classificateurs binaires  $\mathfrak{C}_k$  retournent  $\hat{z}_k = 1$ , si bien que  $\hat{k}^* = k_{\max}$  pour presque tous les profils. Cette stratégie, bien que conservatrice, engendre une perte d'exactitude notable, puisque seule une fraction minoritaire des profils correspond effectivement à un niveau maximal  $k^* = k_{\max}$ . Cela explique que le taux d'exactitude, bien que faible, reste strictement positif. Inversement, à l'extrême droite, un seuil très élevé, lorsque  $\tau$  tend vers 1, désactive la quasi-totalité des sorties binaires  $\hat{z}_k$ , menant à  $\hat{k}^* = 1$  dans tous les cas. Cette stratégie conduit à des sous-prédictions systématiques, sauf pour les profils dont la solution exacte est effectivement atteinte avec le niveau minimal  $k^* = 1$ , ce qui explique à nouveau la présence d'un taux d'exactitude non nul, mais très dégradé. Ce profil de forme concave valide la sensibilité de la prédiction à la calibration du seuil  $\tau$ , et suggère qu'une plage modérée assure un compromis entre conservatisme et exactitude avec une nécessité de reprises quasi-nulle.

La figure 4.7 met en évidence ces interprétations en détaillant la répartition des erreurs de classification avant l'application du mécanisme de reprise, en fonction du seuil de décision  $\tau$ .

Nous observons les deux types d'erreurs de manière distinguée avec d'une part, la sous-prédiction, lorsque  $\hat{k}^* < k^*$ , et d'autre part, la sur-prédiction, lorsque  $\hat{k}^* > k^*$ .

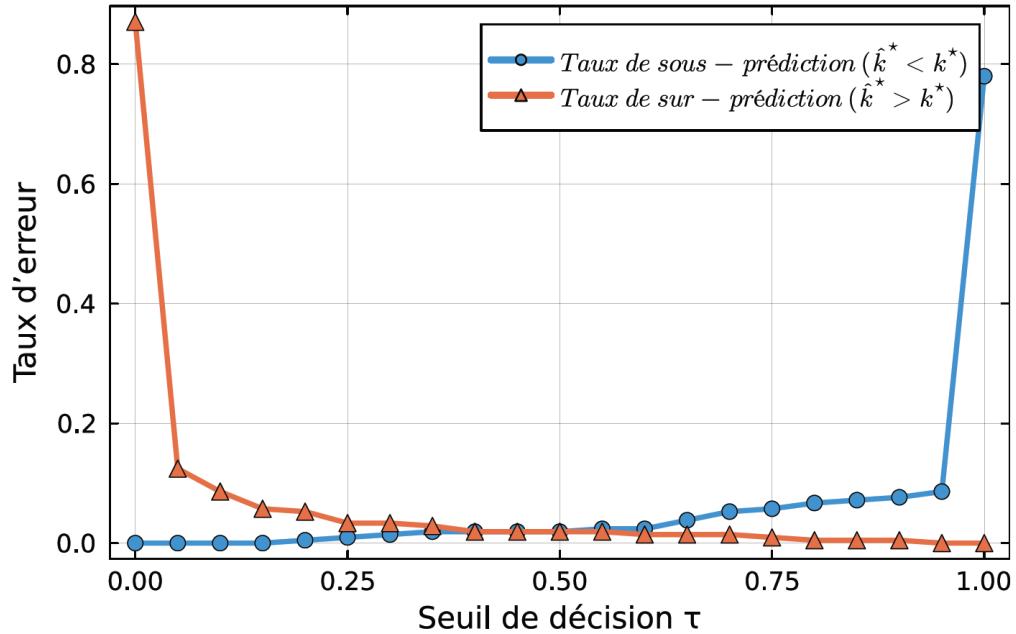
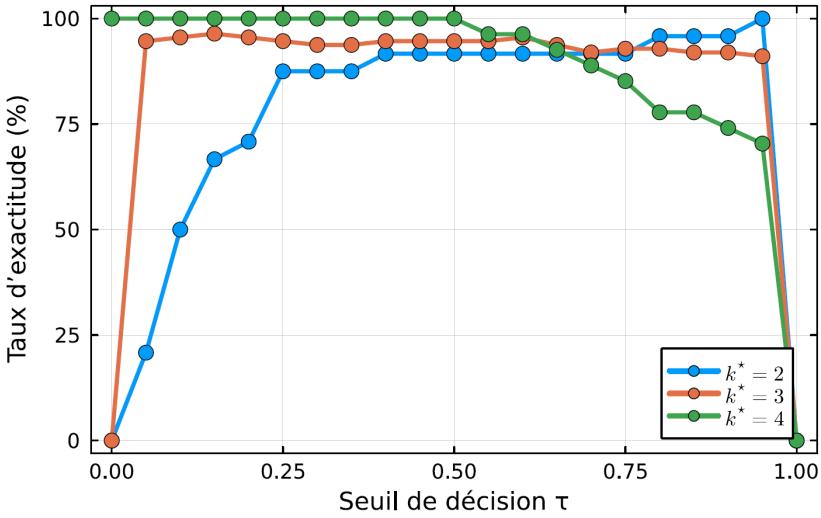
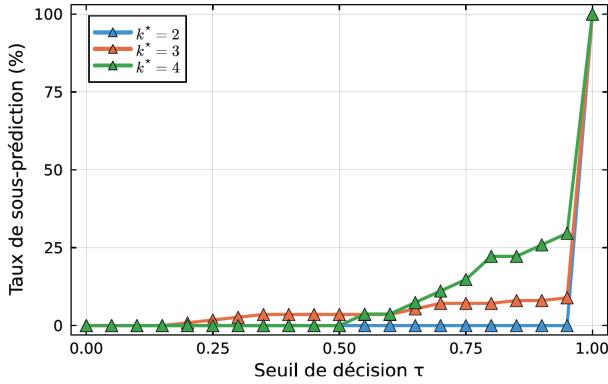
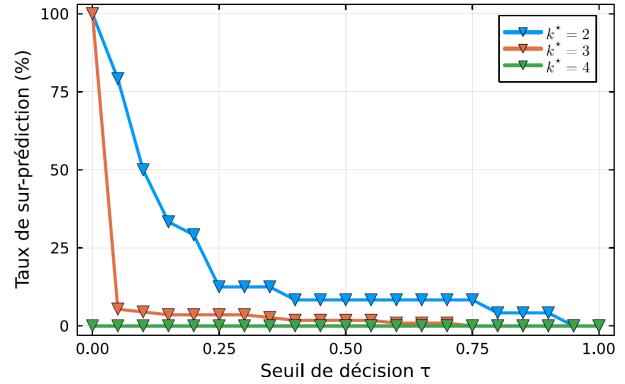


FIGURE 4.7 Impact du seuil  $\tau$  sur les erreurs de classification avant reprise.

