UNDERSTANDING THE LMP-LOAD COUPLING: THEORY, EXAMPLES AND AN SVM-BASED DATA-DRIVEN APPROACH

A Thesis

by

XINBO GENG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Le Xie
Committee Members,	Robert Balog
	Srinivas Shakkottai
	Guofei Gu
Head of Department,	Miroslav M. Begovic

December 2015

Major Subject: Electrical Engineering

ABSTRACT

This thesis investigates the fundamental coupling between loads and locational marginal prices (LMPs) in security-constrained economic dispatch (SCED). Theoretical analysis based on multi-parametric programming theory points out the unique one-to-one mapping between load and LMP vectors. Such one-to-one mapping is depicted by the concept of system pattern region (SPR) and identifying SPRs is the key to understanding the LMP-load coupling. The SPR identification problem is considered from a market participant's viewpoint. Built upon the characteristics of SPRs, this paper proposes a data-driven approach to identifying SPRs. It is shown that even without the knowledge of system topology and parameters, the SPRs can be estimated by learning from historical load and price data. A 3-bus system serves as an illustrative example. SPRs and posterior probabilities are visualized and discussed in details for better understandings. The proposed data-driven approach is also examined on the IEEE 24-bus system and 118-bus system. The simulation results on both systems illustrate that the proposed method is effective, and some computational issues are discussed.

DEDICATION

To my parents.

ACKNOWLEDGEMENTS

I wish to thank my committee members who were more than generous with their expertise and precious time.

A special thanks to Dr. Le Xie, my advisor, for providing me a pressure-free environment and numerous hours discussing various topics together.

I want to thank my friends and colleagues who always support and inspire me.

Very special thanks go to the swimming pool and spa of Texas A&M University, these places are where I resurge after consecutive thinking.

NOMENCLATURE

$c_i(P_{G_i})$: the generation cost function of generator i

$\mathbb{D}$: the load space

$\mathcal{D}$: the set of all feasible vectors of loads

η^+, η^- : multipliers related with generation capacity constraints

H : shift factor matrix

$\mathcal{J}$: set of indices of all the constraints $\mathcal{J} = \{1, 2, \cdots, n_c\}$

λ : LMP vector

λ_1 : LMP of the slack bus (energy component)

LMP : Locational Marginal Price

μ^+, μ^- : multipliers related with transmission capacity constraints

n_b : number of buses in the system

n_c : number of constraints in the SCED formulatio

n_g : number of generators in the syste

n_l : number of transmission lines in the system

MLP : Multi-parametric Linear Programming

π : system pattern $\pi = (\mathcal{B}, \mathcal{N})$

$\mathcal{B}$: the set of the indices of binding constraints in primal SCED

$\mathcal{N}$: the set of the indices of non-binding constraints in primal SCED

P_G, P_D : generation vector, load vector

s : vector slack variables in the primal SCED problem

$\mathcal{S}_\pi$: SPR related with system pattern π

SCED : Security-constrained Economic Dispatch

SPR : System Pattern Region

SVM : Support Vector Machine

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

A fundamental issue with electricity market operation is to understand the impact of operating conditions (e.g. load levels at each bus) on the locational marginal prices (LMPs). This paper examines this key issue of the relationship between nodal load levels and LMPs. This issue is further compounded by the increasing levels of demand response and variable resources in the grid.

1.1 Literature Review

In the power systems literature, reference [5] is among the pioneering work that uses perturbation techniques to compute the sensitivities of the dual variables in SCED (e.g. LMPs) with respect to parameters (e.g. the nodal load levels). This sensitivity calculation method is widely used in subsequent researches. However, this approach is only valid for small changes and the marginal generator stays the same. Reference [15] observed the "step changes" of LMPs with respect to increasing system load level and discovered that new binding constraints (transmission or generation) are the reason of the "step changes". This is followed by further analysis on identifying the critical load levels (CLLs) that trigger such step changes of LMPs [16][2][3]. This line of work assumes that the system load change is distributed to each bus proportional to the base case load, which, in many instances, do not necessarily represent the real-world situations. Reference [21] analyzed this problem using quadratic-linear programming (QLP) and the concepts of *system patterns* and *system pattern regions (SPRs)* were first introduced. The SPRs depict the relationship between loads and LMPs in the whole load space, which is not confined in a small neighborhood of an operating point or constrained by a specific load distribution pattern.

This paper is inspired by [21] but focuses on the case of piecewise linear generation costs, instead of the quadratic cost case in [21]. The reasons that we study the piecewise linear cost case are

- piecewise linear cost curves are often quite representative of the market practice in the real world.
- some new theoretical results based on piecewise linear cost curves are derived, and are generalizable towards quadratic cost cases.

Characterizing the SPRs would provide important insights to both system operators and market participants. Reference [13] advances the theory of SPR from system operator's perspective where the knowledge of system topology and parameters is available. For market participants, such knowledge is not necessarily available. Our previous work [10] examines the issue from market participant's viewpoint and applies the geometric features of SPRs to identify them.

1.2 Contribution of This Paper

This paper significantly advances our previous work [10] by

1. completing the theoretical characterization of SPRs as a function of nodal load levels;
2. proposing a computational algorithm to identify SPRs using historical data;
3. introducing the posterior probabilities of SPRs with the presence of uncertain system parameters such as transmission limits.

The rest of the paper is organized as follows. Section 2 formulates SCED in the form of multi-parametric linear programming. Theoretical analysis of relationship between nodal load levels and LMPs is presented in Section 3. A 3-bus system serves

as a illustrative example. A data-driven algorithm for market participants to identify SPRs is described in Section 4. Section 5 illustrates the performance of the algorithm on the IEEE 118-bus system. Section 6 presents the concluding remarks and future work.

2. PRELIMINARIES

2.1 Notations

The notations of this paper are summarized below: mathematical symbols in hollowed-out shapes (e.g. $\mathbb{R}$) represent spaces and symbols in Calligra font (e.g. $\mathcal{S}_\pi$) stand for sets. The superscript "*" indicates the variable is optimal, " $\hat{}$ " denotes estimated values (e.g. $\hat{\lambda}$). Variables with " $\bar{}$ " are expectations or average values (e.g. $\bar{\lambda}$). "$\intercal$" denotes the transpose of a vector or matrix (e.g. $\mathbf{1}_n^\intercal$). The subscript "$_i$" represents the ith element of the vector (e.g. P_{G_i}), and the superscript "$^{(i)}$" represents the ith element in a set (e.g. $P_D^{(i)}$). The vector of $n \times 1$ ones, matrix of $m \times n$ zeros and the $n \times n$ identity matrix are denoted by $\mathbf{1}_n$ and $\mathbf{0}_{m\times n}$ and $\mathbf{I}_n$ respectively.

2.2 Security Constrained Economic Dispatch

In real-time energy market operations, the LMPs are the results from the security-constrained economic dispatch (SCED), which is formulated as follows:

$$\begin{aligned}
\underset{P_G}{\text{minimize}} \quad & \sum_{i=1}^{n_b} c_i(P_{G_i}) \\
\text{subject to} \quad & \sum_{i=1}^{n_b} P_{G_i} = \sum_{j=1}^{n_b} P_{D_j} && : \lambda_1 \\
& -F^+ \le H(P_G - P_D) \le F^+ && : \mu^+, \mu^- \\
& P_G^- \le P_G \le P_G^+ && : \eta^+, \eta^-
\end{aligned} \tag{2.1}$$

where P_G is the generation vector, and P_D is the load vector. Without loss of generality, we assume there are both generation and load at each bus[1]. Let n_b

[1]For the discussions about this assumption, please refer to Appendix A.

denote the number of buses and n_l denote the number of transmission lines, then $P_G, P_D \in \mathbb{R}^{n_b}$. $H \in \mathbb{R}^{n_l \times n_b}$ is the shift factor matrix. In Eqn. (2.1), bus 1 is assumed to be the slack bus. For the sake of simplicity, line losses are not modeled explicitly in this formulation.

The objective of SCED is to minimize the total generation cost and satisfy the transmission and generation capacity constraints while keeping the real-time balance between supply and demand. The generation cost function $c_i(P_{G_i})$ of generator i is increasing and convex, and it is usually regarded as a quadratic function or approximated by a piecewise linear function. To better reflect the current practice in electricity markets, this paper studies the SCED problem with piecewise linear generator bidding functions. And for the consideration of simplicity, the simplest form, i.e. $\sum_{i=1}^{n_b} c_i(P_{G_i}) = c^\intercal P_G$ is being considered in this paper[2]. The associated Lagrangian Multipliers $[\lambda_1; \mu^+; \mu^-; \eta^+; \eta^-]$ play a pivotal role in market pricing, and the vector of locational marginal prices (LMPs) λ can be calculated as follows [20]:

$$\lambda = \lambda_1 \mathbf{1}_{n_b} + H^\intercal(\mu^+ - \mu^-) \tag{2.2}$$

2.3 Multi-parametric Linear Programming

In real-world market operations, the parameters associated with the SCED above are typically time-varying. Therefore, it is essential to understand the effects of parameters on the optimality of the problem. Multi-parametric Programming (MP) problem aims at exploring the characteristics of an optimization problem which depending on a *vector of parameters* [4]. Multi-parametric Linear Programming (MLP) theory, which is the foundation of this paper, pays special attention to Linear Programming (LP) problems. The standard form of the MLP problem is stated as

[2]For the discussions about this assumption, please refer to Appendix A.

follows:

$$\text{Primal:} \qquad \min_x\{c^\intercal x : Gx \leq W + F\theta\} \tag{2.3}$$

$$\text{Dual:} \quad \max_y\{-(W + F\theta)^\intercal y : G^\intercal x = -c, x \geq 0\} \tag{2.4}$$

where $\theta \in \mathbb{R}^s$ is the vector of parameters, $c \in \mathbb{R}^n$ is the vector of costs. And θ belongs to some parameter space Θ.

In other references (e.g. [1, 7]), the primal form of the MLP problem is different. For the consideration of convenience of analyzing SCED problem, we follow the formulations in [4]. Those two forms are interchangeable.

In this paper, we would like to understand the impact of parameters (i.e., load levels, line capacities, etc) on the outcome of SCED (namely, the prices). In the next Section, we pose the problem in view of MLP, and analyze the theoretical properties. An important assumption we have is that unit commitment results stay unchanged during the SCED time frame. Future work will investigate the impact of change of unit commitment on the prices.

3. SCED ANALYSIS VIA MLP

3.1 Definitions and Theorems

Consider the SCED problem in MLP form[1]:

$$\text{Primal:} \quad \min\{c^\intercal P_G : AP_G + s = b + WP_D, s \geq 0\} \tag{3.1}$$

$$\text{Dual:} \quad \max\{-(b + WP_D)^\intercal y : A^\intercal y = -c, y \geq 0\} \tag{3.2}$$

where:

$$A = \begin{bmatrix} \mathbf{1}_{n_b}^\intercal \\ -\mathbf{1}_{n_b}^\intercal \\ H \\ -H \\ \mathbf{I}_{n_b} \\ -\mathbf{I}_{n_b} \end{bmatrix}, b = \begin{bmatrix} 0 \\ 0 \\ F^+ \\ -F^+ \\ P_G^+ \\ -P_G^- \end{bmatrix}, W = \begin{bmatrix} \mathbf{1}_{n_b}^\intercal \\ -\mathbf{1}_{n_b}^\intercal \\ H \\ -H \\ \mathbf{0}_{n_b \times n_b} \\ \mathbf{0}_{n_b \times n_b} \end{bmatrix} \tag{3.3}$$

The load vector P_D is the vector of parameters θ, and the load space $\mathbb{D}$ is the parameter space Θ. Since not every P_D in the load space leads to a feasible SCED problem, $\mathcal{D} \in \mathbb{D}$ denotes the set of all feasible vectors of loads. [7] shows that $\mathcal{D}$ is a convex polyhedron in $\mathbb{D}$.

Definition 1 (Optimal Partition/System Pattern)**.** For a load vector $P_D \in \mathcal{D}$, we could find a finite optimal solution P_G^* and s^*. Let $\mathcal{J} = \{1, 2, \cdots, n_c\}$ denote the index set of constraints where $n_c = 2+2n_l+2n_g$ for Eqn. (3.1). The *optimal partition*

[1]The supply and demand balance constraint is rewritten as two equivalent inequalities (first two rows of matrices A, W and vector b).

$\pi = (\mathcal{B}, \mathcal{N})$ of the set $\mathcal{J}$ is defined as follows:

$$\mathcal{B}(P_D) := \{i : s_i^* = 0 \text{ for } P_D \in \mathcal{D}\} \tag{3.4}$$

$$\mathcal{N}(P_D) := \{j : s_j^* > 0 \text{ for } P_D \in \mathcal{D}\} \tag{3.5}$$

Or in the dual form (Eqn. (3.2)):

$$\mathcal{B}(P_D) := \{i : y_i^* > 0 \text{ for } P_D \in \mathcal{D}\} \tag{3.6}$$

$$\mathcal{N}(P_D) := \{j : y_j^* = 0 \text{ for } P_D \in \mathcal{D}\} \tag{3.7}$$

Obviously, $\mathcal{B} \cap \mathcal{N} = \emptyset$ and $\mathcal{B} \cup \mathcal{N} = \mathcal{J}$. The optimal partition $\pi = (\mathcal{B}, \mathcal{N})$ divides the index set into two parts: binding constraints $\mathcal{B}$ and non-binding constraints $\mathcal{N}$. In SCED, the *optimal partition* represents the status of the system (e.g. congested lines, marginal generators), and is called *system pattern*.

Definition 2 (Critical Region/System Pattern Region)**.** The concept *critical region* refers to the set of vectors of parameters which lead to the same optimal partition (system pattern) $\pi = (\mathcal{B}_\pi, \mathcal{N}_\pi)$:

$$\mathcal{S}_\pi := \{P_D \in \mathcal{D} : \mathcal{B}(P_D) = \mathcal{B}_\pi\} \tag{3.8}$$

For the consideration of consistency, the *critical region* is called *system pattern region* (SPR) in this paper.

According to the definitions, each SPR is one-to-one mapped to a system pattern, the SPRs are therefore disjoint and the union of all the SPRs is the feasible set of vectors of loads: $\cup_i \mathcal{S}_{\pi_i} = \mathcal{D}$. All the SPRs together represent a specific partition of the load space. The features of SPRs, which directly inherit from critical regions in

MLP theory, are summarized as follows:

Theorem 1. *The load space could be decomposed into many SPRs. Each SPR is a convex polytope. The relative interiors of SPRs are disjoint convex sets and each corresponds to a unique system pattern [21]. There exists a separating hyperplane between any two SPRs [10].*

Lemma 1. *If the problem is not degenerate, then the partition of the load space is unique, and S_π is an open polyhedron of the same dimension as $\mathbb{D}$ [6].*

Lemma 2. *The optimal value function $f^* = c^\intercal P_G^*(P_D)$ is convex and piecewise affine over $\mathcal{D}$, and affine in each SPR. The optimal solution P_G^* within an SPR is an affine function of the load vector P_D [8].*

Lemma 3 (Complementary Slackness)**.** *According to complementary slackness:*

$$A_{\mathcal{B}} P_G^* = (b + W P_D)_{\mathcal{B}} \tag{3.9}$$

$$A_{\mathcal{N}} P_G^* < (b + W P_D)_{\mathcal{N}} \tag{3.10}$$

$$A_{\mathcal{B}}^\intercal y_{\mathcal{B}} = -c, y_{\mathcal{B}} > 0 \tag{3.11}$$

$$y_{\mathcal{N}} = 0 \tag{3.12}$$

where the $(\cdot)_{\mathcal{B}}$ is the sub-matrix or the sub-vector whose row indices are in set $\mathcal{B}$, same meaning applies for $(\cdot)_{\mathcal{N}}$.

Remark 1. The supply-demand balance equality constraint in SCED is rewritten as two inequalities (Eqn. (3.1)). These two inequalities will always be binding and appear in the binding constraint set $\mathcal{B}$ at the same time. One of them is redundant and therefore eliminated from set $\mathcal{B}$. *In the remaining part of the paper, set $\mathcal{B}$ denotes the set after elimination.*

Remark 2. If the problem is not degenerate, the cardinality of binding constraint set $\mathcal{B}$ is equal to the number of decision variables (i.e. number of generators n_g) [2]. The matrix $A_{\mathcal{B}}$ is invertible and P_G^* is uniquely determined by $A_{\mathcal{B}}^{-1}(b+WP_D)_{\mathcal{B}}$. This is equivalent with Lemma 2 that the optimal solution is an affine function of the load vectors within the same SPR.

