In *Advanced Data Analytics for Power Systems,* Cambridge University Press, pp. 28-49 2021.

1 Probabilistic Forecasting of Power System and Market Operations

Yuting Ji[a], Lang Tong[b], and Weisi Deng[c]

Abstract

This article presents an online learning approach to probabilistic forecasting of real-time system and market operations. Specifically, the proposed approach produces a forecast of marignal or joint probability distributions of variables of interest such as the locational marginal prices (LMPs), power flows on lines, and the level of reserves.

A fundamental challenge in probabilistic forecasting for large systems is the scalability. As the size of the system and the complexity of stochasticity increase, standard techniques based on direct Monte Carlo simulations and classical methods of statistical inference become intractable. This article presents an alternative approach based on an online dictionary learning that overcomes the curse of dimensionality.

1.1 Introduction

The increasing penetration of renewable resources has changed the operational characteristics of the power systems and electricity markets, from one relying on deterministic and static planning to one involving highly stochastic and dynamic operations. For instance, the net load profile in some areas now follows a so-called "duck curve" [1] where there is a steep down-ramp during the hours when a large amount of solar power is injected into the network followed by a sharp up-ramp when the solar power drops in late afternoon hours.

While the duck curve phenomenon represents an *average* net-load behavior, it is the highly stochastic and spatial-temporal dependent ramp events that present some of the most difficult and costly operational challenges to system operators. In such new operation regimes, the ability to adapt to changing environments

[a] Stanford University
[b] Cornell University
[c] China Southern Power Grid Co., Ltd.
[d] This work is supported in part by the National Science Foundation under Awards 1809830 and 1816397

and managing risks arising from complex scenarios of contingencies is essential. For this reason, there is a need for informative and actionable characterization of the overall operation uncertainty over an extended planning horizon, one that reveals interdependencies of power flows, congestions, reserves, and locational marginal prices (LMPs).

Some system operators are providing real-time price forecasts currently. The Electric Reliability Council of Texas (ERCOT) [2], for example, offers one-hour ahead real-time LMP forecasts, updated every 5 minutes. Such forecasts signal potential shortage/oversupply caused by the anticipated fall/rise of renewable supplies or the likelihood of network congestions. The Alberta Electric System Operator (AESO) [3] provides two short-term price forecasts with prediction horizons of 2 hours and 6 hours, respectively. The Australia Energy Market Operator (AEMO) [4] provides 30 minutes to 32-hour spot price and load forecasts.

Most LMP forecasts, especially those provided by system operators, are *point forecasts* that predict directly future LMP values. Often they are generated by substituting actual realizations of stochastic load and generation by their expected (thus deterministic) trajectory in the calculation of future generation and price quantities. Such *certainty equivalent* approaches amount to approximating the expectation of a function of random variables by the function of the expectation of random variables; they can lead to gross mischaracterization of the behavior of the system operation and systematic error in forecasting.

We consider in this article the problem of *online probabilistic forecasting* in real-time wholesale electricity market from an operator perspective. Specifically, we are interested in using real-time SCADA and PMU measurements along with operating conditions known to the operator to produce short-term probabilistic forecasts of nodal prices, power flows, generation dispatches, operation reserves and discrete events such as congestions and occurrences of contingencies. Unlike a point forecast that predicts the realization of a random variable, say the LMP at a particular bus one hour ahead, probabilistic forecasting produces the probability distribution of the random variable one hour ahead conditional on the real-time measurements obtained at the time of prediction.

For an operator, probabilistic forecasting provides essential information for evaluating operational risks and making contingency plans. It is only with probabilistic forecasts that stochastic optimization becomes applicable for unit commitment and economic dispatch problems [5]. For the market participants, on the other hand, probabilistic forecasts of prices are essential signals to elicit additional generation resources when the probability of high prices is high and to curtail otherwise. Indeed, in the absence of a day-ahead market, continuing price and load forecasts by the independent market operator that covers 30 minutes and longer forecasting horizons play an essential role in the Australian national electricity market [6].