Nous observons un comportement opposé strictement croissant et décroissant. Pour de faibles valeurs de seuil lorsque  $\tau$  tend vers 0, tous les classificateurs binaires  $\mathfrak{C}_k$  renvoient  $\hat{z}_k = 1$ , entraînant une prédiction maximale  $\hat{k}^* = k_{\max}$  dans la majorité des cas. Cela génère un taux de sur-prédiction élevé, atteignant plus de 80%, tandis que les sous-prédictions deviennent nulles. À l'inverse, lorsque  $\tau$  tend vers 1, aucun des  $\hat{z}_k$  n'est activé, ce qui induit  $\hat{k}^* = 1$ , causant une forte sous-prédiction, tandis que les sur-prédictions deviennent négligeables.

Entre ces deux extrêmes, la zone  $\tau \in [0.3, 0.6]$  constitue une région de compromis favorable, avec des taux d'erreurs globalement faibles et stables.

Les figures 4.8a à 4.8c détaillent l'effet du seuil de décision  $\tau \in [0, 1]$  sur la précision de prédiction par niveau hiérarchique réel  $k^* \in \{2, 3, 4\}$ . Contrairement à une analyse agrégée, cette étude conditionnelle permet d'observer la sensibilité propre de chaque classificateur binaire  $\mathfrak{C}_k$  à la variation du seuil, ainsi que la distribution des erreurs selon le niveau cible.

(a) Exactitude de prédiction avant reprise par niveau  $k^*$  selon  $\tau$ .(b) Sensibilité aux sous-prédition par niveau  $k^*$  selon  $\tau$ .(c) Sensibilité aux sur-prédition par niveau  $k^*$  selon  $\tau$ .FIGURE 4.8 Analyse des performances de prédiction par le niveau  $k^*$ , en fonction du seuil de décision  $\tau$ .

Pour le niveau  $k^* = 2$ , la figure 4.8a révèle une exactitude très faible pour les seuils bas ( $\tau \leq 0.1$ ), liée à une sur-prédition systématique illustrée dans la figure 4.8c, proche de 100 %. Cependant, la figure 4.8b montre que aucune sous-prédition n'est observée pour ce niveau tant que  $\tau < 1$ . Cela signifie que la prédiction initiale respecte toujours  $\hat{k}^* \geq 2$ , même lorsque le classifieur est trop permissif. L'exactitude croît ensuite rapidement au-delà de  $\tau \approx 0.2$ , pour atteindre un plateau supérieur à 90 % entre  $\tau = 0.3$  et 0.9, avant de décroître à  $\tau = 1$ , où toutes les prédictions sont rejetées. À l'inverse, le niveau  $k^* = 4$  présente des résultats marqués par une absence totale de sur-prédictions pour l'ensemble des valeurs du seuil  $\tau$ , comme le montre la figure 4.8c. Cette absence ne reflète pas une performance optimale du modèle, mais provient

du fait qu'aucun profil de charge d'entraînement n'a été étiqueté avec un niveau strictement supérieur à 4. En l'absence de telles observations, le modèle ne peut apprendre à distinguer les situations justifiant une prédiction  $\hat{k}^* > 4$ , ce qui bloque par construction toute possibilité de sur-prédiction pour  $k^* = 4$ . En revanche, la sous-prédiction devient dominante dès que  $\tau$  dépasse 0,65, atteignant 100 % à  $\tau = 1$ , comme l'illustre la figure 4.8b. Ce phénomène souligne le rôle critique de la distribution des niveaux dans les données d'apprentissage. En effet, une couverture insuffisante des cas rares de haut niveau peut induire un biais structurel, où des contraintes potentiellement actives sont systématiquement omises dans les prédictions. Ceci justifie la mise en place du mécanisme de reprise. Le niveau  $k^* = 3$ , qui concerne plus de la moitié des profils comme nous l'avons montré dans le tableau 4.3), présente une position intermédiaire. Il montre une bonne robustesse dans laquelle l'exactitude reste supérieure à 93 % pour un large intervalle de seuils  $\tau \in [0.15, 0.85]$ , et les erreurs de sous et sur-prédictions y sont modérées. Cette stabilité traduit une bonne calibration du modèle  $\mathfrak{C}_3$  sur la majorité des cas.

L'interprétation de ces figures met en évidence l'absence d'un seuil unique optimal pour tous les niveaux. Les seuils bas minimisent les sous-prédictions pour  $k^* = 4$ , mais induisent des sur-prédictions critiques pour  $k^* = 2$ . À l'inverse, des seuils plus élevés corrige efficacement ce dernier cas, mais au prix d'un rejet systématique des niveaux supérieurs. Ces tensions confirment que le choix du seuil  $\tau$  est un compromis entre efficacité et conservatisme.

En pratique, un seuil global dans l'intervalle [0.3, 0.6] semble constituer un bon compromis, assurant une exactitude élevée tout en limitant les erreurs critiques. Néanmoins, ces résultats suggèrent une possibilité d'amélioration future, par la mise en œuvre d'une stratégie de calibration différenciée, où chaque niveau hiérarchique  $k$  disposerait de son propre seuil  $\tau_k$ , permettant d'adapter finement la sensibilité de chaque classifieur à la distribution conditionnelle de ses échantillons.

Finalement, nous évaluons, sur une instance de 4265 échantillons, la performance de la procédure complète de résolution dans laquelle le niveau hiérarchique  $\hat{k}^*$  est prédit à partir du profil de charge nodale  $\mathbf{p}_d \in \mathbb{R}^{|\mathcal{N}|}$ , puis utilisé pour résoudre un SCED restreint aux contraintes hiérarchiques  $\mathcal{I}_{1:\hat{k}^*}$ . La figure 4.9 illustre la proportion de profils de charge pour lesquels la prédiction initiale du niveau hiérarchique  $\hat{k}^*$  a nécessité une ou plusieurs reprises adaptatives, selon la valeur du seuil  $\tau$  utilisé pour la classification. Pour chaque seuil, nous comptabilisons la fréquence relative d'activation du mécanisme de reprise ( $\hat{k}^* \mapsto \hat{k}^* + 1 \mapsto \dots$ ), lequel vise à garantir l'exactitude de la solution en injectant progressivement les contraintes hiérarchiques absentes. Cette reprise est déclenchée dès qu'un dépassement de contrainte est détectée à

l'issue de la résolution restreinte, selon la condition de validation a posteriori définie en (3.8) :

$$\Phi_i \mathbf{p}^* < \bar{\mathbf{s}}_i,$$

où  $\mathbf{p}^*$  est une solution du SCED restreint, et  $\Phi_i$  désigne la  $i$ -ème ligne de la matrice de contraintes d'écoulement de puissance non considérée dans le niveau hiérarchique prédit.