Lemma 4 (Analytical Form of SPRs)**.** *Let* $\mathbf{I}_{\mathcal{B}} \cdot (b + WP_D)$ *represent the sub-vector* $(b+WP_D)_{\mathcal{B}}$*, where* $\mathbf{I}_{\mathcal{B}}$ *is the sub-matrix of the identity matrix whose row indices are in set* $\mathcal{B}$*. Then the analytical form of the SPRs could be solved from Eqn. (3.9) and Eqn. (3.10) as follows:*

$$(\mathbf{I}_{\mathcal{N}}A \cdot (\mathbf{I}_{\mathcal{B}}A)^{-1}\mathbf{I}_{\mathcal{B}} - \mathbf{I}_{\mathcal{N}})(b + WP_D) < 0 \tag{3.13}$$

Lemma 5. *Within each SPR, the vector of LMPs is unique.*

The proof of this lemma follows Eqn. (3.11) (dual form of system pattern definition) . Since the system pattern π is unique within an SPR S_π, therefore the solution y^* is unique for any $P_D \in \mathcal{S}_\pi$ And the vector of LMPs can be calculated using Eqn. (2.2). This lemma also illustrates that the LMP vectors are discrete by nature in the case of linear costs.

Theorem 2. *Different SPRs have different LMP vectors.*

Lemma 2 shows that the optimal value function is piecewise affine, which means the optimal value function is composed of several hyperplanes and each hyperplane corresponds to an SPR. The LMP vectors are the partial derivatives of the optimal value function f^* over the load vector P_D. If two SPRs have the same LMP vector,

[2]This is consistent with the statement that the number of marginal generators equals to the number of congested lines plus one.

then those two hyperplanes are parallel to each other. But a piecewise affine function with two parallel segments cannot be a convex function, this is contradictory to Lemma 2 which also proves that f^* is convex. Therefore different SPRs have different LMP vectors. A more rigorous proof can be found in Appendix F.

3.2 SPRs with Varying Parameters

Section 3.1 shows construction properties of the load space with fixed parameters of the system (e.g. transmission constraints). However, these parameters might also be time-varying (e.g. dynamic line ratings). This section reveals more features of SPRs with respect to varying factors in the system.

Remark 3. Eqn. (3.13) could be written as:

$$(\mathbf{I}_{\mathcal{N}}A(\mathbf{I}_{\mathcal{B}}A)^{-1}\mathbf{I}_{\mathcal{B}} - \mathbf{I}_{\mathcal{N}}) \cdot WP_D < (\mathbf{I}_{\mathcal{N}} - \mathbf{I}_{\mathcal{N}}A(\mathbf{I}_{\mathcal{B}}A)^{-1}\mathbf{I}_{\mathcal{B}})b \tag{3.14}$$

This indicates the shape of the SPR $\mathcal{S}_\pi$ only depends on two factors: (1) the corresponding system pattern $\pi = (\mathcal{B}, \mathcal{N})$; (2) matrices A and W, namely the shift factor matrix H according to Eqn. (3.3). Small change of vector b (e.g. changing transmission limits) only leads to a parallel shift of the SPRs' boundaries.

3.3 3 Bus System Example

An illustrative example is provided in this section. All the parameters of the 3-bus system are shown in Fig. 3.1. The 3-bus system in Fig. 3.1 is analyzed using Multi-Parametric Toolbox 3.0 (MPT 3.0) [12].

Fig. (3.2a) and Fig. (3.2b) show that the optimal solution P_G^* and the optimal value function is piecewise affine over $\mathcal{D}$.

Fig. (3.3a) shows the SPRs when the vector of generation costs c is $[20; 50; 100]$. Within each SPR, the LMPs are identical. The analytical form of SPRs and corre-

sponding LMP vectors are calculated using Eqn. (3.13) and Eqn. (3.9). Details of Fig. (3.3a) are summarized in [9].

Notice that P_{D_2} and P_{D_3} in Fig. (3.3a) could be negative. This is for the consideration of renewable resources in the system, which are typically considered as negative loads.

The features of SPRs with varying transmission limits can be observed from Fig. (3.3a) and Fig. (3.3b). With decreasing transmission limits, the shape of all the SPRs remain the same; SPR#2 and SPR#6 shrink, SPR#3, #4 and #5 expand.

3.4 Practical Issues with SPRs

The theoretical analysis above reveals that the load space could be partitioned into many regions with fascinating features. This overall structure of the load space could help solve the SCED problem and shift part of the online computational burden to offline [14]. The number of SPRs in the load space could help evaluate the offline computational burden.

The number of distinct system pattern regions is finite, an upper bound is $2^{n_g-1} \times C_{n_g+n_l}^{n_g-1}$, which is the number of all the possible system patterns. But this number is actually a very loose upper bound. Some benchmark systems are analyzed using MPT 3.0 [12]. All the test systems except the 3-bus one are IEEE standard test systems, and all the settings are from the case files of Matpower. The load vector is the vector of parameter for analysis.

As shown in Table 3.1[3], the number of SPRs grows rapidly with the scale of systems. There are around 10^4 buses in a real power system (e.g. 1.2×10^4 in ERCOT), consequently there would be too many SPRs to be considered. This is a major issue to analyze the SCED problem via MLP theory.

[3]The MPT toolbox also analyzed the IEEE 24-bus system. It took 7 days to find 2.7×10^6 SPRs but still could not explore all the SPRs in the load space.

Fig. 3.4 also demonstrates the growing of the number of SPRs. The x-axis of Fig. 3.4 is the number of buses of test systems, and the y-axis is the log of the total number of SPRs. The black circles are calculated numbers of SPRs, the red line is the upper bound provided by $2^{n_g-1} \times C_{n_g+n_l}^{n_g-1}$. Although this upper bound is relatively loose, it satisfyingly describes the trend of number of SPRs. The green line indicates an attempt to fit the actual number of SPRs using the upper bound. The equation of the green line is $2^{n_g-1} \times C_{n_g+n_l}^{n_g-1}/1.12^{n_g+n_l}$.

To avoid this issue, [14] focuses on the major stochastic factors (e.g. stochastic generation) in the system. And [21] points out that only some subsets of the complete theoretical load space could be achievable thus helpful in practice. This paper follows the assumptions in [21], namely, starting from the load data and exploring the practical SPRs instead of the whole load space.

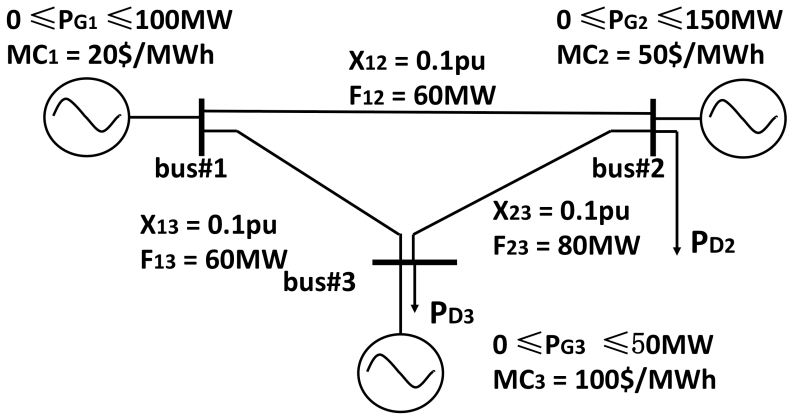
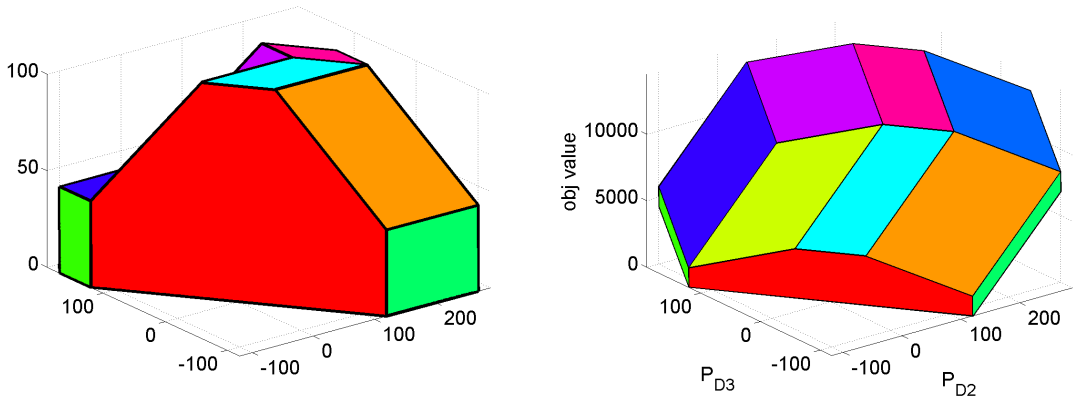


Figure 3.1: 3 Bus System



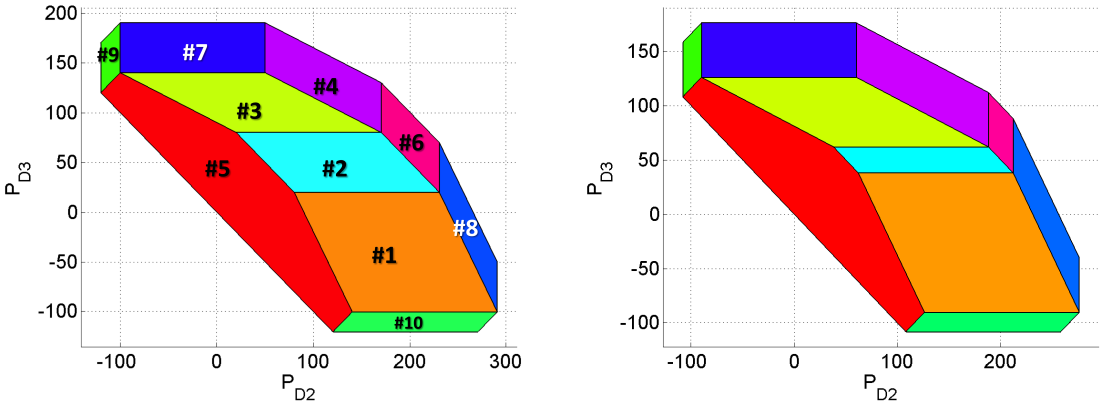
(a) Primal Solution

(b) Optimal Value Function

Figure 3.2: Primal Solution and Optimal Value Function (3-bus System)

Table 3.1: Number of SPRs in Test Systems

System Info	**Number of SPRs**	**Computation Time (s)**
3 Bus System (Fig.3.1)	10	0.7
IEEE 6 Bus System	20	2.1
IEEE 9 Bus System	15	1.5
IEEE 14 Bus System	1470	31



(a) Line Limits: (60, 60, 80)

(b) Line Limits: (54, 54, 72)

Figure 3.3: System Pattern Regions with Varying Transmission Limits

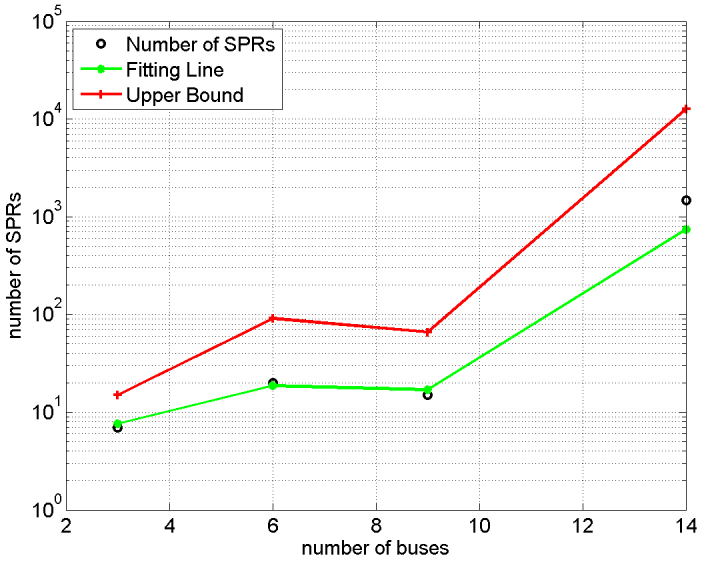


Figure 3.4: Number of SPRs

4. A DATA-DRIVEN APPROACH TO IDENTIFYING SPRS*

The SPRs depict the fundamental coupling between loads and LMP vectors. Massive historical data could help market participants estimate SPRs, understand the load-LMP coupling and then forecast LMPs. This section proposes a data-driven method to identify SPRs, which is a significant improvement of the basic method in [10][1] by considering varying system parameters and the probabilistic nature of system parameters.

4.1 The SPR Identification Problem

4.1.1 SPR Identification as a Classification Problem

A *classifier* is an algorithm to give a *label* y to each *feature* vector x. The feature vectors sharing the same labels belong to the same *class*. The objective of the classification problem is to find the best classifier which could classify each feature vector accurately. For the parametric classifiers, there is always a training set, i.e. a group of feature vectors whose labels are known. There are usually two steps in a classification problem: training and classifying. *Training* usually means solving an optimization problem over the training set to find the best parameters of the classifier. And *classifying* is to classify a new feature vector with the trained classifier.

According to Section 3.1, the load vectors in an SPR share many common features (e.g. vectors of LMPs). Theorem 2 proved that the LMP vectors are distinct for different SPRs. Therefore, one SPR can be regarded as a *class* and the LMP vector is

*The description of the basic method is based on a previous publication: "A Data-driven Approach to Identifying System Pattern Regions in Market Operations" by Xinbo Geng and Le Xie, in Proceedings of Power and Energy Society General Meeting, Denver, CO, 2015. With permission from IEEE.

the *label* of each class. Theorem 1 proves the existence of the separating hyperplanes. Since each separating hyperplane labels two SPRs with different sides, it turned out that the separating hyperplanes are classifiers and the key of identifying SPRs is to find optimal hyperplanes, which is exactly the objective of Support Vector Machine (SVM).

4.1.2 SPR Identification with SVM

Suppose there is a set of labeled load vectors for training and those load vectors belong to only *two* distinct SPRs (labels $y^{(i)} \in \{1, -1\}$). Then the SPR identification problem with a *binary* SVM classifier (separable case) is stated below:

$$\min_{w,b} \quad \frac{1}{2} w^\intercal w \tag{4.1}$$

$$\text{s.t} \quad y^{(i)}(w^\intercal P_D^{(i)} - b) \geq 1, y^{(i)} \in \{-1, 1\} \tag{4.2}$$

The word "binary" here specifies only two classes (i.e. SPRs) are being considered. Eqn. (4.2) is feasible only when the two SPRs are not overlapping and there exists at least one hyperplane thoroughly separating them (separable case). For any load vector P_D in the load space, $w^\intercal P_D - b = 0$ represents the separating hyperplane where w is the norm vector to the hyperplane. Two lines satisfying $w^\intercal P_D - b = \pm 1$ separate all the training data and formulate an area with no points inside. This empty area is called *margin*. The width of the margin is $2/||w||$, which is the distance between those two lines. The optimal solution refers to the separating hyperplane which maximizes the width of the margin $2/||w||$, therefore the objective of the binary SVM classifier is to minimize the norm of vector w.

Due to the existence of multiple SPRs, multi-class classifiers are needed. Since Theorem 1 guarantees the existence of separating hyperplanes between every pair of

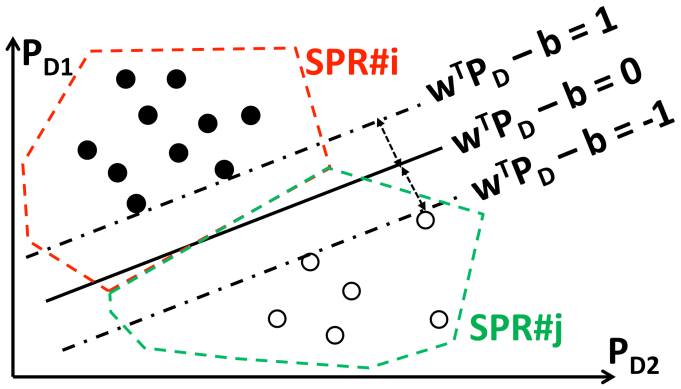


Figure 4.1: SPR Identification Problem with SVM

SPRs, the "one-vs-one" multi-class SVM classifier is incorporated in the data-driven approach to identifying SPRs. Detailed procedures are summarized in Section 4.2.

4.2 A Data-driven Approach

4.2.1 SPR Identification with Varying System Parameters

When the system parameters are varying (e.g. dynamic line ratings), two SPRs may overlap with each other (shown in Section 3.2, Fig. 3.3 and Fig. 5.1). The SPR identification problem is no longer a separable case as in [10] (Eqn. (4.1) and Eqn. (4.2)). The SVM classifier needs to incorporate *soft margins* to allow some tolerance of classification error. The slack variable s_i is added to Eqn. (4.1) and Eqn. (4.2) representing tolerant errors. Penalties of violation $C\sum_i s_i$ are added in the objective function. Large C means that the extent of tolerance is low.

$$\min_{w,b,s} \quad \tfrac{1}{2}w^\intercal w + C\sum_i s^{(i)} \tag{4.3}$$

$$\text{s.t} \quad y^{(i)}(w^\intercal P_D^{(i)} - b) \geq 1 - s^{(i)} \tag{4.4}$$

$$s^{(i)} \geq 0, y^{(i)} \in \{-1, 1\}$$

4.2.2 Fitting Posterior Probabilities of SVM Classifier

Estimating the posterior probability $P(\text{class}|\text{input})$ is very helpful in practical problems [18]. When identifying SPRs, knowing the probability $P(y = i|P_D$ and $y \in \{1, 2, \cdots, n\})$ is not only about knowing the result $y = i$ (P_D belongs to SPR#i), but also understanding the confidence or possible risk. The market participants could accordingly adjust their bidding strategy and reduce possible loss.