1.1.1 Related Work

There is an extensive literature on point forecasting techniques from perspectives of external market participants. See [7, 8] for a review. These techniques typically do not assume having access to detailed operating conditions. Results focusing on probabilistic forecasting by an operator are relatively sparse. Here we highlight some of the relevant results in the literature applicable to probabilistic forecasting by system operators.

A central challenge of obtaining probabilistic forecasting is estimating the (joint or marginal) conditional probability distributions of future system variables. One approach is to approximate the conditional distributions based on point estimates. For instance, a probabilistic LMP forecasting approach is proposed in [9] based on attaching a Gaussian distribution to a point estimate. While the technique can be used to generalize point forecasting methods to probabilistic ones, the Gaussian distribution is, in general, a poor approximation as there typically multiple distribution modalities caused by different realizations of binding constraints in the optimization. The authors of [10] and [11] approximate the probabilistic distribution of LMP using higher order moments and cumulants. These methods are based on representing the probability distribution as an infinite series involving moments or cumulants. In practice, computing or estimating higher order moments and cumulants are challenging; lower order approximations are necessary.

A more direct approach is to estimate the conditional distributions directly. Other than some simple cases, however, probabilistic forecasting in a large complex system can only be obtained by Monte Carlo simulations where conditional distributions are estimated from sample paths generated either according to the underlying system model or directly from measurements and historical data. In this context, the problem of probabilistic forecasting is essentially the same as online Monte Carlo simulations. To this end, there is a premium on reducing computation costs.

The idea of using simulation techniques for LMP forecasting was first proposed in [12], although probabilistic forecasting was not considered. Min *et al.* proposed in [13] a direct implementation of Monte Carlo simulations to obtain short-term forecasting of transmission congestion. For M Monte Carlo runs over a T-period forecasting horizon, the computation cost is dominated by the computation of $M \times T$ direct current optimal power flow (DCOPF) solutions that are used to generate the necessary statistics. For a large scale system with a significant amount of random generations and loads, such computation costs may be too high for such a technique to be used for online forecasting.

A similar approach based on a nonhomogeneous Markov chain modeling of real-time LMP was proposed in [14]. The Markov chain technique exploits the discrete nature of LMP distributions and obtains LMP forecasts by the product of transition matrices of LMP states. Estimating the transition probabilities, however, requires roughly the same number of Monte Carlo simulations, thus requiring approximately the same number of DCOPF computations.

A significant reduction of simulation cost is achieved by exploiting the structure of economic dispatch by which random variables (such as stochastic generations) enter the optimal power flow (OPF) problem. In [15], a multiparametric programming formulation is introduced where random generations and demands are modeled as the right-hand side parameters in the constraints of the DCOPF problem. From the parametric linear/quadratic programming theory, the (conditional) probability distributions of LMP and power flows, given the current system state, reduce to the conditional probabilities that realizations of the random demand and generation fall into one of the *finite* number of *critical regions* in the parameter space. The reduction of the modeling complexity from high dimensional continues random variables of stochastic generations and loads to a discrete random variable with a finite number of realizations represents a fundamental step toward a scalable solution to probabilistic forecasting.

Although the multiparametric programming approach in [16] reduces the modeling complexity to a finite number of critical regions in the parameter space, the cardinality of the set of critical regions grows exponentially with the number of constraints in the DCOPF problem. For a large power system with many constraints, the cost of characterizing these critical regions, even if performed offline, have exponential complexity in computation and storage.

In [17], an online dictionary learning (ODL) approach is proposed. The main idea is to avoid computing and store all critical regions ahead of time, using instead online learning to acquire and update sequentially a dictionary that captures the parametric structure of DCOPF solutions. In particular, each entry of the dictionary corresponds to an *observed* critical region within which a sample of random generation/demand has fallen. A new entry of the dictionary is produced only when the realization of the renewable generation and demand does not fall into one of the existing critical regions in the dictionary. This avoids costly DCOPF computations and recalls the solution directly from the dictionary. Because renewable generation and load processes are physical processes, they tend to be bounded and concentrated around the mean trajectory. As a result, despite that there is potentially an exponentially large number of potential entries in the dictionary, only a tiny fraction of the dictionary entries are observed in the simulation process.