Nous observons une courbe globalement plate pour  $\tau \in [0.05, 0.8]$ , avec un taux de reprise modéré compris entre 10 % et 15 %, traduisant un bon compromis entre précision et conservatisme dans la prédition initiale. Ce plateau témoigne d'une certaine robustesse du classifieur, et témoigne que la majorité des prédictions sont suffisamment proches du niveau réel pour éviter la nécessité de corrections coûteuses.

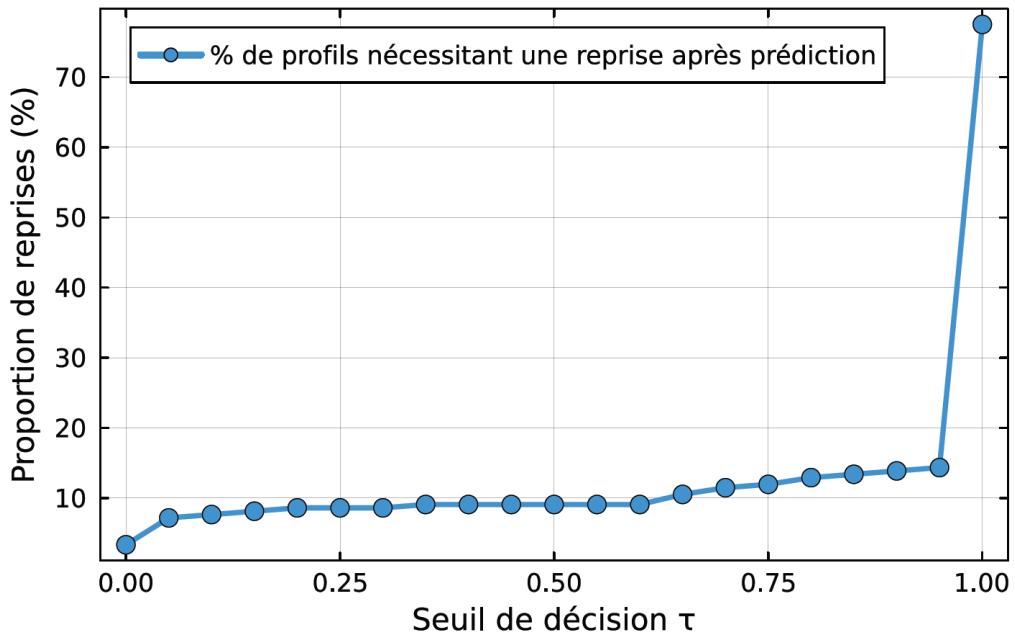


FIGURE 4.9 Proportion de profils nécessitant au moins une reprise adaptative, en fonction du seuil de décision  $\tau$ .

À l'opposé, les extrémités du domaine sont marquées par des comportements extrêmes. Lorsque  $\tau$  est proche de 0, le modèle tend à considérer presque tous les niveaux, ce qui explique un taux de reprise initialement bas ( $< 5\%$ ). À l'autre extrême, lorsque  $\tau \rightarrow 1$ , la prédition devient excessivement conservatrice : aucun niveau n'est activé, entraînant un taux de reprise élevé dépassant 70 %. Ces résultats confirment l'intérêt de calibrer le seuil  $\tau$  dans un intervalle intermédiaire, e.g., [0.3, 0.6], où la précision de prédition est élevée et la nécessité de reprises faible.

Afin d'évaluer l'impact global du seuil de décision  $\tau$  sur la performance opérationnelle de la

méthode proposée, nous analysons le compromis entre la précision de la prédiction initiale du niveau hiérarchique  $\hat{k}^*$  et le gain de temps moyen engendré par la résolution restreinte du SCED. La figure 4.10 illustre cette relation.

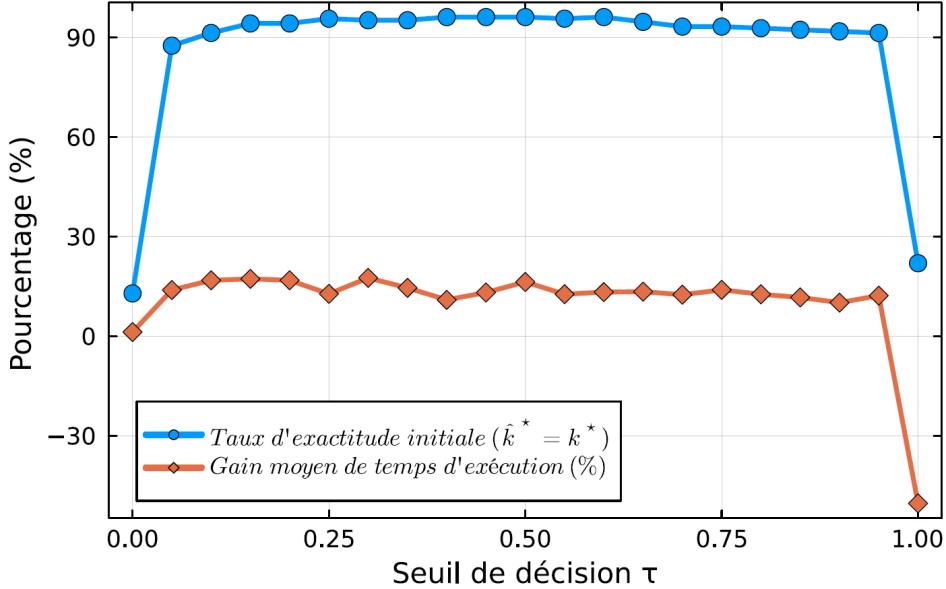


FIGURE 4.10 Compromis entre l'exactitude de la prédiction initiale ( $\hat{k}^* = k^*$ ) et le gain de temps moyen obtenu par rapport à la résolution complète.

Dans l'intervalle  $\tau \in [0.05, 0.6]$ , le système atteint un compromis particulièrement favorable, combinant une exactitude initiale élevée (supérieure à 95 % pour  $\tau = 0.4$  à 0.6) et des gains de temps moyens significatifs, allant jusqu'à plus de 17.5 % pour  $\tau = 0.3$ . Cette zone correspond à une prédiction suffisamment confiante tout en évitant les reprises coûteuses. Lorsque  $\tau$  est trop faible ( $\tau \leq 0.05$ ), la prédiction initiale implique que toutes les contraintes sont injectées d'emblée dans la formulation restreinte, ce qui rapproche le problème du cas complet, réduisant ainsi le gain de temps. Ce n'est donc pas la correction *a posteriori* qui est coûteuse, mais bien l'absence de réduction effective lors de la résolution initiale. À l'inverse, lorsque  $\tau$  tend vers 1, la sévérité du seuil bloque quasi systématiquement l'activation des classificateurs binaires, menant à des prédictions erronées importantes avec une exactitude approximative de 22 %) et à des reprises systématiques. Le temps total de calcul excède alors celui du modèle complet, ce qui se traduit par une chute drastique du gain de temps moyen de -50 %.