Although the posterior probabilities are desired, the standard SVM algorithm provides an uncalibrated value which is not a probability as output [18]. Modifications are needed to calculate the *binary* posterior probabilities $P(y = i|P_D$ and $y \in \{i, j\})$. Common practice is to add a link function to the binary SVM classifier and train the data to fit the link function. Some typical link functions include Gaussian approximations [11] and sigmoid functions [18]. In this paper, the sigmoid link function is selected due to its general better performance than other choices [18].

In general, there are more than two SPRs. What we really want to know is the *multi-class* posterior probabilities $P(y = i|P_D$ and $y \in \{1, 2, \cdots, n\})$. For short, we will use $P(y = i|P_D)$ to represent multi-class posterior probabilities. [11] proposed a well-accepted algorithm to calculate multi-class posterior probabilities from pairwise binary posterior probabilities. This algorithm is incorporated in our approach and briefly summarized in Appendix B.

4.2.3 The Data-driven Approach

There are three steps in the proposed data-driven approach (Fig. 4.2):

4.2.3.0.1 Training Suppose there are n different SPRs in the training data set. Each time two SPRs are selected, trained and we get a *binary* SVM classifier. This pairwise training procedure is repeated $C_n^2 = n(n-1)/2$ times and we collect $n(n-1)/2$ binary classifiers, namely the $n(n-1)/2$ separating hyperplanes between any

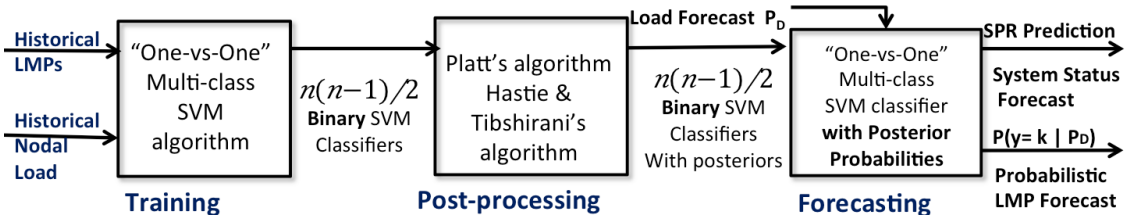


Figure 4.2: The Data-driven Approach

two out of n SPRs.

4.2.3.0.2 Data Post-processing Calculate posterior probabilities $P(y = i|P_D)$ for $i = 1, 2, \cdots, n$ by applying Platt's algorithm and then Hastie and Tibshirani's algorithm.

4.2.3.0.3 Classifying/Predicting Given load forecast P_D, there are two different approaches to predicting, which are termed as "max-vote-wins" and "max-posterior-wins", respectively:

- *(max-vote-wins:)* each binary classifier provides a classification result (vote) for the load forecast P_D, the SPR which collects the most votes will be the final classification result. The load forecast P_D is therefore pinpointed to an SPR, whose label vector (LMP vector) is the forecast of LMPs. This was proposed in [10] and it does not require data post-processing (step (b)).

- *(max-posterior-wins:)* step (b) provides $n(n-1)/2$ trained binary SVM classifiers with posterior probabilities, the classification result SPR#i^* is the class with the largest multi-class posterior probability: $i^* = \arg\max_i P(y = i|P_D)$.

It is worth noting that the proposed approach considered but not limited to varying system parameters. It could be generalized to many other scenarios with overlapping SPRs in the data, e.g. estimating SPRs with nodal load data of one area instead of

the whole system.

4.2.4 Probabilistic LMP Forecast

With the "max-posterior-wins" method, we obtain not only the SPR that P_D belongs to, but also the distribution of LMPs. The probabilistic LMP forecast is built on the estimated multi-class posterior probabilities $P(y = i|P_D)$:

$$\hat{\lambda}(P_D) = \mathbb{E}[\lambda] = \sum_{i=1}^{n} \lambda^{(i)} P(y = i|P_D) \tag{4.5}$$

where $\lambda^{(i)}$ is the vector of LMPs of SPR$\#i$.

The "max-vote-wins" methods forecasts the LMP vector in a deterministic way: $\hat{\lambda}(P_D) = \lambda^{(i^*)}$ where i^* is the index of the SPR winning the most votes.

5. CASE STUDY

In this section, we illustrate the proposed data-driven approaches in two systems. We first introduce the performance metrics for the proposed algorithm.

5.1 Performance Metrics

5.1.1 3-fold Cross Validation

To evaluate the performance of the model to an independent data set and avoid overfitting, the *k-fold cross validation* technique is being used. In k-fold cross-validation, the overall data set is randomly and evenly partitioned into k subsets. Every time a subset is chosen as validation data set, and the remaining $k-1$ subsets are used for training. This cross-validation process is repeated k times (k folds), and each subset serves as the validation data set once. The 3-fold cross validation is being used in this paper.

5.1.2 Classification Accuracy

Classification accuracy is the major criteria to evaluate the performance of the method. The classification accuracy α is the ratio of the correctly classified points in the validation data set. When incorporating 3-fold cross validation, the classification accuracy of each fold $(\alpha_1, \alpha_2, \alpha_3)$ is calculated first, then the overall performance of the method is evaluated by the average classification accuracy $\bar{\alpha} = (\alpha_1 + \alpha_2 + \alpha_3)/3$.

5.1.3 LMP Forecast Accuracy

The performance of the LMP forecast is evaluated by *LMP forecast accuracy*, which is the average forecast accuracy of all the validation data points ($j = 1, 2, \cdots, n_v$)

over all the buses ($i = 1, 2, \cdots, n_b$):

$$\beta = \frac{1}{n_b}\frac{1}{n_v}\sum_{i=1}^{n_b}\sum_{j=1}^{n_v}\frac{|\hat{\lambda}_i[j] - \lambda_i[j]|}{\lambda_i[j]} \tag{5.1}$$

5.2 An Illustrative Example

We start with an illustrative 3-bus system example as shown in Fig. 3.1. This succinct example provides key insights and visualization of the proposed method.

5.2.1 Data

The data set is generated using Matpower [22] with the following assumptions: (1) the load vector is evenly distributed in the load space; (2) the transmission limits F is time-varying. For simplicity, F is assumed to satisfy Gaussian distribution: $F \sim N(\bar{F}, \Sigma)$, where $\bar{F} = [60, 60, 80]^\intercal$ and 10% standard deviation $\Sigma = [6, 0, 0; 0, 6, 0; 0, 0, 8]$. All the points generated are visualized in Fig. (5.1). Each color represents an SPR. As shown in Fig. (5.1), some SPRs (e.g. blue and red) are overlapping.

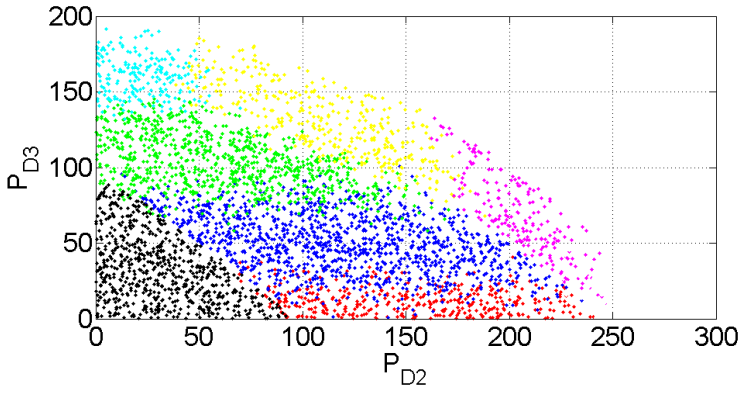


Figure 5.1: Generated Dataset

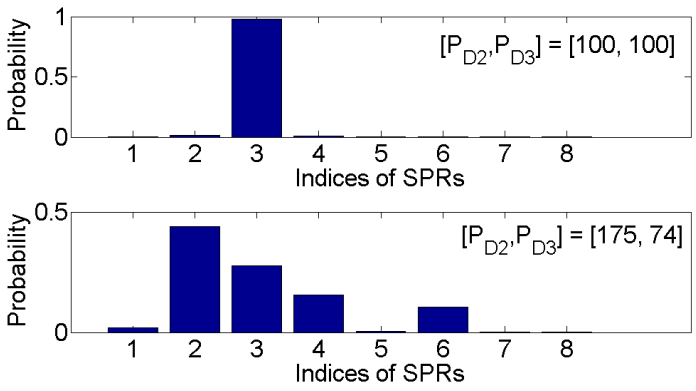


Figure 5.2: Posterior Probabilities of Given Load Vectors

5.2.2 Simulation Results

Table 5.1 summarizes the classification accuracies. Both approaches, Max-Vote-Wins and Max-Posterior-Wins, could accurately estimate SPRs and perform prediction.

Table 5.1: Classification Accuracy (3 Bus System)

Method	**Fold#1**	**Fold#2**	**Fold#3**	**Average**
Max-Vote-Wins	88.40%	89.92%	89.77%	89.36%
Max-Posterior-Wins	88.63%	89.69%	90.00%	89.44%

5.2.3 Posterior Probabilities

The estimated posterior probabilities are visualized in Fig. (5.3). There are 8 surfaces in the figure, which relates with the overall 8 SPRs and represents $P(y = i|P_D)$ for $i = 1, 2, \cdots, 8$ respectively. Due to varying line limits and overlapping SPRs, the eight surfaces intersect with each other and formulate some "mountains"

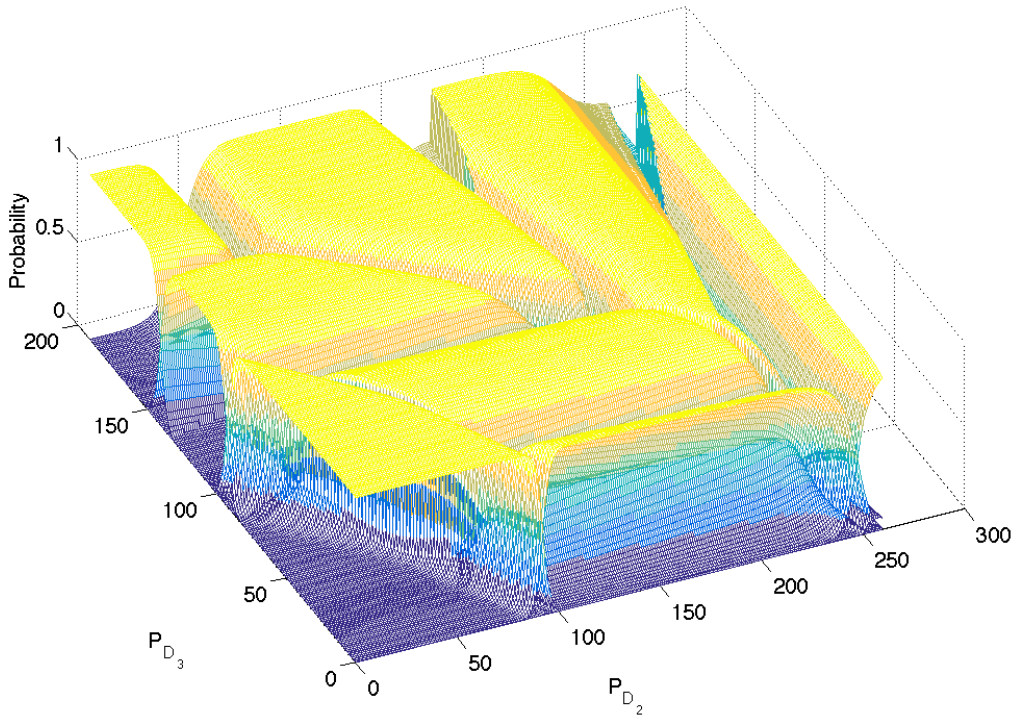


Figure 5.3: Posterior Probability Surfaces

and "valleys". The "mountains" correspond to the inner parts of SPRs, where the overlapping of SPRs is almost impossible to happen. And the "valleys" always locate at the boundaries among SPRs. Two examples are provided in Fig. (5.2). When $P_{D_2} = P_{D_3} = 100$MW, the load vector is in the middle of SPR#3, $P(y = 3|P_D = [100; 100]) \approx 1$ and other posterior probabilities are close to zero. However, when $P_{D_2} = 175$MW and $P_{D_3} = 74$MW, the load vector is close to the boundaries of SPR#2,#3,#4 and #6. In this case, $P(y = 2|P_D = 0.439)$, $P(y = 2|P_D = 0.276)$, $P(y = 2|P_D = 0.156)$, $P(y = 2|P_D = 0.104)$ and other probabilities are close to zero. With the estimated posterior probabilities, market participants could estimate the risks associated with LMPs given different combinations of load levels without running numerous simulations.

5.3 118 Bus System

A more comprehensive case study is conducted on the IEEE 118 bus system for the following objectives: (1) to evaluate the performance and computational burden of the data-driven approach on a complex system with realistic settings; (2) to demonstrate a possible application (LMP forecast) for market participants with estimated posterior probabilities.

5.3.1 System Configuration

Most of the system settings follow the *IEEE 118-bus, 54-unit, 24-hour system* in [17] but with the following changes: the lower bounds of generators are set to zero, but the upper bounds of generators remain the same as in [17]. Exact details of generation cost are summarized in [9]. To consider dynamic line ratings, we assume the line limits are random and satisfy normal distribution $N(\mu, \sigma)$. The expectation of line limits μ is the same as [17], only the limit of line 127 (from bus 81 to bus 80) is revised from 500MW to 100MW; the standard deviation σ is set to be 5% of the

expectation, which means 95% of the time the line limits vary within 10% from the mean μ.

5.3.2 Load

[17] also provides an hourly system load profile and a bus load distribution profile (Table 5 and Table 6 in [17]). With linear interpolation, the hourly system load profile is modified to be 5min-based. Since the load is stochastic by nature, we assume the load at each bus satisfies normal distribution $\mathcal{N}(\mu, \sigma)$. The expectation μ of each nodal load is calculated from the system load profile and bus load distribution profile, the standard deviation σ is set to be 5% of the expectation, i.e. $\sigma = 5\% \times \mu$. 8640 (30 days, 5 min-based) load vectors are generated, then Matpower solved these 8640 SCED problems and recorded 8640 LMP vectors. The 8640 load vectors and LMP vectors are the training and validation data for the 118 Bus system. These settings of load focus on a small but practically meaningful portion of the load space, and the calculation burden is significantly reduced.

5.3.3 Performance

The algorithm is implemented in Matlab, with the function *fitcsvm* and *fitSVMPosterior* in the Statistics and Machine Learning Toolbox. Table 5.2 summarizes the computation time of each step in the data-driven approach on a PC with Intel i7-2600 8-core CPU@3.40GHz and 16GB RAM memory. The classification and LMP forecast accuracy are summarized in Table 5.3 and Table 5.4, respectively.

5.4 Discussions on Posterior Probabilities

According to Table 5.1, 5.3 and 5.4, the methods "max-vote-wins" and "max-posterior-wins" have comparable performance in classification accuracy. This observation convinces us that the posterior probability calculation is correct. This is based

Table 5.2: Average Computation Time of the Data-driven Approach (average of 3 folds, in seconds)

Steps	**Computation Time**
(a) training	4.93s
(b) data post-processing	45.47s
(c1) max-vote-wins	184.79/2880 = 0.064s per forecast
(c2) max-posterior-wins	1033.78/2880 = 0.359s per forecast

Table 5.3: Classification Accuracy (118 Bus System)

Method	**Fold#1**	**Fold#2**	**Fold#3**	**Average**
Max-Vote-Wins	95.14%	94.31%	94.32%	94.59%
Max-Posterior-Wins	95.24%	93.99%	94.79%	94.67%

on the understanding that all the load vectors satisfying $P(y = i|P_D) = P(y = j|P_D)$ should be approximately on the optimal separating hyperplane between SPR#i and SPR#j, which is solved from Eqn. (4.3) and (4.4).