1.1.2 Summary and Organization

This article highlights and extends recent advances in simulation-based probabilistic forecasting techniques [16, 17]. We focus on simulation-based approaches for two reasons. First, simulation-based forecasting is so far the only type of techniques for large complex power systems that can yield *consistent estimates* of conditional probability distributions. Second, although computation costs of such methods have long been recognized as the primary barrier for their applications in large systems, novel computation and machine learning techniques and new computation resources such as cloud computing and graphical processing units (GPUs) will lead to a scalable online forecasting platform for system operators.

This chapter is organized as follows. Section 1.2 presents a general simulation-based approach to probabilistic forecasting. Examples of real-time operation models are provided in Section 1.3. Section 1.4 summarizes key structural results on multiparametric programming for developing the probabilistic forecasting algorithm. Details of the proposed forecasting approach is presented in Section 1.5, followed by case studies in Section 1.6 and concluding remarks in Section 1.7.

1.2 Probabilistic Forecasting via Monte Carlo Simulations

We highlight in this section a general approach to probabilistic forecasting based on Monte Carlo simulation. *Probabilistic forecast*, in contrast to point forecast, aims to provide (marginal or joint) probability distributions of future quantities of interest such as load, power flow, reserve, LMP, etc.

Two critical components of simulation-based probabilistic forecasting are (i) a model that captures the physical characteristics of the system and (ii) a probability model that characterizes the underlying random phenomena in the system. For power system and market operations, the former includes a physical model of the underlying power system and an algorithmic model of the decision process for unit commitment, economic dispatch, and LMP. The latter comes from probabilistic load/generation forecasts and probability models for contingencies.

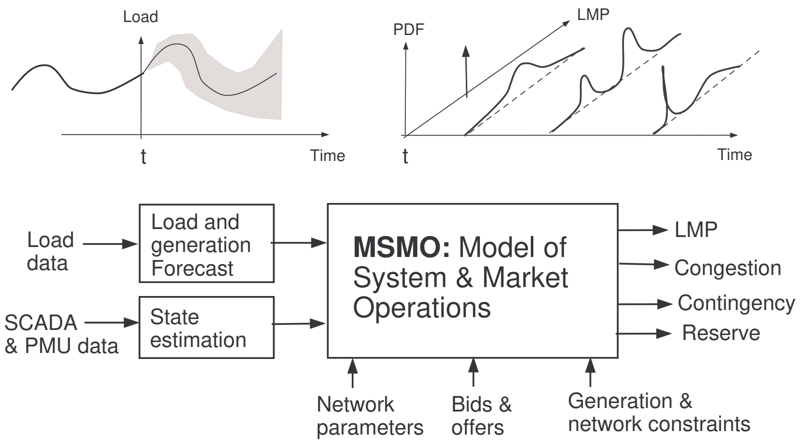

Figure 1.1 Schematics of simulation based probabilistic forecasting.

Fig. 1.1 illustrates a general schematic of simulation-based probabilistic forecasting, similar to that in [12]. The main engine of the probabilistic forecasting is a computational model of real-time system and market operations (MSMO). Details of MSMO are subjects of Sections 1.3.

One set of inputs to MSMO are the exogenous *random processes* from real-time measurements from SCADA or PMUs that define the system state S_t at the time of forecasting t. Another set of inputs are the probabilistic forecasts of generations and loads. Shown in Fig. 1.1 is an illustration of probabilistic load forecast at time t where the grey area represents possible load trajectories of the future. Probabilistic forecasting of load and renewable generation has been studied extensively. See review articles [18, 19, 20] and references therein.

MSMO also imports a set of *deterministic parameters* that characterize the network condition (topology and electrical parameters), bids and offers from market participants, and operation constraints such as generation and line capacities. Using a physical model of the power system, the MSMO simulates the actual system operations from generated scenarios using probabilistic load and generation forecasts. Statistics of variables of interests such as LMP, congestion patterns, levels of the reserve, etc. are collected and presented to the operator.

The output of MSMO is the set of the conditional probability distributions of the variables of interest. In particular, given the system state $S_t = s$ estimated from PMU/SCADA data, MSMO outputs the conditional probability distribution $f_{t+T|t}$ of LMP at time $t + T$ in the form of a histogram or parametric distributions learned from LMP samples generated by MSMO.