Ces résultats confirment que la performance de la méthode repose fortement sur une calibration adéquate du seuil  $\tau$ . Une valeur trop conservatrice ou trop permissive augmente le recours au mécanisme de reprise, dont le déclenchement est coûteux et guidé par la condition de validation *a posteriori* (3.8). Une extension naturelle consisterait à considérer une fonction de

seuil adaptatif, optimisée en fonction du niveau hiérarchique ou des caractéristiques du profil de charge.

L'évaluation finale a été menée sur un échantillon de 4265 profils de charge, en appliquant la procédure complète de prédiction et résolution adaptative décrite précédemment. Le seuil de décision  $\tau$  a été fixé à sa valeur optimale  $\tau = 0.5$ , identifiée lors de la phase de calibration sur les données de validation.

TABLEAU 4.4 Bilan agrégé de la procédure prédictive à seuil  $\tau = 0.5$

Indicateur	Valeur	Unité
Nombre total de profils analysés	4265	—
Nombre cumulé de reprises	852	—
Durée moyenne résolution complète	0.121	s
Durée moyenne résolution réduite	0.096	s
Gain moyen de temps absolu	0.024	s
Gain moyen de temps relatif	19.72	%

Le tableau 4.4 synthétise les performances observées. Le nombre total de reprises, comptabilisé de manière cumulative sur tous les profils, s'élève à 852. Cela inclut les cas où plusieurs incrémentations du niveau ont été nécessaires pour satisfaire la condition (3.8). Ce résultat témoigne de la capacité des classificateurs hiérarchiques  $\mathfrak{C}_k$ , dont les décisions binaires cumulées convergent dans la majorité des cas vers un niveau admissible.

En moyenne, la procédure adaptative permet un gain de temps relatif de 19.72 % par rapport à la résolution complète. Ce gain reste d'autant plus significatif que les économies de temps sont principalement réalisées sur les profils associés à des niveaux hiérarchiques faibles. En effet, plus le niveau hiérarchique requis est élevé, plus le nombre de contraintes injectées dans la résolution est important, rendant le temps de calcul de la formulation réduite proche de celui du problème complet. Ainsi, les profils nécessitant un niveau hiérarchique élevé apportent un gain marginal, tandis que les profils associés à des niveaux hiérarchiques faibles bénéficient pleinement de la réduction. Ces résultats confirment la pertinence de l'approche proposée dans un cadre opérationnel, en particulier dans des conditions normales de fonctionnement du réseau, où les profils de charge observés s'éloignent rarement des cas extrêmes. En pratique, les niveaux hiérarchiques les plus élevés, associés aux contraintes les plus marginales, ne sont sollicités que dans des situations extrêmes (épisodes climatiques sévères, défaillances multiples, pics de consommation rares). La capacité à s'en tenir, dans la majorité des cas, à un sous-ensemble réduit de contraintes rend la méthode particulièrement adaptée aux

environnements temps réel, où la rapidité de calcul et la robustesse décisionnelle sont critiques.

#### 4.5 Conclusion des résultats

L'ensemble des résultats présentés dans cette section atteste de la robustesse et de la performance algorithmique de la méthodologie proposée. Sur les trois réseaux de test considérés, la stratégie hiérarchique guidée par apprentissage supervisé permet, sans compromis sur l'exactitude de la solution, une réduction substantielle du nombre de contraintes explicitement mobilisées, ainsi qu'un gain significatif en temps de calcul. Cette compression structurelle s'avère d'autant plus marquée que le système est de grande dimension, témoignant de la bonne mise à l'échelle de l'approche.

L'analyse détaillée du modèle prédictif montre que les erreurs de sous-prédition sont rares, circonscrites à des niveaux hiérarchiques élevés et systématiquement corrigées par le mécanisme de reprise incrémentale. La précision moyenne atteint plus de 94 %, avec une erreur absolue moyenne inférieure à 0.1 niveau, ce qui confirme l'aptitude du modèle à capturer les classes associées aux contraintes de limite d'écoulement de puissance. Les courbes de convergence du classifieur LightGBM, ainsi que son comportement en fonction du seuil de décision  $\tau$ , révèlent une bonne régularisation, capable de concilier précision et prudence, conformément aux exigences de sûreté des réseaux électriques.

Enfin, les analyses par niveau hiérarchique montrent que les performances du modèle sont particulièrement élevées dans les niveaux bas et intermédiaires, qui concentrent la majorité des cas rencontrés en pratique. Cette observation souligne la cohérence du cadre hiérarchique car, dans la grande majorité des cas, le problème optimal peut être résolu en mobilisant uniquement un ensemble réduit de contraintes. Ce résultat souligne l'intuition physique selon laquelle les configurations extrêmes, nécessitant l'activation de niveaux hiérarchiques supérieurs, sont peu fréquentes dans l'exploitation normale du réseau.

L'ensemble de ces éléments suggère que la méthode proposée constitue une voie intéressante pour la réduction adaptative de contrainte dans les formulations pénalisées du SCED, compatible avec une intégration réaliste dans les contextes décisionnels temps réel. Les perspectives ouvertes par cette approche seront discutées en conclusion générale du mémoire.

## CHAPITRE 5 CONCLUSION

Ce chapitre conclut les travaux menés dans le cadre de ce mémoire. Il en synthétise les principales contributions méthodologiques et expérimentales, discute les limites identifiées de l'approche proposée, et propose plusieurs améliorations pour des recherches futures.

### 5.1 Synthèse des travaux

Les travaux présentés dans ce mémoire s'inscrivent dans le cadre de l'optimisation du SCED pour les réseaux électriques de grande dimension, en réponse aux défis computationnels croissants rencontrés par les opérateurs de marché face à la densification des contraintes et à la multiplication des scénarios de contingence. Ce problème, au cœur des architectures de marché modernes, s'inscrit à la fois dans un contexte de forte exigence de précision opérationnelle et d'un besoin impératif de réduction de complexité algorithmique. Dans ce contexte, l'objectif de ce mémoire est de concevoir une méthode de réduction de contraintes efficace, compatible avec les formulations pénalisées du SCED utilisées en pratique, i.e., dans sa forme avec des contraintes souples, et capable de s'adapter à la structure du problème pour chaque profil de charge observé.