In the 118-bus system, the "max-posterior-wins" algorithm takes 45.47 seconds fitting the posterior probabilities, which have many potential benefits. One example of such benefits is about SPRs with high LMPs. In the 118-bus system, when load vectors fall into SPR#8, the LMPs rocket up from \$50/MWh to \$200/MWh due to the tripping of an expensive generator. In the validation dataset (2880 points), there are 74 load vectors associated with this SPR. Both the "max-vote-wins" and "max-posterior-wins" methods provide 61 correct classification results out of 74 vali-

Table 5.4: LMP Forecast Accuracy (118 Bus System)

Method	**Fold#1**	**Fold#2**	**Fold#3**	**Average**
Max-Vote-Wins	95.24%	99.15%	96.55%	96.98%
Max-Posterior-Wins	98.68%	98.70%	97.02%	98.13%

dation points. The remaining 13 points are false negative points, i.e. there is a price spike but we failed to predict it. However, the "max-posterior-wins" method provides posterior probabilities, which clearly show the confidence of our forecast. The posterior probabilities of one false negative point is as follows: $P(y = 8|P_D) = 0.42$, $P(y = 9|P_D) = 0.56$ and $P(y = \text{others}|P_D) = 0.02$. Although the classification result claims the load vector will not fall into the high price region (SPR#8), the possibility of having high prices is around 50%, the market participants should be cautious and take actions. Another example of the benefits of calculating posterior probabilities is stated in Table. 5.4, the LMP forecast could be improved with posterior probabilities.

Quantification of the posterior probabilities could yield many interesting applications. For example, load serving entities (LSEs) could consider demand response mechanisms to partially change the load vector and thus shift from high price SPRs. Market participants could also estimate the price volatilities due to renewables in a system.

6. CONCLUSIONS

In this paper, we examine the fundamental coupling between nodal load levels and LMPs in real-time SCED. It is shown that the load space can be partitioned into convex system pattern regions, which are one-to-one mapped with distinct LMP vectors. Based on the theoretical results, we propose a data-driven learning algorithm for market participants to identify SPRs. Identifying SPRs is modeled as a classification problem, and the proposed data-driven approach is built upon a "one-vs-one" multi-class SVM classifier. The proposed algorithm is shown to be capable of estimating SPRs solely from historical data without knowing confidential system information such as network topology and bidding curves. Given many uncertainties associated with SCED (e.g. transmission line capacities), we further quantify the probabilistic distributions of SPRs by use of posterior probabilities. The posterior probabilities could benefit market participants in various aspects such as improving LMP forecast and quantifying risks. Simulation results based on the IEEE 118-bus system demonstrates that the proposed algorithm is effective in understanding the past and predicting the future.

This paper is a first step towards developing theoretically rigorous and computationally feasible algorithms to analyzing the market prices as a result of varying loading levels. Future work should investigate the impact of unit commitment results on the system pattern regions. Another important avenue of research is to develop efficient learning algorithm to process a large amount of historical data in near real-time market operations.

REFERENCES

[1] Ilan Adler and Renato D. C. Monteiro. A geometric view of parametric linear programming. *Algorithmica*, 8(1-6):161–176, December 1992.

[2] Rui Bo and Fangxing Li. Probabilistic LMP forecasting considering load uncertainty. *IEEE Transactions on Power Systems*, 24(3):1279–1289, 2009.

[3] Rui Bo and Fangxing Li. Efficient Estimation of Critical Load Levels Using Variable Substitution Method. *IEEE Transactions on Power Systems*, 26(4):2472–2482, November 2011.

[4] F Borrelli, A Bemporad, and M Morari. Geometric algorithm for multiparametric linear programming. *Journal of Optimization Theory and Applications*, 118(3):515–540, 2003.

[5] AJ Conejo and Enrique Castillo. Locational marginal price sensitivities. *IEEE Transactions on Power Systems*, 20(4):2026–2033, 2005.

[6] C Filippi. On the geometry of optimal partition sets in multiparametric linear programming. 1997.

[7] T Gal and J Nedoma. Multiparametric linear programming. *Management Science*, 18(7):406–422, 1972.

[8] Tomas Gal. Postoptimal analyses, parametric programming and related topics. 1995.

[9] Xinbo Geng. *Understand LMP-Load Coupling from A Market Participant Perspective: Theory, Examples and An SVM-based Data-driven Approach*. Master thesis, Texas A&M University, College Station, 2015.

[10] Xinbo Geng and Le Xie. A Data-driven Approach to Identifying System Pattern Regions in Market Operations. In *IEEE Power and Energy Society General Meeting*, 2015.

[11] Trevor Hastie, Robert Tibshirani, and Others. Classification by pairwise coupling. *The annals of statistics*, 26(2):451–471, 1998.

[12] Martin Herceg, Michal Kvasnica, C Jones, and Manfred Morari. Multi-Parametric Toolbox 3.0. In *Proceedings of the European Control Conference*, pages 502–510, 2013.

[13] Yuting Ji, RJ Thomas, and Lang Tong. Probabilistic Forecast of Real-Time LMP via Multiparametric Programming. In *Proc. the 48th Annual Hawaii International Conference on System Sciences*, 2015.

[14] Yuting Ji, Lang Tong, Robert J Thomas, and Life Fellow. Probabilistic Forecast of Real-Time LMP and Network Congestion. *arXiv preprint arXiv:1503.06171*, 2015.

[15] Fangxing Li. Continuous locational marginal pricing (CLMP). *IEEE Transactions on Power Systems*, 22(4):1638–1646, 2007.

[16] Fangxing Li and Rui Bo. Congestion and price prediction under load variation. *IEEE Transactions on Power Systems*, 24(2):911–922, 2009.

[17] Illinois Institute of Technology. IEEE 118-bus, 54-unit, 24-hour system.

[18] J Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10:61–74, 1999.

[19] P M Subcommittee. IEEE reliability test system. *IEEE Transactions on Power Apparatus and Systems*, pages 2047–2054, 1979.

[20] Felix Wu, P Varaiya, P Spiller, and S Oren. Folk theorems on transmission access: Proofs and counterexamples. *Journal of Regulatory Economics*, 23:5–23, 1996.

[21] Qun Zhou, Leigh Tesfatsion, and CC Liu. Short-term congestion forecasting in wholesale power markets. *IEEE Transactions on Power Systems*, 26(4):2185–2196, 2011.

[22] Ray Daniel Zimmerman, Carlos Edmundo Murillo-Sánchez, and Robert John Thomas. MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education. *Power Systems, IEEE Transactions on*, 26(1):12–19, 2011.

APPENDIX A

ON THE ASSUMPTIONS

In Section 2, we made the following assumptions:

1. The generation cost is linear.

2. There is only one generator at each bus.

This section aims at illustrating these assumptions are reasonable.

A.1 From Piecewise Linear Cost to Linear Cost

In general, the generation cost of generators is *quadratic*. In current electricity market, the quadratic generation cost is approximated by a *piecewise linear* function. In our paper, the cost function is further simplified from *piecewise linear* to *linear*. The reason in details are summarized below:

For a generator with piecewise linear cost function $f(P_G)$:

$$f(P_G) = \begin{cases} f_1(P_G) = c_1^\intercal P_G & P_{G_0} \le P_G \le P_{G_1} \\ f_2(P_G) = c_2^\intercal P_G & P_{G_1} \le P_G \le P_{G_2} \\ \vdots & \vdots \\ f_n(P_G) = c_n^\intercal P_G & P_{G_{n-1}} \le P_G \le P_{G_n} \end{cases} \tag{A.1}$$

$f(P_G)$ has n linear segments $(f_i(P_G), i = 1, 2, \cdots, n)$.

In the formulation of SCED, it is obvious that a piecewise-linear-cost generator with n linear segments is equivalent with n generators, each one has the same cost and

upper/lower bound as one segment of the piecewise-linear-cost generator. Namely:

$$\begin{array}{lll} \text{Generator 1:} & f^{(1)}(P_G = c_1^{\intercal} P_G) & \text{and } P_{G_0} = P_G^{(1)\min} \leq P_G \leq P_G^{(1)\max} = P_{G_1} \\ \text{Generator 2:} & f^{(2)}(P_G = c_2^{\intercal} P_G) & \text{and } P_{G_1} = P_G^{(2)\min} \leq P_G \leq P_G^{(2)\max} = P_{G_2} \\ \vdots & \vdots & \vdots \\ \text{Generator n:} & f^{(n)}(P_G = c_n^{\intercal} P_G) & \text{and } P_{G_{n-1}} = P_G^{(n)\min} \leq P_G \leq P_G^{(n)\max} = P_{G_n} \end{array} \tag{A.2}$$

So we can only consider the scenario that there are many generators at one bus, but all of they have **linear** cost functions. Next section argues that we could further assume that there is only one generator at each bus.

A.2 From Multi Generators At A Bus to Only One Generator At A Bus

The **key idea** is: if there are several generators at one bus, there is an equivalent network with different shift factor matrix. In the new network, there is only one generator at each bus. We can analyze the equivalent network instead of the original network.

A 3-bus system example is provided (Figure. A.1). The line flow of the 3-bus

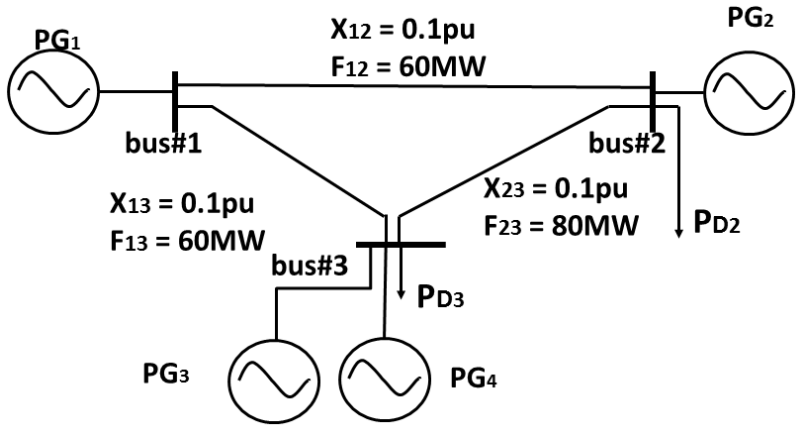


Figure A.1: 3-bus System (4 Generators)

system [1] can be expressed as:

$$F = \begin{bmatrix} F_{12} \\ F_{13} \\ F_{23} \end{bmatrix} = \begin{bmatrix} 0 & -2/3 & -1/3 \\ 0 & -1/3 & -2/3 \\ 0 & 1/3 & -1/3 \end{bmatrix} \begin{bmatrix} P_{G_1} \\ P_{G_2} \\ P_{G_3} + P_{G_4} \end{bmatrix} \tag{A.3}$$

$$= \begin{bmatrix} 0 & -2/3 & -1/3 \\ 0 & -1/3 & -2/3 \\ 0 & 1/3 & -1/3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} P_{G_1} \\ P_{G_2} \\ P_{G_3} \\ P_{G_4} \end{bmatrix} \tag{A.4}$$

We can further define a new shift factor matrix H_{new} as:

$$H_{new} = \begin{bmatrix} 0 & -2/3 & -1/3 \\ 0 & -1/3 & -2/3 \\ 0 & 1/3 & -1/3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \tag{A.5}$$

$$= \begin{bmatrix} 0 & -2/3 & -1/3 & -1/3 \\ 0 & -1/3 & -2/3 & -2/3 \\ 0 & 1/3 & -1/3 & -1/3 \end{bmatrix} \tag{A.6}$$

H_{new} is the shift factor matrix of a new network. This new network only has exactly one generator at one bus, and its cost function is linear.

[1] consider generations only, we can use superposition to consider loads later.

APPENDIX B

POSTERIOR PROBABILITY CALCULATION

First step: estimating posterior probabilities of every *binary* SVM classifier $P(y = i|P_D \text{ and } y \in \{i, j\})$. According to [18], training data is fitted to a sigmoid function (Eqn. (B.2)) by minimizing the negative log likelihood function (Eqn. (B.1)).

$$\min_{A,B} \quad -\sum_k t_k \log(r_k) + (1 - t_k) \log(1 - r_k) \tag{B.1}$$

$$\text{where} \quad r_k = P(y = i|P_D^{(k)} \text{ and } y \in \{i, j\})$$

$$= (1 + e^{AP_D^{(k)} + B})^{-1} \tag{B.2}$$

The parameter t_k associated with $P_D^{(k)}$ is calculated in Eqn. (B.3), where $P_D^{(k)}$ is the kth load vector in the training set, N_+ (N_-) is the number of positive (negative) examples.

$$t_k = \begin{cases} \frac{N_+ + 1}{N_+ + 2} & \text{if } y_k = +1 \\ \frac{1}{N_- + 2} & \text{if } y_k = -1 \end{cases} \tag{B.3}$$

Second step: estimating *multi-class* posterior probabilities $\hat{p_i} = P(y = i|P_D \text{ and } i \in \{1, 2, \cdots, n\})$ from the *binary* posterior probabilities, where n is the total number of classes. The algorithm proposed in [11] is widely being used. If the multi-class posterior probabilities $\hat{p_i}$s are correctly estimated, then the estimated binary posterior probabilities $\hat{\mu_{ij}} = \frac{\hat{p_i}}{\hat{p_i} + \hat{p_j}}$ should be identical to the observed binary posterior probabilities r_{ij}s by solving Eqn. (B.1) and (B.2). Therefore the objective of the algorithm in [11] is to minimize the Kullback-Leibler distance between $\hat{\mu_{ij}}$ and r_{ij}. Detailed algorithm is summarized below:

1. Start with the initial guess for the $\hat{p}_i$ and $\hat{\mu_{ij}} = \frac{\hat{p}_i}{\hat{p}_i + \hat{p}_j}$.

2. Repeat this $(i = 1, 2, \cdots, n, 1, 2, \cdots)$ until convergence:

$$\hat{p}_i \leftarrow \hat{p}_i \sum_{j \neq i} n_{ij} r_{ij} / \sum_{j \neq i} n_{ij} \mu_{ij} \tag{B.4}$$

   Then renormalize $\hat{p}_i \leftarrow \hat{p}_i / \sum_{j=1}^{n} \hat{p}_j$ and recompute $\hat{\mu_{ij}} = \frac{\hat{p}_i}{\hat{p}_i + \hat{p}_j}$.

3. If $\hat{\mathbf{p}} / \sum \hat{p}_i$ converges to the same $\hat{\mathbf{p}}$, then the algorithm stops, the vector $\hat{\mathbf{p}}$ will be the estimated multi-class posterior probabilities.

APPENDIX C

MORE RESULTS OF THE 3-BUS SYSTEM

C.1 System Pattern Regions in the 3D Space

In the 3-bus 2-generator system, we assume there are loads $(P_{D_1}, P_{D_2}, P_{D_3})$ at each bus. Therefore the system pattern regions (SPRs) are in the 3D space. We visualize the SPRs by monte-carlo simulation.

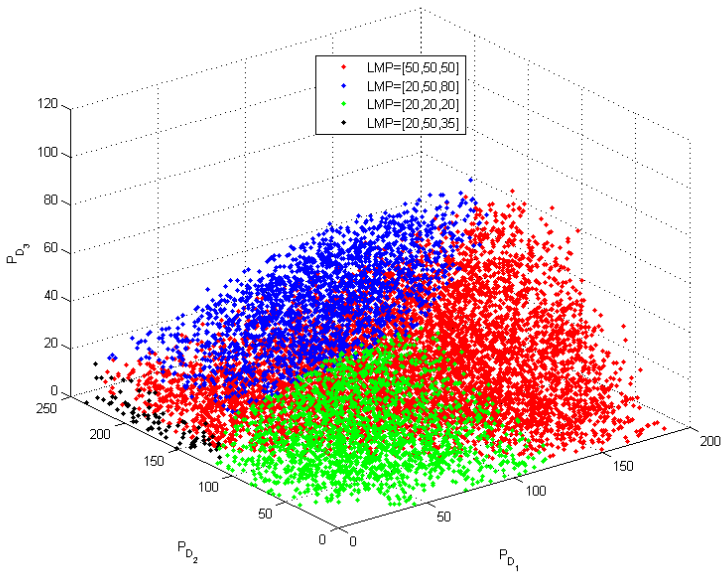


Figure C.1: 3D System Pattern Regions

C.2 Analytical Results of the 3-bus System

Table C.1 summarizes the details of SPRs in Fig. (3.3a).