1.3 Models of System and Market Operations

Most wholesale electricity markets consist of day-ahead and real-time markets. The day-ahead market enables market participants to commit to buy or sell wholesale electricity one day ahead of the operation. The day-ahead market is only binding in that the market participants are obligated to pay or be paid for the amount cleared in the day ahead market at the day-ahead clearing price. The amount of energy cleared in the day ahead market is not related to the actual power delivery in real-time.

The real-time system operation determines the actual power delivery based on the actual demand and supply. The real-time market, in contrast to the day ahead market, is a *balancing market* that prices only the differences between day-ahead purchases and the actual real-time demand and production.

Modern power systems and electricity markets are highly complex. Here we present a stylized parametric model for real-time system and market operations, which captures essential features of the optimal economic dispatch, energy-reserve co-optimization, and related LMP computations.

1.3.1 A Multi-parametric Model for Real-time System and Market Operations

We defer to later sections for the detailed specification of several real-time operations. Here we present a generic multiparametric optimization model that underlies primary real-time system and market operations.

Specifically, we consider the multi-parametric linear or quadratic program of the following form

$$\underset{x}{\text{minimize}}\ z(x) \text{ subject to } Ax \leq b + E\theta \qquad (y) \tag{1.1}$$

where x is the decision variable typically representing the dispatch of generation or flexible load, $z(\cdot)$ the overall cost function, the inequalities the generation and network constraints, and y the vector of Lagrangian multipliers (dual variables) from which the energy prices are calculated. Special to this optimization is the parameter θ that captures the realized exogenous stochastic generations and demands.

The real-time system is operated in discrete time periods, each of duration, say, one to five minutes. For a twenty-four hour operation, let $\theta_t, t = 1, 2, \cdots, T$ be the sequence of realized stochastic demands and (renewable) generations. The *single period operation model* is to solve a sequence optimizations of the form (1.1) for each realized θ_t to determine the optimal generation dispatch x_t^* and related LMP y_t^*. When θ_t are drawn repeatedly from a probability distribution based on load and generation forecasts, we obtain samples of dispatch levels and LMPs from which their distributions can be estimated. Sections 1.3.2, 1.3.4 describes techniques under the single-period operation model.

In practice, there are the so-called *ramping constraints* for generators on how much the generation level can change from one interval to the next. This means that, given a sequence of realized demand θ_t, obtaining optimal dispatch independently one interval at a time according to the single operation model may violate the ramping constraints, leading to an infeasible dispatch sequence.

One approach to deal with significant ramping events is to call up the reserve in cases of shortage, which is a costly proposition. A more economic approach is to schedule generations based on a *multi-period operation* model in which generation levels of the entire operation horizon are considered jointly. For instance, at time t, if the future load and stochastic generation can be forecasted perfectly as $\theta = (\theta_t, \theta_{t+1}, \cdots, \theta_{t+T'})$, the problem of jointly determining generation levels $x^* = (x_t^*, x_{t+1}^*, \cdots, x_{t+T'}^*)$ can be solved from (1.1).

In practice, one does not have the perfect prediction of $\theta_{t+1}, \cdots \theta_{t+T'}$, the problem of determining the optimal sequence of dispatch becomes on of multi-period stochastic (or robust) optimizations, for which the computation complexity is substantially higher. Practical solutions based on certainty equivalent heuristics or model predictive control (MPC) are used in practice. We present one such approach in Sec. 1.3.3.

We should note that, unlike computing (and forecasting) LMP under the single period operation model, pricing dispatch under the multi-period operation model is highly nontrivial and is an area of active research. See [21] and references therein.

1.3.2 Single Period Economic Dispatch

Here we consider the problem of economic dispatch in the energy-only real-time market under the single period operation model. The system operator sets generation adjustments by solving a so-called DCOPF problem in which the one-step-ahead real-time demand is balanced subject to system constraints.