Les travaux présentés dans ce mémoire s'articulent autour de deux contributions scientifiques complémentaires, actuellement en cours de rédaction sous forme d'articles. La première contribution prend la forme d'une revue de l'état de l'art, consacrée aux approches de réduction de contraintes et d'optimisation structurelle dans les formulations du SCED à contraintes strictes. La seconde contribution, qui constitue l'apport central de ce mémoire, propose une méthode de réduction de contraintes guidée par apprentissage supervisé, articulée autour de deux innovations conceptuelles :

1. **La hiérarchisation des contraintes dans le SCED pénalisé.** Nous introduisons une généralisation du concept classique de contrainte redondante au cadre des formulations à contraintes souples, dans lesquelles la faisabilité est toujours assurée par des variables de pénalités. Cette extension repose sur une proposition d'équivalence formelle entre les formulations complète et réduite. Elle permet la construction d'une hiérarchie d'ensembles de contraintes non redondantes, extraits récursivement à l'aide de l'algorithme `HIERARCHICALCONSTRAINTEXTRACTION` (Algorithme 3). Chaque niveau de cette hiérarchie définit une région admissible  $\mathcal{F}_k$  strictement incluse dans la suivante, assurant par construction une couverture progressive de l'espace réalisable.
2. **La prédiction du niveau hiérarchique minimal dans un cadre de classification**

**hiérarchique cumulative.** En exploitant la structure imbriquée des régions  $\mathcal{F}_k$ , nous formulons la tâche d’identification du plus petit niveau  $k^*$  suffisant pour garantir l’exactitude de la solution comme un problème de classification hiérarchique cumulative. Chaque niveau  $k \in \{1, \dots, k_{\max}\}$  est associé à un classifieur binaire  $\mathfrak{C}_k$ , entraîné à estimer la probabilité que les contraintes de niveaux 1 à  $k$  engendrent une région admissible contenant l’optimum du SCED complet. La prédiction finale est reconstruite par agrégation des sorties, à l’aide d’un mécanisme de seuillage contrôlant le compromis entre conservatisme et précision. Une reprise incrémentale est intégrée pour garantir la robustesse vis-à-vis des erreurs de sous-prédiction.

Les résultats expérimentaux démontrent un potentiel d’efficacité de la stratégie proposée. En particulier, la formulation restreinte à un sous-ensemble hiérarchique minimal  $\mathcal{I}_{1:k^*}$ , sans recours à l’apprentissage, permet déjà de reproduire à l’identique la solution optimale du SCED complet, tout en réduisant significativement le temps de résolution et le nombre de contraintes considérées. Pour les réseaux `case9`, `case39` et `case118`, la taille du système d’optimisation est divisée respectivement par 63 %, 18.9 % et 6.1 %, pour un gain de temps allant jusqu’à 41 %, sans aucune dégradation de la solution économique. Cette première validation structurelle atteste de la pertinence du découpage hiérarchique introduit.

Sur cette base, l’intégration d’un modèle de prédiction LightGBM permet d’automatiser, pour chaque profil nodal  $\mathbf{p}_d$ , l’estimation du niveau minimal  $k^*$  requis. L’évaluation globale du modèle pour le cas `case39` montre une précision de 94.3 %, avec un taux de sous-prédictions inférieur à 3.5 %, et une faible erreur absolue moyenne. Ce cadre prédictif, couplé à un mécanisme de reprise adaptative, permet d’atteindre un gain moyen de temps de 19.7 % sur plus de 4000 profils de test, sans perte de qualité de solution. Ces résultats confirment une bonne capacité de la méthode à maintenir un compromis favorable entre exactitude, robustesse et performance algorithmique.

Ces résultats prometteurs supportent la validité du cadre méthodologique proposé. Néanmoins, certaines limitations que nous abordons dans la section suivante.

## 5.2 Limitations de la solution proposée

Malgré les avancées conceptuelles introduites dans ce travail, plusieurs limitations subsistent, tant sur le plan computationnel que sur celui de la structure algorithmique.

Une première limitation est que la méthode proposée n’a pu, à ce stade, être testée sur des réseaux de très grande taille, i.e, plusieurs milliers de nœuds ou lignes, du fait de contraintes matérielles sur les ressources de calcul disponibles. La phase de génération supervisée, nécessi-

tant de résoudre plusieurs milliers d'instances du SCED complet, induit une consommation mémoire et temporelle difficilement maîtrisable sans infrastructure dédiée. En conséquence, les résultats présentés sont limités à des réseaux de taille intermédiaire, ce qui limite la généralisation empirique des observations. Il convient de noter que, dans un contexte opérationnel réel, les profils de charge à partir desquels la classification est effectuée peuvent être directement extraits de données historiques mises à disposition par les opérateurs de marché, tels que les *State Estimator Load Reports* publiés par ERCOT [78]. Cette disponibilité réduit considérablement le coût d'intégration de la méthode proposée dans un cadre industriel, en évitant la phase de génération synthétique des profils, qui est précisément celle qui constitue le principal facteur de coût computationnel de la méthodologie, dans la mesure où elle requiert l'exécution du SCED complet un très grand nombre de fois pour produire un ensemble d'entraînement suffisamment diversifié.

D'autre part, la méthode repose sur une représentation dense et explicite des matrices  $\Phi$ , stockées en mémoire pour chaque scénario de contingence. Or, ces matrices sont massives, d'autant plus que le nombre de lignes et de contingences augmente. Leur gestion constitue, dans la version actuelle de notre méthode, un goulot d'étranglement majeur, à la fois en mémoire vive et en bande passante, notamment lors des étapes de génération de données, où chaque instance du SCED requiert une reconstruction complète du modèle.

Au-delà de leur coût de stockage, les matrices  $\Phi^{(c)}$ , représentant les sensibilités d'écoulement pour chaque scénario de contingence  $c \in \mathcal{C}$ , sont supposées fixes sur l'horizon d'analyse. Cette hypothèse, implicite dans la construction hiérarchique, revient à considérer une topologie constante du réseau. Or, en pratique, des évolutions planifiées ou aléatoires de la structure peuvent altérer la validité des régions admissibles  $\mathcal{F}_k$  construites à partir d'un état nominal. L'approche reste néanmoins pertinente dans un cadre à horizon glissant ou pour des scénarios stationnaires, mais sa généralisation à des contextes dynamiques exigerait un raffinement adaptatif de la hiérarchie.

La hiérarchie  $\{\mathcal{I}_k\}_{k=1}^{k_{\max}}$ , bien qu'indépendante des profils de charge et construite une seule fois en amont, s'appuie sur une procédure récursive de suppression de redondance structurelle. Cette étape, fondée sur des appels successifs au solveur, mobilise de manière intensive les ressources de calcul et reste peu parallélisable dans sa version actuelle. Pour des réseaux de grande taille, elle peut représenter un coût de prétraitement non négligeable. Toutefois, ce coût est amorti dès lors que la hiérarchie est réutilisée sur plusieurs instances, et s'inscrit dans une logique d'optimisation hors ligne compatible avec les contraintes opérationnelles.

Ces considérations ouvrent la voie à des améliorations qui font l'objet de la section suivante.