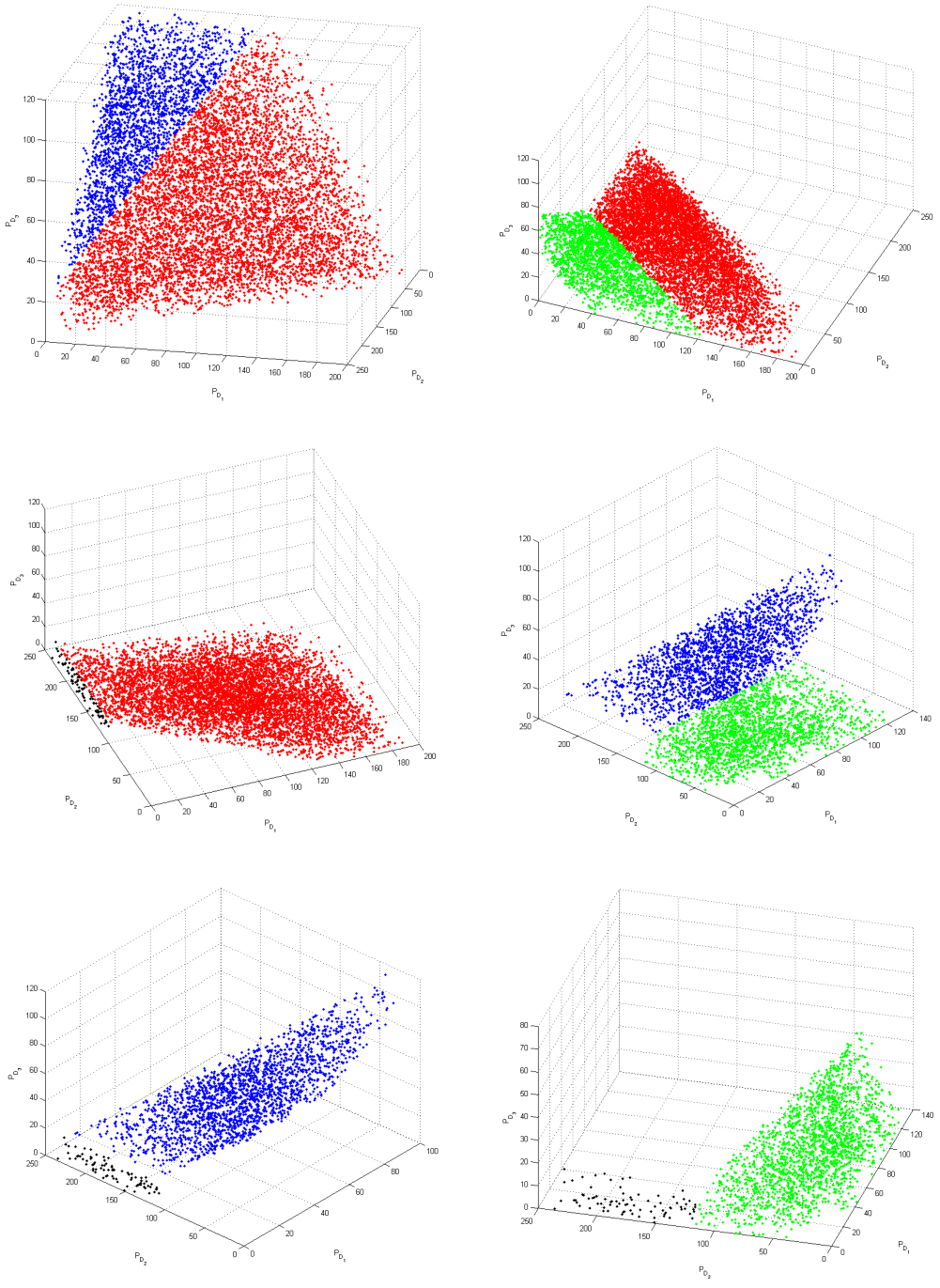


Figure C.2: Visualization of Any Two System Pattern Regions

Table C.1: Details of SPRs in Fig. (3.3a)

No.	SP	Analytical Form of The SPRs	LMPs
1	$\begin{bmatrix}1\\2\\3\\14\end{bmatrix}$	$\begin{bmatrix}1.0000 & 0\\0.7071 & -0.7071\\-0.7071 & 0.7071\\-1.0000 & 0\end{bmatrix}\begin{bmatrix}P_{D_2}\\P_{D_3}\end{bmatrix}\leq\begin{bmatrix}-100.0000\\-169.7056\\205.0610\\120.0000\end{bmatrix}$	$\begin{bmatrix}20\\50\\35\end{bmatrix}$
2	$\begin{bmatrix}1\\2\\9\\14\end{bmatrix}$	$\begin{bmatrix}0 & -1.0000\\0 & 1.0000\\0.7071 & 0.7071\\-0.7071 & -0.7071\end{bmatrix}\begin{bmatrix}P_{D_2}\\P_{D_3}\end{bmatrix}\leq\begin{bmatrix}-20.0000\\80.0000\\176.7767\\-70.7107\end{bmatrix}$	$\begin{bmatrix}50\\50\\50\end{bmatrix}$
3	$\begin{bmatrix}1\\2\\4\\14\end{bmatrix}$	$\begin{bmatrix}0 & 1.0000\\0 & -1.0000\\0.4472 & 0.8944\\-0.4472 & -0.8944\end{bmatrix}\begin{bmatrix}P_{D_2}\\P_{D_3}\end{bmatrix}\leq\begin{bmatrix}140.0000\\-80.0000\\147.5805\\-80.4984\end{bmatrix}$	$\begin{bmatrix}20\\50\\80\end{bmatrix}$
4	$\begin{bmatrix}1\\2\\4\\10\end{bmatrix}$	$\begin{bmatrix}-1.0000 & 0\\-0.4472 & -0.8944\\1.0000 & 0\\0.4472 & 0.8944\end{bmatrix}\begin{bmatrix}P_{D_2}\\P_{D_3}\end{bmatrix}\leq\begin{bmatrix}-50.0000\\-147.5805\\170.0000\\192.3018\end{bmatrix}$	$\begin{bmatrix}20\\60\\100\end{bmatrix}$
5	$\begin{bmatrix}1\\2\\13\\14\end{bmatrix}$	$\begin{bmatrix}0.8944 & 0.4472\\0.4472 & 0.8944\\-0.7071 & 0.7071\\0.7071 & -0.7071\\0.7071 & 0.7071\\-0.7071 & -0.7071\end{bmatrix}\begin{bmatrix}P_{D_2}\\P_{D_3}\end{bmatrix}\leq\begin{bmatrix}80.4984\\80.4984\\169.7056\\169.7056\\70.7107\\0\end{bmatrix}$	$\begin{bmatrix}20\\20\\20\end{bmatrix}$

6	$\begin{bmatrix}1\\2\\9\\10\end{bmatrix}$	$\begin{bmatrix}1.0000 & 0\\-1.0000 & 0\\-0.7071 & -0.7071\\0.7071 & 0.7071\end{bmatrix}\begin{bmatrix}P_{D_2}\\P_{D_3}\end{bmatrix} \leq \begin{bmatrix}230.0000\\-170.0000\\-176.7767\\212.1320\end{bmatrix}$		$\begin{bmatrix}100\\100\\100\end{bmatrix}$
7	$\begin{bmatrix}1\\2\\4\\5\end{bmatrix}$	$\begin{bmatrix}0 & -1\\1 & 0\\0 & 1\\-1 & 0\end{bmatrix}\begin{bmatrix}P_{D_2}\\P_{D_3}\end{bmatrix} \leq \begin{bmatrix}-140.0000\\50.0000\\190.0000\\100.0000\end{bmatrix}$		$\begin{bmatrix}20\\50\\100\end{bmatrix}$
8	$\begin{bmatrix}1\\2\\3\\10\end{bmatrix}$	$\begin{bmatrix}1.0000 & 0\\-0.8944 & -0.4472\\-1.0000 & 0\\0.8944 & 0.4472\end{bmatrix}\begin{bmatrix}P_{D_2}\\P_{D_3}\end{bmatrix} \leq \begin{bmatrix}290.0000\\-214.6625\\-230.0000\\237.0232\end{bmatrix}$		$\begin{bmatrix}20\\180\\100\end{bmatrix}$
9	$\begin{bmatrix}1\\2\\5\\13\end{bmatrix}$	$\begin{bmatrix}1.0000 & 0\\0.7071 & -0.7071\\-0.7071 & 0.7071\\-1.0000 & 0\end{bmatrix}\begin{bmatrix}P_{D_2}\\P_{D_3}\end{bmatrix} \leq \begin{bmatrix}-100.0000\\-169.7056\\205.0610\\120.0000\end{bmatrix}$		$\begin{bmatrix}20\\-60\\100\end{bmatrix}$
10	$\begin{bmatrix}1\\2\\8\\14\end{bmatrix}$	$\begin{bmatrix}0 & 1.0000\\0.7071 & -0.7071\\0 & -1.0000\\-0.7071 & 0.7071\end{bmatrix}\begin{bmatrix}P_{D_2}\\P_{D_3}\end{bmatrix} \leq \begin{bmatrix}-100.0000\\275.7716\\120.0000\\-169.7056\end{bmatrix}$		$\begin{bmatrix}20\\50\\-10\end{bmatrix}$

C.3 Simulation Results on the 3-bus System (Linear Separable Case)

The 3-bus system used in this section (Fig. C.3) is different from previous sections. There are only two generators in the system.

The training and testing data sets are separately generated from the Monte-Carlo simulation. There are 500 points used for training and 2000 points used for testing.

Optimal separating hyperplanes are visualized in Fig. C.4 $\sim$ Fig. C.6. The overall

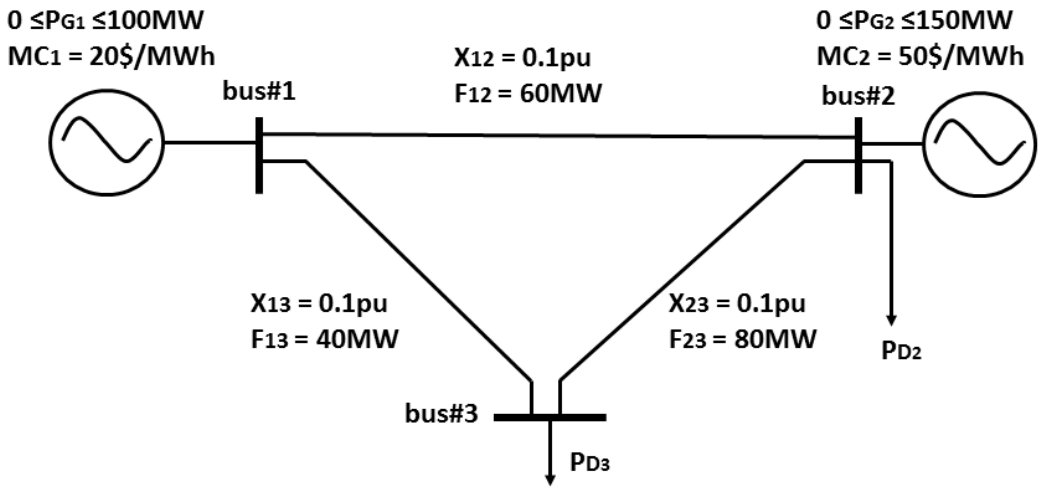


Figure C.3: 3 Bus System

classification accuracy is 99.1%, and the error points (0.9%) are plotted in Fig. C.7. All the error points locate beside the separating hyperplanes (straight lines in the two-dimensional space); there are no error points inside an SPR. Fig. C.7 also indicates there are some significant points, called support vectors, having huge influence on the separating hyperplanes. The support vectors are circled in Fig. C.4 $\sim$ Fig. C.6. In the separable case, the optimal hyperplane is solely determined by the support vectors, which means more support vectors could lead to a better identification of the SPRs. This feature could be utilized to improve the performance of the algorithm.

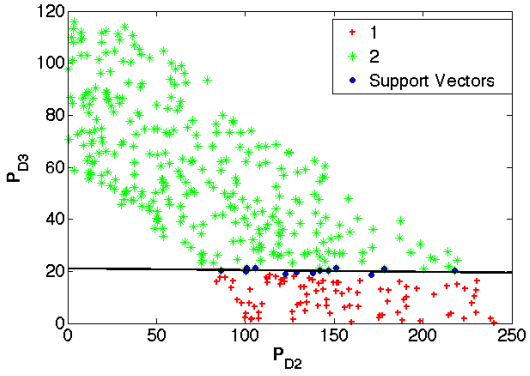


Figure C.4: SPR Class #1 and #2

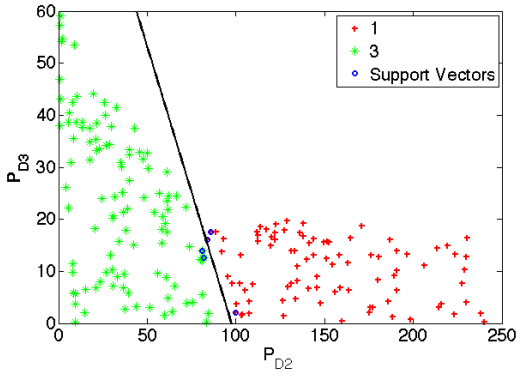


Figure C.5: SPR Class #1 and #3

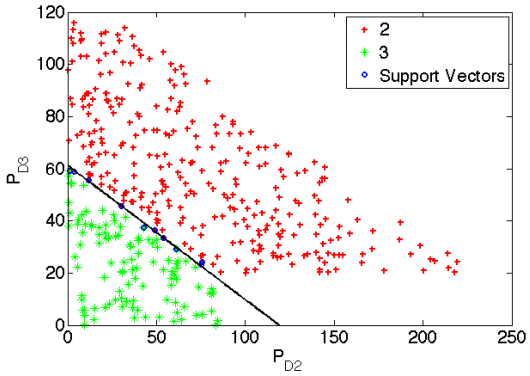


Figure C.6: SPR Class #2 and #3

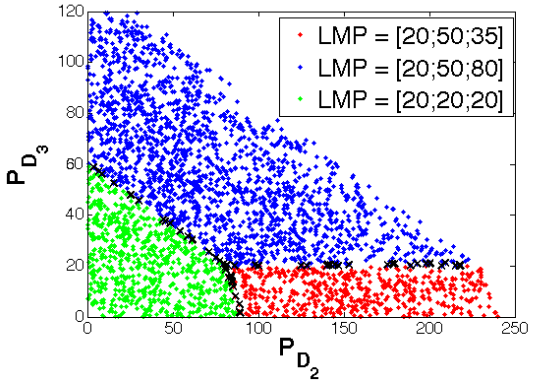


Figure C.7: Error Points

APPENDIX D

DETAILED SETTINGS OF THE IEEE 118-BUS SYSTEM

D.1 The Modified IEEE 118 Bus System

The lower bounds of generators are set to zero, but the upper bounds of generators remain the same as in [17]. Exact details of generation cost are summarized in the next section. To consider dynamic line ratings, we assume the line limits are random and satisfy normal distribution $N(\mu, \sigma)$. The expectation of line limits μ is the same as [17], only the limit of line 127 (from bus 81 to bus 80) is revised from 500MW to 100MW; the standard deviation σ is set to be 5% of the expectation, which means 95% of the time the line limits vary within 10% from the mean μ.

D.2 Settings of Generation Costs

Table D.1: Settings of Generation Costs

Index	Bus No.	$P_{\min}$	$P_{\max}$	cost c
1	4	5	30	26.2438
2	6	5	30	26.2438
3	8	5	30	26.2438
4	10	150	300	12.8875
5	12	100	300	12.8875
6	15	10	30	26.2438
7	18	25	100	17.8200
8	19	5	30	26.2438
9	24	5	30	26.2438

10	25	100	300	12.8875
11	26	100	350	10.76
12	27	8	30	26.24
13	31	8	30	26.24
14	32	25	100	17.82
15	34	8	30	26.24
16	36	25	100	17.82
17	40	8	30	26.24
18	42	8	30	26.24
19	46	25	100	17.82
20	49	50	250	12.32
21	54	50	250	12.32
22	55	25	100	17.82
23	56	25	100	17.82
24	59	50	200	13.29
25	61	50	200	13.29
26	62	25	100	17.82
27	65	100	420	8.339
28	66	100	420	8.339
29	69	80	300	12.88
30	70	30	80	15.47
31	72	10	30	26.2438
32	73	5	30	26.2438
33	74	5	20	37.6968
34	76	25	100	17.8200

35	77	25	100	17.8200
36	80	150	300	12.8875
37	82	25	100	17.8200
38	85	10	30	26.2438
39	87	100	300	10.7600
40	89	50	200	12.8875
41	90	8	20	37.6968
42	91	20	50	22.9423
43	92	100	300	12.8875
44	99	100	300	12.8875
45	100	100	300	12.8875
46	103	8	20	37.6968
47	104	25	100	17.8200
48	105	25	100	17.8200
49	107	8	20	37.6968
50	110	25	50	22.9423
51	111	25	100	17.8200
52	112	25	100	17.8200
53	113	25	100	17.8200
54	116	25	50	22.9423

APPENDIX E

SOME UNPUBLISHED RESULTS

E.1 The Impacts of Different Generation Costs on SPRs

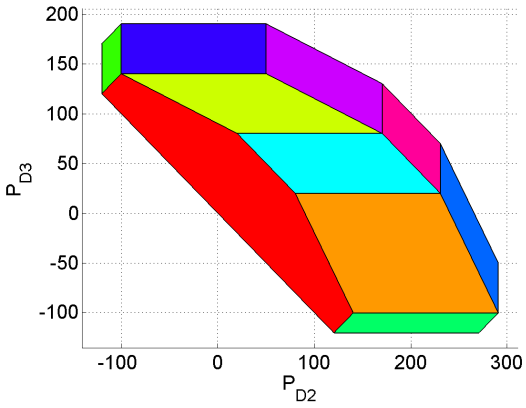
Remark 4. For a system pattern $\pi = (\mathcal{B}, \mathcal{N})$, its SPR would remain the same as long as the vector of costs c satisfies:

$$A_{\mathcal{B}}^{-1} c < 0 \tag{E.1}$$

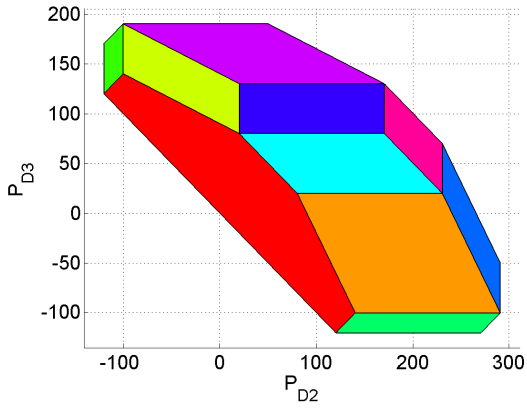
This is a direct conclusion from Eqn. (3.11).