For simplicity, we assume that each bus has a generator and a load. The single-period DCOPF problem solves, in each time period t, for the optimal generation dispatch vector g_t^* given the forecast demand d_t in period t subject to generation and network constraints from the following optimization:

$$
\begin{aligned}
\underset{g}{\text{minimize}} \quad & C_g(g) && && (1.2)\\
\text{subject to} \quad & \mathbf{1}^\top(g - d_t) = 0 && (\lambda_t) && (1.3)\\
& S(g - d_t) \leq F^{\max} && (\mu_t) && (1.4)\\
& G^{\min} \leq g \leq G^{\max} && && (1.5)\\
& \hat{g}_{t-1} - R^{\max} \leq g \leq \hat{g}_{t-1} + R^{\max} && && (1.6)
\end{aligned}
$$

where

$C_g(\cdot)$	real-time generation cost function;
$\mathbf{1}$	the vector of all ones;
d_t	vector of net load forecast at time t;
g	vector of ex-ante dispatch at time t;
$\hat{g}_{t-1}$	vector of generation estimate at time $t-1$;
d_t	vector of one-step net load[1] forecast at time t;
$F^{\max}$	vector of transmission capacities;
$G^{\max}$	vector of maximum generator capacities;
$G^{\min}$	vector of minimum generator capacities;
$R^{\max}$	vector of ramp limits;
S	power transfer distribution factor matrix;
λ_t	shadow price for the energy balance constraint at time t;
μ_t	shadow prices for transmission constraints at time t.

In the above optimization, the first equality constraint (1.3) represents the power balance dictated by the Kirchhoff law, the second (1.4) on the maximum power transfer over each branch, and the third (1.5) on the maximum and minimum restrictions on generations. The last constraint (1.6 is on the up and down ramping capabilities of generators from the previously set dispatch levels.

The above model clearly is an instance of the general parametric program defined in (1.1). In this model, the generation costs can be linear or strictly convex quadratic. The real-time LMP π_{it} at bus i and time t is defined by the marginal cost of demand d_{it} at that bus. In other words, the LMP is the total cost increase induced by an ϵ increase of demand d_{it}. In the limit,

$$\pi_{it} = \frac{\partial}{\partial d_{it}} C(g^*(d_t)).$$

By the envelop theorem, the LMP vector π_t can be computed from the dual variable y^* as

$$\pi_t = \lambda_t^* \mathbf{1} - S^\top \mu_t^*, \tag{1.7}$$

where the first term corresponds to the sum of energy prices λ^* and weighted sum of congestion prices μ^*. Note that the ith entry of μ^* corresponds to the constraint associated with the ith branch of the system. Thus we have $\mu_i^* > 0$ only if the ith branch is *congested, i.e.,* the power flow constraint on branch i is binding.

1.3.3 Multi-Period Economic Dispatch with Ramping Products

With increasing levels of variable energy resources and behind the meter generation, the operational challenge of ramping capability becomes more prominent. ISOs [22, 23] are adopting market-based "ramp products" to address the operational challenges of maintaining the power balance in the real-time dispatch. Here we present a multi-period economic dispatch model [24] based on the so-called *flexible ramping product (FRP)* [22] recently adopted in the California Independent System Operator (CAISO).

Given the load forecast $\bar{d}_t$ for the next T period, the following optimization, again in the general form of (1.1), produces a sequence of dispatch levels (g_t^*) and up and down ramping levels $(r_t^{\text{up}}, r_t^{\text{down}})$:

$$\min_{\{g_t, r_t^{\text{up}}, r_t^{\text{down}}\}} \quad \sum_{t=t_0+1}^{t_0+T} C_g(g_t) + C_r^{\text{up}}\left(r_t^{\text{up}}\right) + C_r^{\text{down}}\left(r_t^{\text{down}}\right) \tag{1.8}$$