### 5.3 Améliorations futures

La méthode proposée constitue une première démonstration de faisabilité pour la réduction hiérarchique de contraintes guidée par apprentissage supervisé, appliquée aux formulations pénalisées du SCED. L'ensemble des résultats expérimentaux présentés dans ce mémoire repose sur des réseaux de taille intermédiaire, sélectionnés de manière à garantir un compromis entre représentativité structurelle et faisabilité computationnelle. En particulier, la génération exhaustive des échantillons d'entraînement requis pour calibrer les classificateurs hiérarchiques n'a pu être réalisée, par contrainte de temps, sur des réseaux de grande dimension. Dans cette perspective, la généralisation de la méthode à des réseaux de plus grande échelle constitue une première extension naturelle. Elle permettrait de consolider empiriquement les performances de la démarche et d'en évaluer la scalabilité. Au-delà de cette première piste, plusieurs directions concrètes d'amélioration peuvent être envisagées afin d'étendre la portée du cadre proposé et d'en renforcer la robustesse algorithmique.

Une première généralisation naturelle consiste à intégrer des contraintes inter-temporelles, telles que les rampes de production, dans une version multipériode du SCED. Dans ce cadre, chaque période  $t \in \{1, \dots, T\}$  est associée à sa propre matrice  $\Phi^{(t)}$ , reflétant la configuration topologique du réseau à cet instant. Toutefois, dans la pratique, seules certaines contingences modifient effectivement la topologie sous-jacente. Il devient donc pertinent de restreindre la mise à jour de la hiérarchie de contraintes aux seules régions du graphe affectées par ces changements, en s'appuyant par exemple sur des parcours en largeur (*breadth-first search*, BFS) localisés autour des lignes modifiées. Une telle stratégie permettrait d'éviter le recalcul intégral de la hiérarchie et de limiter la charge mémoire en ne stockant que les blocs utiles des matrices  $\Phi^{(t)}$ .

Deuxièmement, la structure cumulative du problème suggère une modélisation par classification ordinale plutôt que par une famille de classificateurs binaires entraînés de manière indépendante. Une direction prometteuse consiste à adapter des méthodes robustes de classification ordinale, telles que celles fondées sur les *robust support vector ordinal regression*, récemment proposées dans [84]. Ces approches permettent de capturer explicitement la structure d'ordre entre les niveaux hiérarchiques tout en intégrant des mécanismes de régularisation et de gestion du bruit sur les étiquettes, particulièrement utiles dans les cas où les niveaux  $k^*$  sont sensibles à de faibles perturbations du profil de charge.

Enfin, à mesure que la taille des réseaux considérés augmente, le besoin de distribuer les calculs devient critique. Une extension naturelle du cadre actuel consisterait à répartir la génération de profils, l'entraînement des classificateurs  $\mathfrak{C}_k$ , et l'évaluation du SCED sur une

infrastructure parallèle ou distribuée. Cette direction ouvrirait la voie à une scalabilité réelle de l'approche, tout en permettant de traiter des réseaux interconnectés à l'échelle continentale.

Ces axes d'amélioration définissent des prolongements méthodologiques clairs, qui pourront faire l'objet de développements futurs tant du point de vue théorique qu'algorithme.

## RÉFÉRENCES

- [1] Federal Energy Regulatory Commission, “Regional transmission organizations/independent system operators,” 2024, retrieved from <https://www.ferc.gov/power-sales-and-markets/rtos-and-isos>.
- [2] New York ISO and Independent Electricity System Operator, “Part 4.3 : Operation of the real-time markets,” NYISO and IESO, Technical report, 2024, market and System Operations – Real-Time Energy and Operating Reserve. Retrieved from <https://www.ieso.ca/-/media/Files/IESO/Document-Library/engage/imrm/mm-4-3-operation-of-the-real-time-markets-20240313.pdf>.
- [3] S. Fiorenzani, *Quantitative Methods for Electricity Trading and Risk Management*. Palgrave Macmillan, 2018.
- [4] F. E. R. Commission, “Order no. 888 : Promoting wholesale competition through open access non-discriminatory transmission services by public utilities and recovery of stranded costs by public utilities and transmitting utilities,” United States Federal Energy Regulatory Commission, Rapport technique Order No. 888, 75 FERC 61080, apr 1996, final rule issued April 24, 1996. <https://www.ferc.gov/industries-data/electric/industry-activities/open-access-transmission-tariff-oatt-reform/history-oatt-reform/order-no-888>
- [5] M. C. Ferris, “Overview of electricity markets,” 2024, presentation, Wisconsin Public Utilities Institute, April 9. [https://www.ferc.gov/sites/default/files/2020-06/energy-primer-2020\\_Final.pdf](https://www.ferc.gov/sites/default/files/2020-06/energy-primer-2020_Final.pdf)
- [6] X. Ma *et al.*, “The security-constrained commitment and dispatch for midwest iso day-ahead co-optimized energy and ancillary service market,” dans *2009 IEEE Power & Energy Society General Meeting*, 2009, p. 1–8.
- [7] Midcontinent Independent System Operator, *Energy and operating reserve markets business practices manual (BPM-002-r24)*, MISO, 2023, effective date : September 30, 2023.
- [8] I. M. Mack, *Energy trading and risk management : A practical approach to hedging, trading, and portfolio diversification*. Wiley, 2014.
- [9] J. A. Taylor, *Convex optimization of power systems*. Cambridge University Press, 2015.
- [10] C. Kroer, “Lecture note 13 : Power flows and equilibrium pricing,” 2022, accessed in 2024. <https://www.columbia.edu/~ck2945/teaching/e4903-spring2022/Lecture13.pdf>
- [11] V. Hinojosa-Mateus, L. Pérez-Andrades et M. Ilić, “DC optimal power flow formulation using the power transmission distribution factors—a DIgSILENT programming language