The last column of Table C.1 provides the analytical results on the 3-bus system.

We could also visualize the SPRs given different generation cost vectors. Fig. (E.1a) shows the SPRs when $c = [20; 50; 81]$. When we reduce c_3 from 81 to 79, we get Fig. (E.1b). Part of the SPRs are different



(a) c= (20, 50, 81)



(b) c= (20, 50, 79)

E.2 All the SPRs Given Different Generation Costs

In this section, we explore all the possible SPRs **given various generation cost vectors**. The system topology, transmission and generation capacity are fixed. The results are summaried as following:

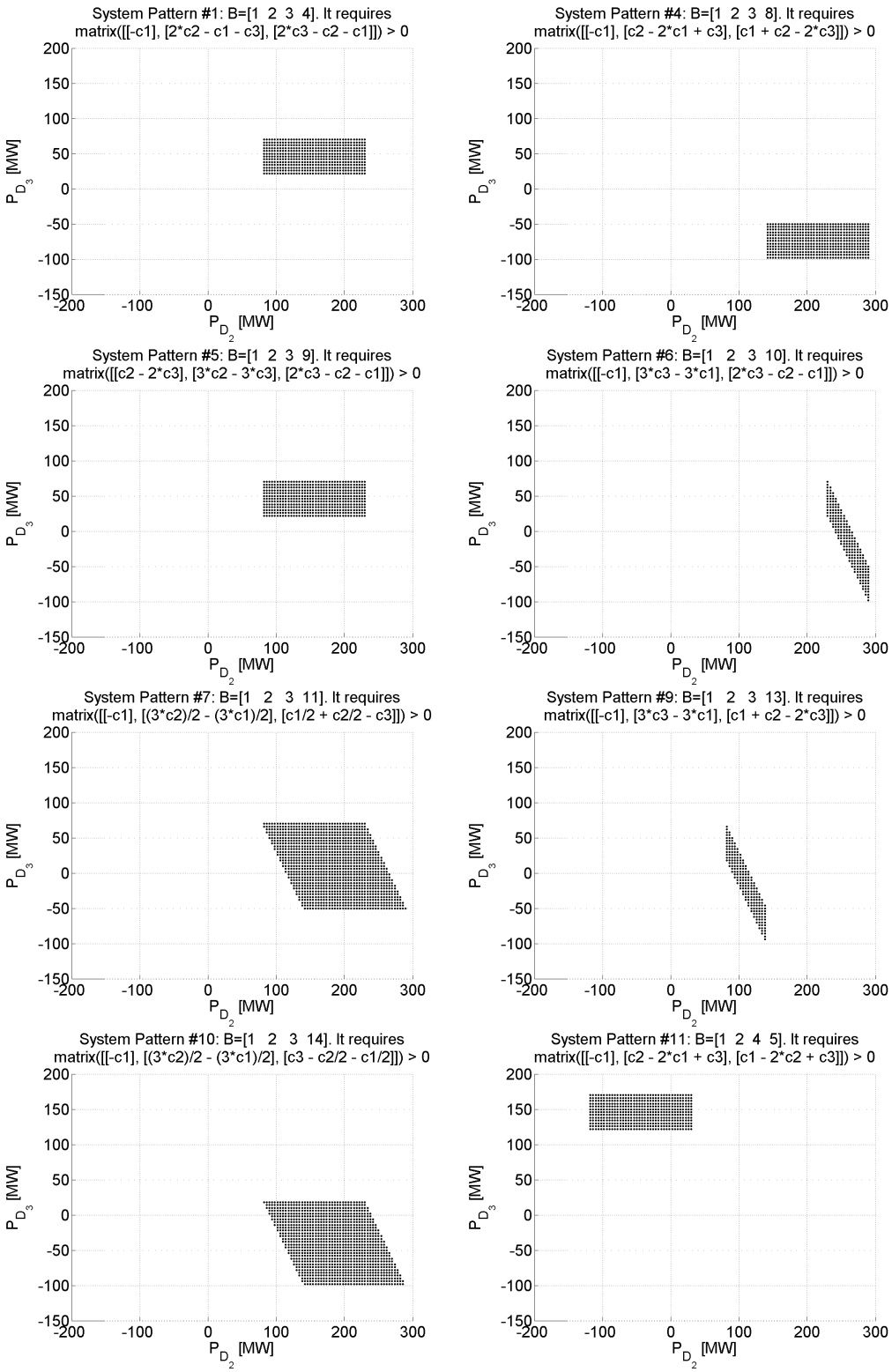


Figure E.1: All the Possible SPRs

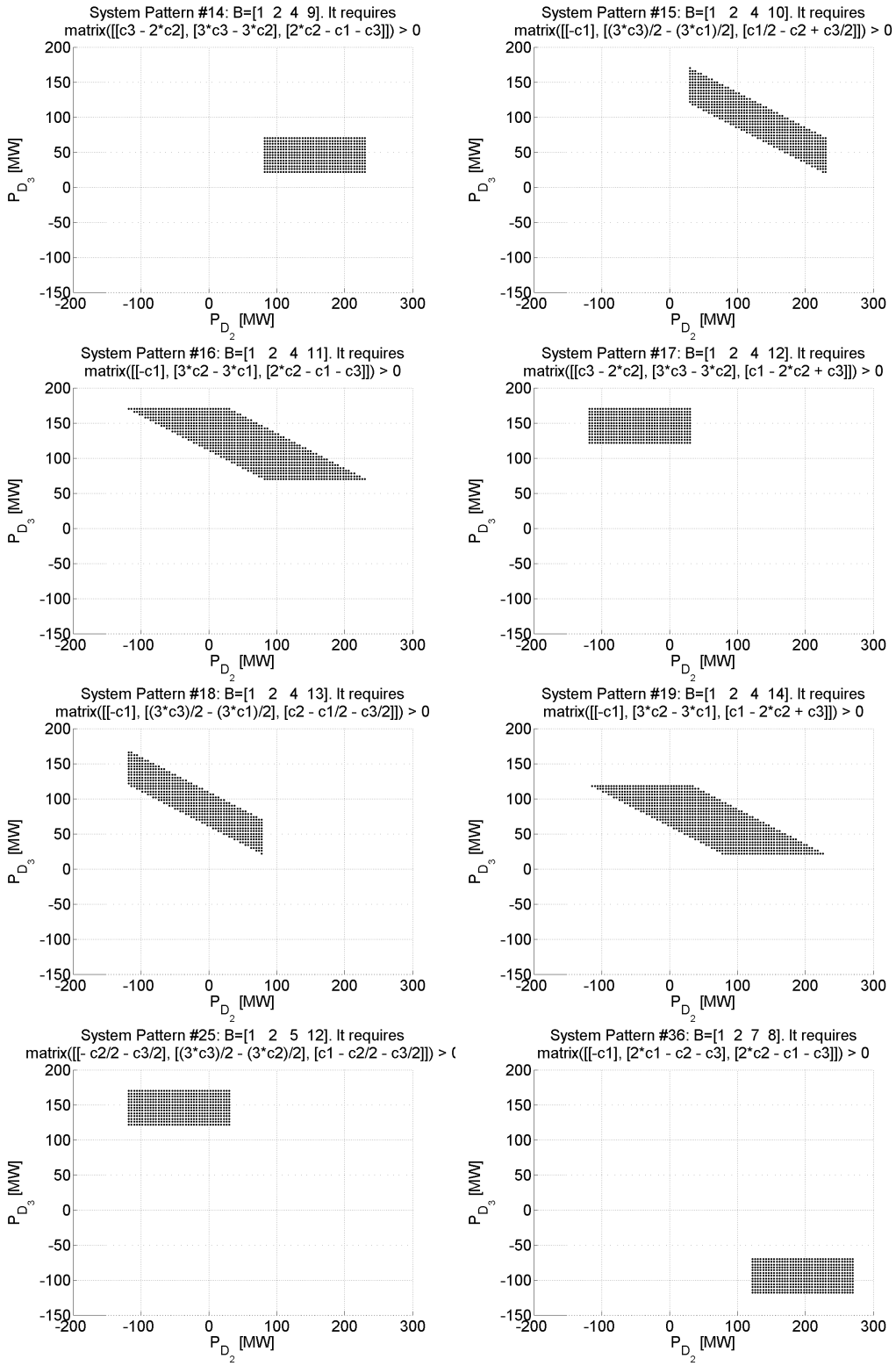


Figure E.1: All the Possible SPRs (Cont'd)

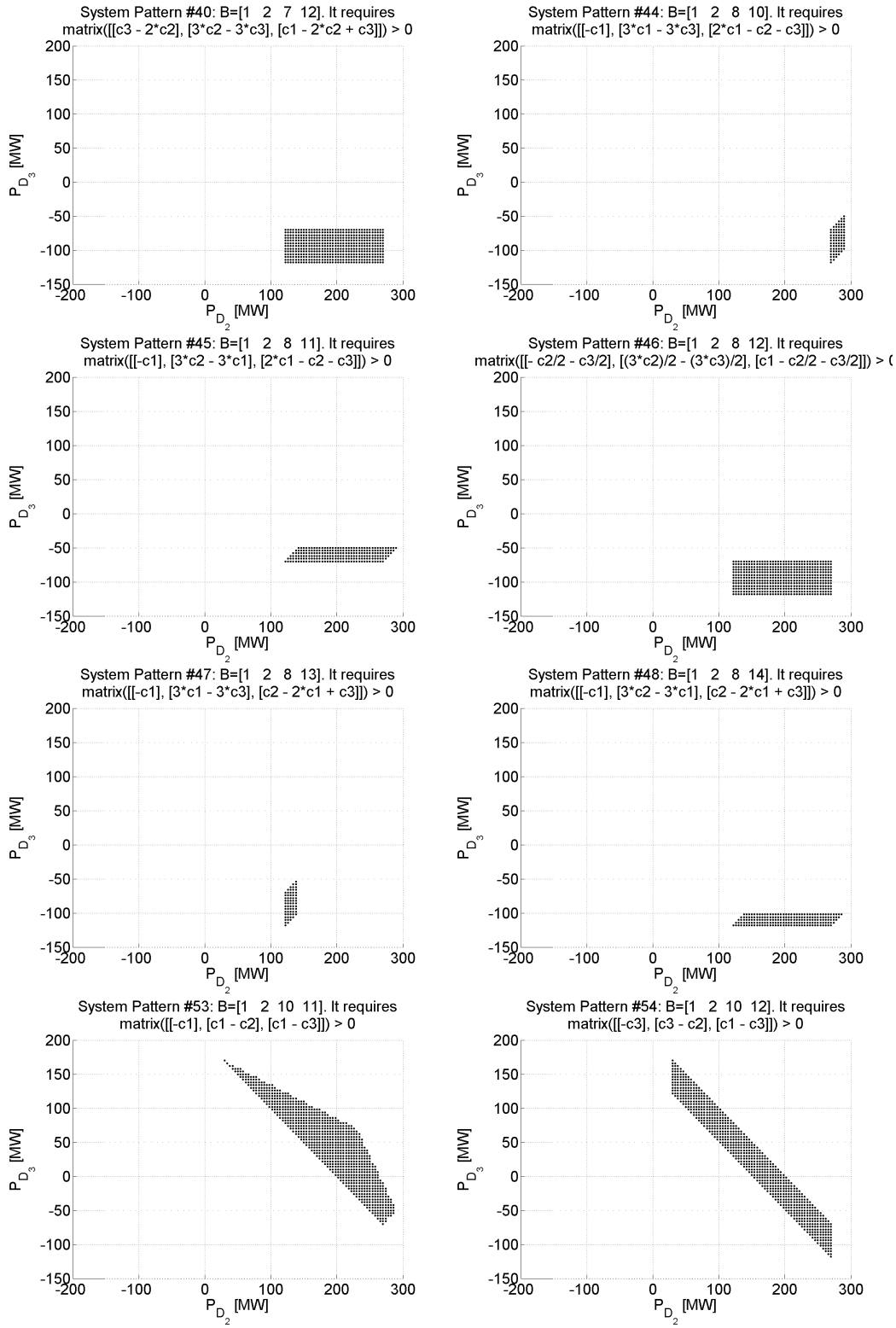


Figure E.1: All the Possible SPRs (Cont'd)

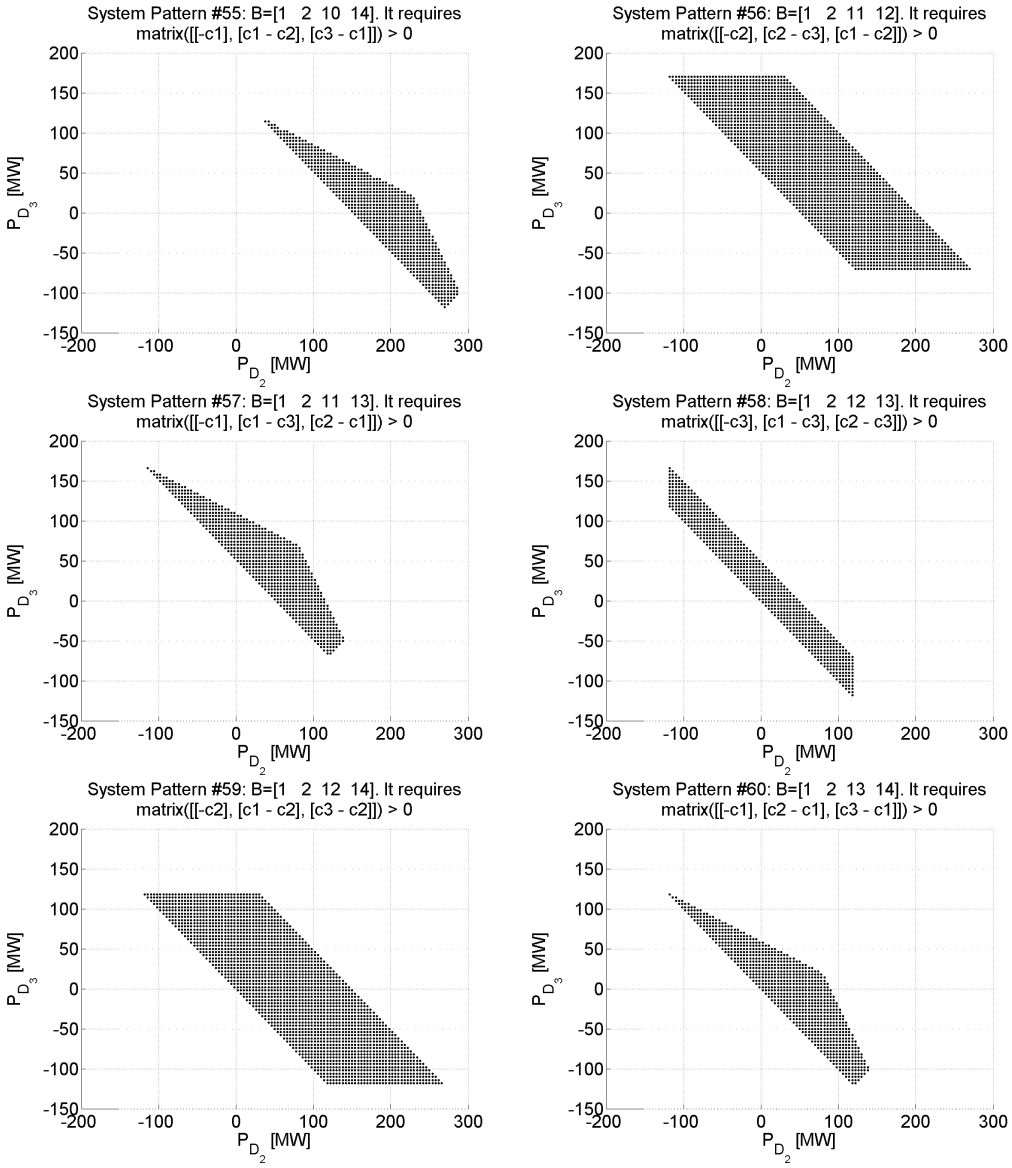


Figure E.1: All the Possible SPRs (Cont'd)

APPENDIX F

PROOF OF THEOREMS

F.1 Proof of the theorem "different SPRs have different LMP vectors"

We want to prove the theorem "different system pattern regions (SPRs) have different LMP vectors". This section will provide a detailed proof. It is worth noting that there are some special cases where different SPRs might have the same LMP vectors. In the following sections, we will describe and discuss those special cases in details. It is shown that the special cases exist only when some unusual conditions are satisfied, and could be further simplified.