- application,” dans *PowerFactory applications for power system analysis*. Springer, 2014, p. 111–134.
- [12] R. Weinhold et R. Mieth, “Fast security-constrained optimal power flow through low-impact and redundancy screening,” *IEEE Transactions on Power Systems*, vol. 35, n°. 6, p. 4574–4584, 2020.
  - [13] J. Guo *et al.*, “Direct calculation of line outage distribution factors,” *IEEE Transactions on Power Systems*, vol. 24, n°. 3, p. 1633–1634, aug 2009.
  - [14] Midcontinent Independent System Operator, “Schedule 28 – demand curves for operating reserves,” 2023, accessed : 2025-07-10. <https://www.misoenergy.org/legal/tariff/>
  - [15] P. L. Joskow et J. Tirole, “Transmission rights and market power on electric power networks,” *The RAND Journal of Economics*, vol. 31, n°. 3, p. 450–487, 2000.
  - [16] Z. Yan et Y. Xu, “A hybrid data-driven method for fast solution of security-constrained optimal power flow,” *IEEE Transactions on Power Systems*, vol. 37, n°. 6, p. 4365–4374, 2022.
  - [17] C. Wang et Y. Fu, “Fully parallel stochastic security-constrained unit commitment,” *IEEE Transactions on Power Systems*, vol. 31, n°. 5, p. 3561–3571, 2016.
  - [18] J. Hörsch *et al.*, “Linear optimal power flow using cycle flows,” *Electric Power Systems Research*, vol. 158, p. 126–135, 2018.
  - [19] B. N. Giraud, A. Rajaei et J. L. Cremer, “Constraint-driven deep learning for n-k security constrained optimal power flow,” *Electric Power Systems Research*, vol. 235, p. 110692, 2024.
  - [20] F. Hasan et A. Kargarian, “Topology-aware learning assisted branch and ramp constraints screening for dynamic economic dispatch,” *IEEE Transactions on Power Systems*, vol. 37, n°. 5, p. 3495–3505, 2022.
  - [21] J. Nocedal et S. J. Wright, *Numerical optimization*, 2<sup>e</sup> éd. Springer, 2006.
  - [22] A. Murli et G. Di Pillo, édit., *High performance algorithms and software for nonlinear optimization*. Kluwer Academic Publishers, 2013.
  - [23] M. Javadi *et al.*, “Identification of simultaneously congested transmission lines in power systems operation and market analysis,” *IEEE Transactions on Power Systems*, vol. 32, n°. 3, p. 1772–1781, 2017.
  - [24] A. J. Conejo *et al.*, *Decomposition techniques in mathematical programming : Engineering and science applications*. Springer, 2006.
  - [25] J. Mohammadi, G. Hug et S. Kar, “A benders decomposition approach to corrective security constrained opf with power flow control devices,” dans *2013 IEEE Power & Energy Society General Meeting*, 2013, p. 1–5.

- [26] Y. Li et J. D. McCalley, "Decomposed SCOPF for improving efficiency," *IEEE Transactions on Power Systems*, vol. 24, n°. 1, p. 494–495, 2009.
- [27] M. Bazrafshan, K. Baker et J. Mohammadi, "Computationally efficient solutions for large-scale security-constrained optimal power flow," arXiv preprint arXiv :2006.00585, 2020. <https://arxiv.org/abs/2006.00585>
- [28] S. Boyd *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, n°. 1, p. 1–122, 2011.
- [29] D. Phan et J. Kalagnanam, "Some efficient optimization methods for solving the security-constrained optimal power flow problem," *IEEE Transactions on Power Systems*, vol. 29, n°. 2, p. 863–872, 2014.
- [30] L. Liu *et al.*, "A distributed parallel approach for big data scale optimal power flow with security constraints," dans *2013 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. IEEE, 2013, p. 774–778.
- [31] C. G. Petra et I. Aravena, "Solving realistic security-constrained optimal power flow problems," arXiv preprint arXiv :2110.01669, 2021. <http://arxiv.org/abs/2110.01669>
- [32] C. Zhao *et al.*, "A marginal equivalent decomposition method and its application to security-constrained economic dispatch," *IEEE Transactions on Power Systems*, vol. 29, n°. 6, p. 2718–2727, 2014.
- [33] A. Velloso, P. Van Hentenryck et E. S. Johnson, "An exact and scalable problem decomposition for security-constrained optimal power flow," *Electric Power Systems Research*, vol. 195, p. 106677, 2021.
- [34] Y. Varawala *et al.*, "A generalised approach for efficient computation of look-ahead security-constrained optimal power flow," *Electric Power Systems Research*, vol. 220, p. 109347, 2023.
- [35] M. Vistnes *et al.*, "A novel approach for solving security-constrained optimal power flow using the inverse matrix modification lemma and benders decomposition," arXiv preprint arXiv :2305.01395, 2023. <https://arxiv.org/abs/2305.01395>
- [36] O. Alsac, B. Stott et W. F. Tinney, "Sparsity-oriented compensation methods for modified network solutions," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-102, n°. 5, p. 1050–1060, 1983.
- [37] A. B. Birchfield *et al.*, "Grid structural characteristics as validation criteria for synthetic networks," *IEEE Transactions on Power Systems*, vol. 32, n°. 4, p. 3258–3265, jul 2017.
- [38] T. Erseghe, "Distributed optimal power flow using ADMM," *IEEE Transactions on Power Systems*, vol. 29, n°. 5, p. 2370–2380, 2014.

- [39] R. Verschae *et al.*, “Distributed SCOPF using convex relaxations and the ADMM,” *IEEE Transactions on Smart Grid*, vol. 10, n°. 2, p. 1880–1890, 2019.
- [40] E. Dall’Anese, G. B. Giannakis et B. F. Wollenberg, “Distributed optimal power flow for smart microgrids,” *IEEE Transactions on Smart Grid*, vol. 4, n°. 3, p. 1464–1475, 2013.
- [41] A. Maneesha *et al.*, “A survey on applications of alternating direction method of multipliers in smart power grids,” *Renewable and Sustainable Energy Reviews*, vol. 152, p. 111687, 2021.
- [42] Z. Chen *et al.*, “A distributed power flow optimization method based on graph partition and alternating direction method of multipliers,” *Energies*, vol. 16, n°. 11, p. 4483, 2023.
- [43] J. Yang *et al.*, “A parallel method for solving the DC security-constrained optimal power flow under load uncertainty,” *Energies*, vol. 11, n°. 5, p. 1098, 2018.
- [44] A. Minot, Y. M. Lu et N. Li, “A parallel primal-dual interior-point method for DC optimal power flow,” dans *2016 Power Systems Computation Conference (PSCC)*, 2016, p. 1–7.
- [45] N. Li *et al.*, “Distributed economic dispatch over transmission and distribution networks,” *IEEE Transactions on Smart Grid*, vol. 8, n°. 6, p. 2614–2624, 2016.
- [46] D. Klos, S. Ropke et C. Weinhardt, “A decomposition approach for zonal configuration of electricity markets,” *European Journal of Operational Research*, vol. 245, n°. 2, p. 622–632, 2015.
- [47] T. Valencia-Zuluaga *et al.*, “A fast decomposition method to solve a security-constrained optimal power flow (SCOPF) problem through constraint handling,” *IEEE Access*, vol. 9, p. 52 812–52 824, 2021.
- [48] S. Li *et al.*, “A review of security-constrained optimal power flow calculation method,” dans *AIIPCC 2022 : The Third International Conference on Artificial Intelligence, Information Processing and Cloud Computing*, 2022, p. 1–6. <https://ieeexplore.ieee.org/document/10025882>
- [49] H. Khaloie *et al.*, “Review of machine learning techniques for optimal power flow,” *Applied Energy*, vol. 388, p. 125637, 2025.
- [50] X. Pan *et al.*, “DeepOPF : A deep neural network approach for security-constrained DC optimal power flow,” *IEEE Transactions on Power Systems*, vol. 36, n°. 3, p. 1725–1735, 2020.
- [51] Y. Yang *et al.*, “Fast economic dispatch in smart grids using deep learning : An active constraint screening approach,” *IEEE Internet of Things Journal*, vol. 7, n°. 11, p. 11 030–11 040, 2020.

- [52] D. Deka et S. Misra, “Learning for DC-OPF : Classifying active sets using neural nets,” dans *2019 IEEE Milan PowerTech*, 2019, p. 1–6.
- [53] A. Jahanbani Ardakani et F. Bouffard, “Prediction of umbrella constraints,” dans *2018 Power Systems Computation Conference (PSCC)*, 2018, p. 1–7.
- [54] A. Velloso et P. Van Hentenryck, “Combining deep learning and optimization for preventive security-constrained DC optimal power flow,” *IEEE Transactions on Power Systems*, vol. 36, n°. 4, p. 3618–3628, 2021.
- [55] W. Chen, M. Tanneau et P. Van Hentenryck, “End-to-end feasible optimization proxies for large-scale economic dispatch,” *IEEE Transactions on Power Systems*, vol. 39, n°. 2, p. 4723–4734, mar 2024.
- [56] A. Stratigakos *et al.*, “Interpretable machine learning for DC optimal power flow with feasibility guarantees,” *IEEE Transactions on Power Systems*, vol. 39, n°. 3, p. 5126–5137, 2024.
- [57] W. Chen *et al.*, “Learning optimization proxies for large-scale security-constrained economic dispatch,” *Electric Power Systems Research*, vol. 213, p. 108566, 2022.
- [58] S. Liu *et al.*, “Predicting active constraints set in security-constrained optimal power flow via deep neural network,” dans *Proceedings of the 2021 IEEE Power & Energy Society General Meeting (PESGM)*, 2021, p. 1–5.
- [59] S. Misra, L. Roald et Y.-H. R. Ng, “Learning for constrained optimization : Identifying optimal active constraint sets,” *INFORMS Journal on Computing*, vol. 34, n°. 1, p. 463–480, 2022.
- [60] H. Khazaei et Y. Zhao, “Physics-aware fast learning and inference for predicting active set of DC-OPF,” dans *2022 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2022, p. 1–5.
- [61] S. Boyd et L. Vandenberghe, *Convex optimization*. Cambridge : Cambridge University Press, 2004. <https://web.stanford.edu/~boyd/cvxbook/>
- [62] K. Fukuda, “Frequently asked questions in polyhedral computation,” 2022, eTH Zurich. Version January 15, 2022. <https://people.inf.ethz.ch/fukudak/>
- [63] P. Sumathi et S. Paulraj, “Identification of redundant constraints in large-scale linear programming problems with minimal computational effort,” *Applied Mathematical Sciences*, vol. 7, p. 3963–3974, 2013.
- [64] S. Paulraj et P. Sumathi, “A comparative study of redundant constraints identification methods in linear programming problems,” *Mathematical Problems in Engineering*, vol. 2010, p. 1–16, 2010.

- [65] E. Lalla-Ruiz et S. Voß, “Improving solver performance through redundancy,” *Journal of Systems Science and Systems Engineering*, vol. 25, n°. 3, p. 303–325, 2016.
- [66] F. Bouffard, F. D. Galiana et J. M. Arroyo, “Umbrella contingencies in security-constrained optimal power flow,” dans *15th Power Systems Computation Conference (PSCC)*, Liège, Belgique, 2005, p. 1, session 21, Paper 4, August 22–26.
- [67] R. Weinhold et R. Mieth, “Power market tool (pomato) for the analysis of zonal electricity markets,” *SoftwareX*, vol. 16, p. 100870, 2021.
- [68] A. Jahanbani Ardakani et F. Bouffard, “Identification of umbrella constraints in DC-based security-constrained optimal power flow,” *IEEE Transactions on Power Systems*, vol. 28, n°. 4, p. 3924–3934, 2013.
- [69] ———, “Acceleration of umbrella constraint discovery in generation scheduling problems,” *IEEE Transactions on Power Systems*, vol. 30, n°. 4, p. 2100–2109, 2015.
- [70] R. Madani, J. Lavaei et R. Baldick, “Constraint screening for security analysis of power networks,” *IEEE Transactions on Power Systems*, vol. 32, n°. 3, p. 1828–1838, 2017.
- [71] A. Agrawal *et al.*, “CVXPYLayers : Differentiable convex optimization layers,” 2020, accessed July 14, 2025. <https://github.com/cvxpy/cvxpylayers>
- [72] R. B. Nelsen, *An introduction to copulas*, 2<sup>e</sup> éd., ser. Springer Series in Statistics. Springer Science & Business Media, 2006.
- [73] T. Pham et X. Li, “Constraints and variables reduction for optimal power flow using hierarchical graph neural networks with virtual node-splitting,” arXiv preprint arXiv :2411.06268, 2025. <https://arxiv.org/abs/2411.06268>
- [74] G. Ke *et al.*, “LightGBM : A highly efficient gradient boosting decision tree,” dans *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [75] H. He et E. A. Garcia, “Learning from imbalanced data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, n°. 9, p. 1263–1284, 2009.
- [76] J. Bezanson *et al.*, “Julia : A fresh approach to numerical computing,” *SIAM Review*, vol. 59, n°. 1, p. 65–98, 2017.
- [77] S. Babaeinejad-sarookolaee *et al.*, “The Power Grid Library for benchmarking AC optimal power flow algorithms,” arXiv preprint arXiv :1908.02788, 2021, version published on arXiv. <https://arxiv.org/abs/1908.02788>
- [78] Electric Reliability Council of Texas, “State estimator load report – total ERCOT generation,” the aggregate generation of real and reactive power (MW and MVAR) for the entire ERCOT system. <https://www.ercot.com/mp/data-products/data-product-details?id=NP6-625-CD>

- [79] Q. Huangfu et J. A. J. Hall, “Parallelizing the dual revised simplex method,” *Mathematical Programming Computation*, vol. 10, n°. 1, p. 119–142, 2018.
- [80] J. A. J. Hall, “HiGHS : High-performance open-source software for linear optimization,” Conference presentation at CO@Work 2020, University of Edinburgh, sep 2020. <https://www.hgzs.dev/>
- [81] M. Lubin *et al.*, “JuMP 1.0 : Recent improvements to a modeling language for mathematical optimization,” *Mathematical Programming Computation*, vol. 15, n°. 3, p. 581–589, 2023.
- [82] H. Mittelmann, “Benchmark of simplex LP solvers,” nov 2021, decision tree for optimization software. Archived on March 31, 2022. <https://plato.asu.edu/ftp/lpsimp.html>
- [83] S. M. Lundberg et S.-I. Lee, “A unified approach to interpreting model predictions,” dans *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, 2017, available at <https://arxiv.org/abs/1705.07874>.
- [84] H. Xiang *et al.*, “Robust support vector ordinal regression,” *Information Sciences*, vol. 717, p. 122277, 2025.