F.1.1 Preparation

We first define *adjacent* sets.

Definition 3 (Adjacent Sets)**.** Given two closed set D_i and D_j, and $\dim(D_i) = \dim(D_j) = d \geq 2$. We say D_i and D_j are *adjacent* if $D_i \cap D_j \neq \emptyset$ and $\dim(D_i \cap D_j) = d - 1$.

An example of the adjacent sets is shown in Fig.F.1. In the 3-dimensional space, there are three sets (cuboids). The black and green cuboids are adjacent sets because their intersection has dimension $2 = 3 - 1$. The red and green cuboids are not adjacent, because their intersection belongs to a line, whose dimension is $1 \neq 3 - 1$.

Then we prove two lemmas.

Lemma 6 (Convex Piecewise Linear Functions With Parallel Segments)**.** *Assume the piecewise linear function $f : \mathbb{R}^n \to \mathbb{R}$ is composed of m linear functions $f_k = c_k^{\intercal} x$ where $k = 1, 2, \cdots, m$. Let $D = domf$, $D_k = dom\ f_k$[1], and assume D and D_i are*

[1]It is obvious that $D = \cup_k D_k$.

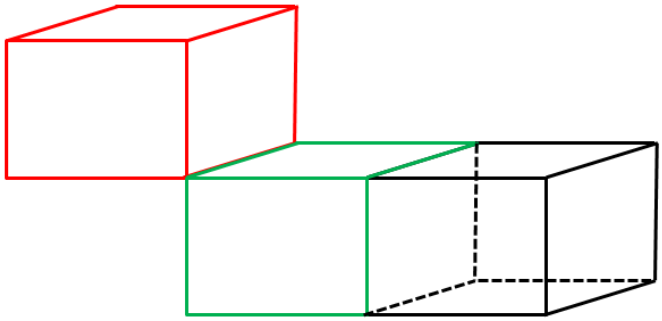


Figure F.1: Adjacent Sets (Example and Counter Example)

closed convex sets [2]. *If* f *is convex and has two parallel segments:* f_i *and* f_j $(i \neq j)$ *with* $\nabla^\intercal f_i = \nabla^\intercal f_j$, *then* f_i *and* f_j *have to be on the same hyperplane. Namely:*

1. $\forall x_i \in$ *relint* $D_i, \forall x_j \in$ *relint* D_j, $f(x_j) = f(x_i) + \nabla^\intercal f(x_i) \cdot (x_j - x_i)$.

2. $\forall x_i \in$ *relint* $D_i, \forall x_j \in$ *relint* D_j, *if the convex combination of* x_i *and* x_j *belongs to* D_k *instead of* D_i *or* D_j $(i \neq j \neq k)$, *then* $\nabla^\intercal f_k = \nabla^\intercal f_i = \nabla^\intercal f_j$ *and* $f(x_k) = f(x_i) + \nabla^\intercal f(x_i) \cdot (x_k - x_i)$.

Proof. f is differentiable in relint D_i.

We first prove (1). Since f is convex, according to the first order condition:

$$f(x_j) \geq f(x_i) + \nabla^\intercal f(x_i) \cdot (x_j - x_i) \tag{F.1}$$

Similarly,

$$f(x_i) \geq f(x_j) + \nabla^\intercal f(x_j) \cdot (x_i - x_j) \tag{F.2}$$

[2]The word "closed" indicates for $i \neq j$: $D_i \cap D_j \neq \emptyset$, but relint $D_i \cap$ relint $D_j = \emptyset$

Because $\nabla^\intercal f_i = \nabla^\intercal f_j$:

$$\begin{aligned} f(x_j) &\leq f(x_i) - \nabla^\intercal f(x_j) \cdot (x_i - x_j) = f(x_i) - \nabla^\intercal f(x_i) \cdot (x_i - x_j) && \text{(F.3)} \\ &= f(x_i) + \nabla^\intercal f(x_i) \cdot (x_j - x_i) && \text{(F.4)} \end{aligned}$$

So $f(x_i) + \nabla^\intercal f(x_i) \cdot (x_j - x_i) \geq f(x_j) \geq f(x_i) + \nabla^\intercal f(x_i) \cdot (x_j - x_i)$.

Therefore $f(x_j) = f(x_i) + \nabla^\intercal f(x_i) \cdot (x_j - x_i)$.

Then we prove (2) According to the convexity of the function f:

$$f(x) = f(\alpha x_i + (1-\alpha)x_j) \leq \alpha f(x_i) + (1-\alpha) f(x_j) \qquad \text{(F.5)}$$

We use the results in (1):

$$\begin{aligned} f(x) &\leq \alpha f(x_i) + (1-\alpha) f(x_j) && \text{(F.6)} \\ &= \alpha f(x_i) + (1-\alpha)[f(x_i) + \nabla^\intercal f(x_i)(x_j - x_i)] && \text{(F.7)} \\ &= f(x_i) + (1-\alpha)\nabla^\intercal f(x_i)(x_j - x_i) && \text{(F.8)} \\ &= f(x_i) + \nabla^\intercal f(x_i)[\alpha x_i + (1-\alpha)x_j - x_i] && \text{(F.9)} \\ &= f(x_i) + \nabla^\intercal f(x_i)(x - x_i) && \text{(F.10)} \end{aligned}$$

Also, according to the first order condition, we get (for x and x_i):

$$f(x) \geq f(x_i) + \nabla^\intercal f(x_i)(x - x_i) \qquad \text{(F.11)}$$

We combine Eqn. (F.10) and Eqn. (F.5):

$$f(x_i) + \nabla^\intercal f(x_i)(x - x_i) \geq f(x) \geq f(x_i) + \nabla^\intercal f(x_i)(x - x_i) \qquad \text{(F.12)}$$

So

$$f(x) = f(x_i) + \nabla^\intercal f(x_i)(x - x_i) \tag{F.13}$$

Therefore

$$\nabla^\intercal f(x) = \nabla^\intercal f(x_i) = \nabla^\intercal f(x_j) \tag{F.14}$$

□

Lemma 7 (System Patterns of Adjacent SPRs)**.** *Given two system pattern regions (SPRs) S_i and S_j and their system patterns $\pi_i = (\mathcal{B}_i, \mathcal{N}_i)$ and $\pi_j = (\mathcal{B}_j, \mathcal{N}_j)$. If S_i and S_j are* adjacent[3]*, then $\mathcal{B}_i$ and $\mathcal{B}_j$ only differ in one entry.*

Proof. Lemma 7 is actually a direct corollary from Definition 3.

If S_i and S_j are *adjacent* but $\mathcal{B}_i$ and $\mathcal{B}_j$ differ in $k \geq 2$ entries. Then $S_i \cap S_j$ is depicted by k linear constraints. If the constraints are linear independent[4], then $\dim S_i \cap S_j = \dim S_i - k < \dim S_i - 1$. This is contradictory with the definition of *adjacent* sets, which requires $\dim S_i \cap S_j = \dim S_i - 1$. □

F.1.2 The LMP vectors of SPRs

F.1.2.1 KKT Conditions

Consider the Security-constrained Economic Dispatch (SCED) problem in the form of Eqn. (2.1). Its Lagrangian $L : \mathbb{R}^{n_b} \times \mathbb{R} \times \mathbb{R}^{n_l} \times \mathbb{R}^{n_l} \times \mathbb{R}^{n_g} \times \mathbb{R}^{n_g} \to \mathbb{R}$ is defined as:

$$\begin{aligned} L(P_G, \lambda_1, \mu_+, \mu_-, \eta_+, \eta_-) &= c^\intercal P_G + \lambda_1(\mathbf{1}^\intercal P_G - \mathbf{1}^\intercal P_D) \\ &+ \mu_+^\intercal (H P_G - H P_D - F_+) - \mu_-^\intercal (H P_G - H P_D - F_-) \\ &+ \eta_+^\intercal (P_G - G_+) - \eta_-^\intercal (P_G - G_-) \end{aligned} \tag{F.15}$$

[3]See Definition 3.

[4]If they are linear dependent, we can always eliminate the redundant constraints, which will not make any difference.

According to KKT conditions:

$$\begin{aligned}
\mu_+, \mu_-, \eta_+, \eta_- &\geq 0 && \text{(F.16)}\\
\mu_{+,i}^\intercal (H_i P_G - H_i P_D - F_{+,i}) &= 0, i = 1, 2, \cdots, n_l && \text{(F.17)}\\
\mu_{-,i}^\intercal (H_i P_G - H_i P_D - F_{-,i}) &= 0, i = 1, 2, \cdots, n_l && \text{(F.18)}\\
\eta_{+,i}^\intercal (P_{G,i} - G_{+,i}) &= 0, i = 1, 2, \cdots, n_g && \text{(F.19)}\\
\eta_{-,i}^\intercal (P_{G,i} - G_{-,i}) &= 0, i = 1, 2, \cdots, n_g && \text{(F.20)}\\
\nabla_{P_G}^\intercal L &= 0 && \text{(F.21)}
\end{aligned}$$

Where $\mu_{+,i}$ $(\mu_{-,i}, \eta_{+,i}, \eta_{-,i})$ is the $i$the entry of the vector μ_+ (μ_-, η_+, η_-). H_i is the ith row of the shift factor matrix H.

And:

$$\nabla_{P_G}^\intercal L = c + \lambda_1 \mathbf{1} + H^\intercal(\mu_+ - \mu_-) + \eta_+ - \eta_- = c + \lambda_1 \mathbf{1} + H^\intercal \mu + \eta = 0 \qquad \text{(F.22)}$$

where $\mu = \mu_+ - \mu_-$ and $\eta = \eta_+ - \eta_-$.

Also we calculate the LMP vector λ:

$$\lambda = -\nabla_{P_D}^\intercal L = \lambda_1 \mathbf{1} + H^\intercal \mu \qquad \text{(F.23)}$$

This is consistent with [20].

F.1.2.2 Two System Pattern Regions Having The Same LMP Vector

Assume there exist two SPRs (i, j) which have the same LMP vector $\lambda^{(i)} = \lambda^{(j)}$.

Notice that this equality $\lambda^{(i)} = \lambda^{(j)}$ is true for each entry. For the slack bus (in our assumption, bus #1), the LMPs are the same: $\lambda_1^{(i)} = \lambda_1^{(j)}$. In other words, the *energy components* of the LMP vectors are the same. According to $H^\intercal \mu = \lambda - \lambda_1 \mathbf{1}$,

the *congestion components* are also the same:

$$H^\intercal \mu^{(i)} = H^\intercal \mu^{(j)} \tag{F.24}$$

According to Eqn. (F.22):

$$\begin{aligned} \eta^{(i)} &= -c - (\lambda_1^{(i)}\mathbf{1} + H^\intercal \mu^{(i)}) = -c - \lambda^{(i)} && \text{(F.25)} \\ &= -c - \lambda^{(j)} = -c - (\lambda_1^{(j)}\mathbf{1} + H^\intercal \mu^{(j)}) = \eta^{(j)} && \text{(F.26)} \end{aligned}$$

$\eta^{(i)} = \eta^{(j)}$ means the marginal generators (which are ON) of the two SPRs (i, j) are also exactly the same.

Also, according to Lemma 6, there has to be two *adjacent* SPRs (say i and k[5]) which have the same LMP vectors. And the system pattern of adjacent SPRs only differ in one entry. This indicates that there is only one different binding constraints (i.e. $\mathcal{B}_i$), either one different congested line or one different marginal generator.

According to the analysis above, there are two possibilities that lead to different SPRs having the same LMP vectors:

1. The congestion pattern (congested lines) are the same: $\mu^{(i)} = \mu^{(k)}$. But the same Lagrange multipliers $\eta^{(i)} = \eta^{(k)}$ represent exactly one different generator status.

2. The generator statuses are exactly the same, the congested lines are different $\mu^{(i)} \neq \mu^{(k)}$ but $H^\intercal(\mu^{(i)} - \mu^{(k)}) = 0$. And there is only one different congested line form SPR i and k.

We will discuss these two possibilities/cases in the following sections, and we will show:

[5] $k \neq j$ is possible.

- **the first case is possible only when some bids from generators (or generation costs) are the same.**
- **the second case is impossible.**

F.1.2.3 Case 1: Congestion Lines are the same, but there is one different marginal generator.

Eqn. (F.25): $\eta^{(i)} = -c - \lambda^{(i)}$. Since $\lambda^{(i)} = \lambda^{(k)}$, then $\eta^{(i)} = \eta^{(k)}$.

This indicates: the same vector $\eta = \eta^{(i)} = \eta^{(k)}$ represents two different sets of marginal generators. This is possible only when there are some "equivalent" generators, namely the generators with the same bids (or generation costs). In this case, we call those generators hits the maximum (minimum) capacity constraint as "max (min) generators". Since their generation costs are the same, increasing the output of one min generator but decreasing the same amount of another one will not change the total system generation cost. It is worth noting that this would not change the congestion pattern, otherwise the η vector would be different due to the cost of congestions. For all those "equivalent" generators, their Lagrange multipliers are the same (generation cost). This means they could be adjusted and therefore not hit either upper or lower bounds. Their generation capacity constraints are not binding.

The system pattern region (SPR) relates with this case is still convex. The reason is as follow: the SPR is constrained by several hyperplanes (linear constraints in the load space) regardless how many inequalities there are in the problem. The halfspace is convex, and the intersection of halfspaces is also convex. Therefore the SPR is still convex.

We could therefore argue that this case could be further reduced by regarding the "equivalent generators" as one generator but with larger capacity.

F.1.2.4 Case 2: Marginal Generators are the same, but there is one different congested line.

We will show this case is not possible. The proof is as follows[6]:

There is one different congested line between SPR i and SPRj. Without loss of generality[7], we assume line 1 is congested in SPR i but not congested in SPR j. Similarly, line 2 is congested in SPR j but not congested in SPR i. And the index set of the lines congested in both SPR i and SPRj is denoted by $\mathcal{C}$.

In our previous settings, the matrix $A_{\mathcal{B}}$, which relates with binding constraints, has the structure as follows:

$$A_{\mathcal{B}} = \begin{bmatrix} \text{supply-demand balance constraint: } \mathbf{1}_{n_b}^{\intercal} \\ \text{shift factor matrix related with congested lines} \\ \text{matrix related with generation upper/lower bounds} \end{bmatrix} \tag{F.27}$$

We rearrange the structure of matrix $A_{\mathcal{B}}$ as follows[8]:

$$A_{\mathcal{B}} = \begin{bmatrix} \text{supply-demand balance constraint: } \mathbf{1}_{n_b}^{\intercal} \\ \text{matrix related with generation upper/lower bounds: } G \\ \text{shift factor matrix related with commonly congested lines: } H_{\mathcal{C}} \\ \text{shift factor matrix related with uniquely congested lines: } H_U \end{bmatrix} \tag{F.28}$$

[6]This proof is so verbose that we cannot put it in the draft of our journal paper, we hope to simplify it in future works.

[7]We can label the congested lines with any non-repetitive numbers.

[8]This step will not make any difference to the theoretical results, but will significantly simplify the notations.

For SPR i and SPR k:

$$A_{\mathcal{B}_i} = \begin{bmatrix} \mathbf{1}_{n_b}^\intercal \\ G \\ H_{\mathcal{C}} \\ h_1 \end{bmatrix} = \begin{bmatrix} E \\ H_{\mathcal{C}} \\ h_1 \end{bmatrix}, A_{\mathcal{B}_k} = \begin{bmatrix} \mathbf{1}_{n_b}^\intercal \\ G \\ H_{\mathcal{C}} \\ h_2 \end{bmatrix} \begin{bmatrix} E \\ H_{\mathcal{C}} \\ h_2 \end{bmatrix} \tag{F.29}$$

where $[\mathbf{1}_{n_b}^\intercal; G]$ is common for both SPR i and SPR k, we use matrix $E = [\mathbf{1}_{n_b}^\intercal; G]$ to represent it. $H_{\mathcal{C}}$ is the shift factor matrix related with lines congested in both SPR i and SPR k. h_1 is the row of line 1 in the shift factor matrix H, h_2 is the row of line 2 in the shift factor matrix H. $h_1, h_2 \in \mathbb{R}^{1\times n_b}$ are row vectors, where n_b is the number of buses.

Let $\Lambda_{\mathcal{B}_i} = A_{\mathcal{B}_i}^{-1}$ and $\Lambda_{\mathcal{B}_k} = A_{\mathcal{B}_k}^{-1}$, and the structure of $\Lambda_{\mathcal{B}_i}$ and $\Lambda_{\mathcal{B}_k}$ is as follows:

$$\Lambda_{\mathcal{B}_i} = A_{\mathcal{B}_i}^{-1} = \begin{bmatrix} B_1^{(i)} & B_2^{(i)} & \beta_3^{(i)} \end{bmatrix}, \Lambda_{\mathcal{B}_k} = A_{\mathcal{B}_k}^{-1} = \begin{bmatrix} B_1^{(k)} & B_2^{(k)} & \beta_3^{(k)} \end{bmatrix} \tag{F.30}$$

where $\beta_3^{(k)}, \beta_3^{(i)} \in \mathbb{R}^{1\times n_b}$

Since $A_{\mathcal{B}_i}\Lambda_{\mathcal{B}_i} = \mathbf{I}$ and $A_{\mathcal{B}_k}\Lambda_{\mathcal{B}_k} = \mathbf{I}$:

$$A_{\mathcal{B}_i} \times A_{\mathcal{B}_i}^{-1} = \begin{bmatrix} EB_1^{(i)} & EB_2^{(i)} & E\beta_3^{(i)} \\ H_{\mathcal{C}}B_1^{(i)} & H_{\mathcal{C}}B_2^{(i)} & H_{\mathcal{C}}\beta_3^{(i)} \\ h_1B_1^{(i)} & h_1B_2^{(i)} & h_1\beta_3^{(i)} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1 \end{bmatrix} \tag{F.31}$$

Similarly,

$$A_{\mathcal{B}_k} \times A_{\mathcal{B}_k}^{-1} = \begin{bmatrix} EB_1^{(k)} & EB_2^{(k)} & E\beta_3^{(k)} \\ H_{\mathcal{C}}B_1^{(k)} & H_{\mathcal{C}}B_2^{(k)} & H_{\mathcal{C}}\beta_3^{(k)} \\ h_2B_1^{(k)} & h_2B_2^{(k)} & h_2\beta_3^{(k)} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1 \end{bmatrix} \tag{F.32}$$

It is worth emphasizing that although $A_{\mathcal{B}_i}$ and $A_{\mathcal{B}_k}$ have common submatrix $[E; H_{\mathcal{C}}]$,

in the inverse matrices $\Lambda_{\mathcal{B}_i}$ and $\Lambda_{\mathcal{B}_k}$:

$$\begin{bmatrix} EB_1^{(i)} & EB_2^{(i)} \\ H_{\mathcal{C}}B_1^{(i)} & H_{\mathcal{C}}B_2^{(i)} \end{bmatrix} \neq \begin{bmatrix} EB_1^{(k)} & EB_2^{(k)} \\ H_{\mathcal{C}}B_1^{(k)} & H_{\mathcal{C}}B_2^{(k)} \end{bmatrix} \tag{F.33}$$

An interesting observation:

$$\begin{aligned} (A_{\mathcal{B}_k}^{\intercal})^{-1}A_{\mathcal{B}_i}^{\intercal} = (A_{\mathcal{B}_k}^{-1})^{\intercal}A_{\mathcal{B}_i}^{\intercal} &= (A_{\mathcal{B}_i}A_{\mathcal{B}_k}^{-1})^{\intercal} \begin{bmatrix} EB_1^{(k)} & EB_2^{(k)} & E\beta_3^{(k)} \\ H_{\mathcal{C}}B_1^{(k)} & H_{\mathcal{C}}B_2^{(k)} & H_{\mathcal{C}}\beta_3^{(k)} \\ h_1B_1^{(k)} & h_1B_2^{(k)} & h_1\beta_3^{(k)} \end{bmatrix}^{\intercal} \\ &= \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ h_1B_1^{(k)} & h_1B_2^{(k)} & h_1\beta_3^{(k)} \end{bmatrix}^{\intercal} \end{aligned} \tag{F.34}$$

Multiply $y_{\mathcal{B}_i}$ on both sides:

$$(A_{\mathcal{B}_k}^{\intercal})^{-1}A_{\mathcal{B}_i}^{\intercal} \times y_{\mathcal{B}_i} = ((A_{\mathcal{B}_k}^{\intercal})^{-1}A_{\mathcal{B}_i}^{\intercal}) \times y_{\mathcal{B}_i} \tag{F.35}$$

$$= \begin{bmatrix} \mathbf{I} & \mathbf{0} & (B_1^{(k)})^{\intercal}h_1^{\intercal} \\ \mathbf{0} & \mathbf{I} & (B_2^{(k)})^{\intercal}h_1^{\intercal} \\ \mathbf{0} & \mathbf{0} & h_1\beta_3^{(k)} \end{bmatrix} \begin{bmatrix} \alpha \\ \mu_{\mathcal{C}}^{(i)} \\ \mu_1 \end{bmatrix} \tag{F.36}$$

$$= \begin{bmatrix} \alpha \\ \mu_{\mathcal{C}}^{(i)} \\ \mathbf{0} \end{bmatrix} + \mu_1 \begin{bmatrix} (B_1^{(k)})^{\intercal}h_1^{\intercal} \\ (B_2^{(k)})^{\intercal}h_1^{\intercal} \\ h_1\beta_3^{(k)} \end{bmatrix} \tag{F.37}$$

Also:

$$(A^{\intercal}_{\mathcal{B}_k})^{-1}A^{\intercal}_{\mathcal{B}_i} \times y_{\mathcal{B}_i} = (A^{\intercal}_{\mathcal{B}_k})^{-1}(A^{\intercal}_{\mathcal{B}_i} \times y_{\mathcal{B}_i}) = (A^{\intercal}_{\mathcal{B}_k})^{-1} \times (-c) = y_{\mathcal{B}_k} = \begin{bmatrix} \alpha \\ \mu^{(k)}_{\mathcal{C}} \\ \mu_2 \end{bmatrix} \quad \text{(F.38)}$$

Therefore:

$$\begin{bmatrix} \alpha \\ \mu^{(k)}_{\mathcal{C}} \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \alpha \\ \mu^{(i)}_{\mathcal{C}} \\ \mathbf{0} \end{bmatrix} + \mu_1 \begin{bmatrix} (B^{(k)}_1)^{\intercal}h^{\intercal}_1 \\ (B^{(k)}_2)^{\intercal}h^{\intercal}_1 \\ h_1\beta^{(k)}_3 \end{bmatrix} \quad \text{(F.39)}$$

$$\begin{bmatrix} \mathbf{0} \\ \mu^{(k)}_{\mathcal{C}} - \mu^{(i)}_{\mathcal{C}} \\ \mu_2 \end{bmatrix} = \mu_1 \begin{bmatrix} (B^{(k)}_1)^{\intercal}h^{\intercal}_1 \\ (B^{(k)}_2)^{\intercal}h^{\intercal}_1 \\ h_1\beta^{(k)}_3 \end{bmatrix} \quad \text{(F.40)}$$

We get the following equations:

$$\mu^{(k)}_{\mathcal{C}} - \mu^{(i)}_{\mathcal{C}} = \mu_1(B^{(k)}_2)^{\intercal}h^{\intercal}_1 \quad \text{(F.41)}$$

$$\mu_2 = \mu_1 h_1 \beta^{(k)}_3 \quad \text{(F.42)}$$

From the assumption $\lambda^{(i)} = \lambda^{(k)}$ we get $H^{\intercal}(\mu^{(i)} - \mu^{(k)}) = \mathbf{0}$. Since the shadow prices (multipliers) of the non-congested lines are zero, we get:

$$\mathbf{0} = H^{\intercal}(\mu^{(i)} - \mu^{(k)}) = H^{\intercal}_{\mathcal{C}}(\mu^{(i)}_{\mathcal{C}} - \mu^{(k)}_{\mathcal{C}}) + \mu_1 h^{\intercal}_1 - \mu_2 h^{\intercal}_2 \quad \text{(F.43)}$$

From Eqn.(F.41), we get

$$\mathbf{0} = H^{\intercal}_{\mathcal{C}}(\mu^{(i)}_{\mathcal{C}} - \mu^{(k)}_{\mathcal{C}}) + \mu_1 h^{\intercal}_1 - \mu_2 h^{\intercal}_2 = -\mu_1 H^{\intercal}_{\mathcal{C}}(B^{(k)}_2)^{\intercal}h^{\intercal}_1 + \mu_1 h^{\intercal}_1 - \mu_2 h^{\intercal}_2 \quad \text{(F.44)}$$

From Eqn.(F.32), we get $H_{\mathcal{C}}^{\intercal}(B_2^{(k)})^{\intercal} = \mathbf{I}$, therefore Eqn.(F.44) becomes:

$$\mathbf{0} = -\mu_1 H_{\mathcal{C}}^{\intercal}(B_2^{(k)})^{\intercal} h_1^{\intercal} + \mu_1 h_1^{\intercal} - \mu_2 h_2^{\intercal} = -\mu_1 h_1^{\intercal} + \mu_1 h_1^{\intercal} - \mu_2 h_2^{\intercal} = -\mu_2 h_2^{\intercal} \qquad \text{(F.45)}$$

So $\mu_2 h_2^{\intercal} = \mathbf{0}$.

Since line 2 is congested, $\mu > 0$. The only possibility is $h_2^{\intercal} = \mathbf{0}$. This is contradictory with the physical meaning of h_2, which is the row in the shift factor matrix corresponding to line 2[9].

[9]If $h_2^{\intercal} = \mathbf{0}$, this simply means this line does not exist, which is impossible.

APPENDIX G

RESULTS OF THE IEEE 24-BUS SYSTEM

The IEEE 24-bus system [19] is shown in Fig. G.1). We pay special attention to the tradeoff between classification accuracy and computational complexity. The SVM algorithm from Matlab's Statistics toolbox is used, with the linear kernel and Sequential Minimal Optimization (SMO) method to find the optimal separating hyperplanes.

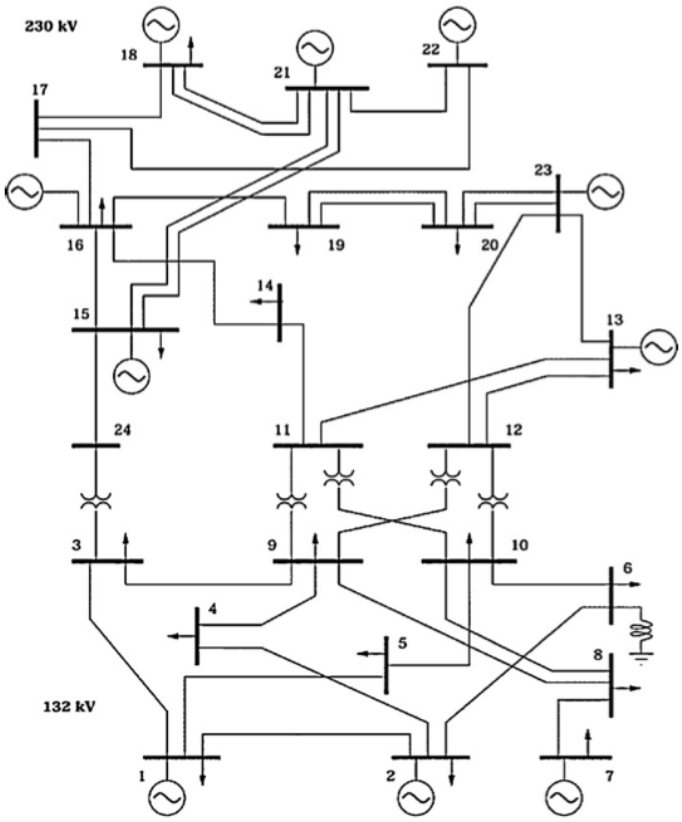


Figure G.1: 24 Bus System [19]

Some of the system patterns represent extreme conditions of the system, e.g. all the lines congested at the same time in the 24 bus system, which is almost impossible to exist in any case. Since the vectors of loads are uniformly (and randomly) generated between 0 and their maxima, the ratio of points corresponding to the system pattern i, denoted by $n_s^{(i)}/n_s$, may sketch the possibility that the system pattern i could happen. The IEEE 24 bus system is taken as an example. The cumulative possibility function and probability density function of the system patterns are plotted (Fig. G.2). As shown in Fig. G.2, around 25 system patterns dominate among all the 445 distinct system patterns. Those dominating system patterns are defined as *major system patterns*. The system is more likely to be operated in the major system patterns, and those major system patterns should attract more attention.

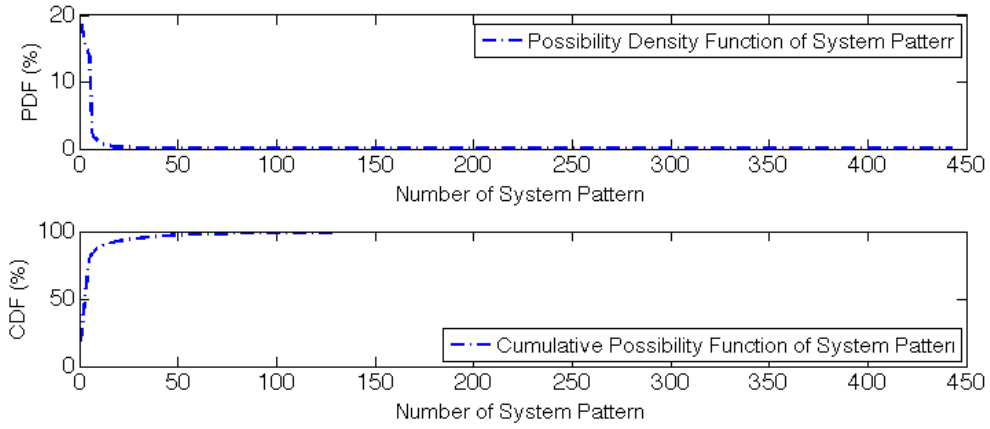


Figure G.2: Cumulative Distribution of System Patterns

The data-driven approach is examined on the IEEE 24 bus system (Fig. G.1). Around 10^5 of testing points are generated from the Monte-Carlo simulation, and training data sets of different sizes are generated in the same way. Fig. G.3 shows

the classification accuracy and training time with the increasing size of training data set.

As shown in Fig. G.2, there are 445 distinct system patterns found in the 24 bus system. If we utilize the "one-vs-one" multi-class SVM on all the 445 system patterns, we need to find $C_{445}^2 = 79800$ optimal separating hyperplanes. There would be both computational time and memory issues. Therefore, we focus on the 25 *major system patterns* which corresponds to 95% of all the system patterns and ignore the other *minor system patterns*. By doing this, we only need to find $C_{25}^2 = 300$ optimal hyperplanes, and the computational time is massively reduced (about 3% of the computational time using all the system patterns). The maximum classification accuracy only degrades from 100% to 95%.

As shown in Fig. G.3, the data-driven approach behaves better with more training points. However, if more than 1×10^4 data points are used for training, the performance is improved slower and will reach the upper limit around 91%. Furthermore, too many training data points would lead to computation time and memory issues. As shown in Fig. G.3, the training time increases almost linearly with the increasing size of training data set. A compromise has to be reached between the accuracy and the training time of the data-driven approach.

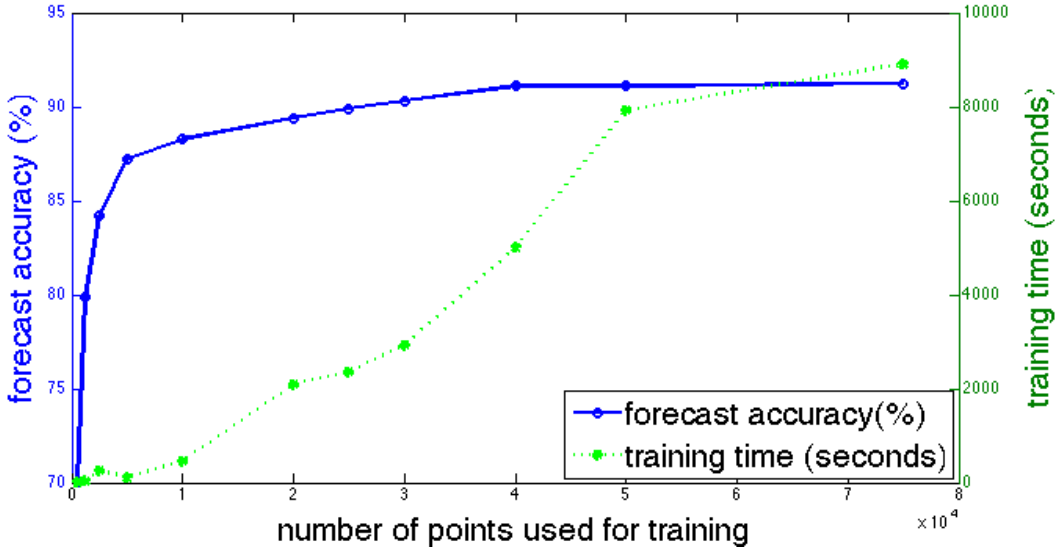


Figure G.3: Forecast Accuracy And Training Time