$$\begin{aligned}
\text{subject to} \quad & \forall t = t_0+1, \cdots, t_0+T, && && (1.9)\\
& \mathbf{1}^\top (g_t - \bar{d}_t) = 0, && (\lambda_t) && (1.10)\\
& S(g_t - \bar{d}_t) \le F^{\max}, && (\mu_t) && (1.11)\\
& g_t + r_t^{\text{up}} \le G^{\max}, && && (1.12)\\
& g_t - r_t^{\text{down}} \ge G^{\min}, && && (1.13)\\
& g_t - g_{t-1} + r_{t-1}^{\text{up}} + r_t^{\text{up}} \le R^{\max}, && && (1.14)\\
& g_t - g_{t-1} + r_{t-1}^{\text{down}} + r_t^{\text{down}} \le R^{\max}, && && (1.15)\\
& \mathbf{1}^\top r_t^{\text{up}} \ge R_t^{\text{up}}, && (\alpha_t) && (1.16)\\
& \mathbf{1}^\top r_t^{\text{down}} \ge R_t^{\text{down}}, && (\beta_t) && (1.17)\\
& r_t^{\text{up}} \ge 0, r_t^{\text{down}} \ge 0. && && (1.18)
\end{aligned}$$

[1] In this model, we use the concept of "net load" d_t. Since renewable generation can be considered as a negative load, we define the net load as the total electrical load plus interchange minus the renewable generation. The interchange schedule refers to the total scheduled delivery and receipt of power and energy of the neighboring areas.

where

$C^g(\cdot)$	energy generation cost function;
$C^{\text{up}}(\cdot)$	cost function to provide upward ramping;
$C^{\text{down}}(\cdot)$	cost function to provide downward ramping;
$\bar{d}_t$	vector of net load forecast at time t *conditioning on the system state at time* t_0;
g_t	vector of generation at time t;
r_t^{up}	vector of upward flexible ramping capacity at time t;
r_t^{down}	vector of downward flexible ramping capacity at time t;
$F^{\max}$	vector of transmission capacities;
$G^{\max}$	vector of maximum generator capacities;
$G^{\min}$	vector of minimum generator capacities;
S	shift factor matrix;
$R^{\max}$	vector of ramp limits;
R_t^{up}	upward ramping requirement of the overall system at time t;
R_t^{down}	downward ramping requirement of the overall system at time t.

This multi-period economic dispatch is performed at time t_0 to minimize the overall system cost, consisting of energy generation cost and generation ramping cost, over time steps $t = t_0+1, t_0+2, \cdots, t_0+T$. The constraints of power balance (1.10) and branch flow capacity (1.11) are the same as in single-period economic dispatch. Constraints (1.12)-(1.15) correspond to the capacity of generations and ramping constraints. The rest of constraints (1.16)-(1.18) enforce the flexible ramping products to satisfy the ramping requirements.

Ramping requirements R_t^{up} and R_t^{down} will ensure there is sufficient ramping capability available to meet the forecasted netload. In practice, ISO uses the historical forecast error to calculate the distribution of ramping needs. The last constraint is the risk-limiting constraint, which implies that the system operator needs to meet the actual demand at all times with the probability of at least p.

The practice used by CAISO [25] to determine the ramping requirements is as follows. The values of R_t^{up} and R_t^{down} are chosen to achieve confidence level of p with respect to the point prediction of load $\bar{d}_t$ at time t:

$$\mathcal{P}[\mathbf{1}^\top \bar{d}_t - R_t^{\text{down}} \leq \mathbf{1}^\top d_t \leq \mathbf{1}^\top \bar{d}_t + R_t^{\text{up}}] \geq p \tag{1.19}$$

where d_t is the actual load at time t.

Because the sequence of demand forecasts ($\bar{d}_t$) is never perfect, the further ahead of the forecast, the higher the forecast error, only the dispatch g_t^* is implemented in reality. The dispatch at time $t_0 + 1$ is determined by solving the above optimization upon receiving the updated forecasts. Such type of sequential decision processes follows the principle of model predictive control (MPC).

The MPC approach seeks to perform the economic dispatch for time steps $t = t_0 + 1, t_0 + 2, \cdots, t_0 + T$ under the condition that ramping capacity needs to be reserved for steps $t = t_0 + 1, t_0 + 2, \cdots, t_0 + T - 1$. Ramping capacity for time $t = t_0$ has been reserved in the previous time step, hence, there are no variables to be determined. Note that the load predictions are updated as time goes by.

Hence, only the energy dispatch profile for $t = t_0$, i.e., g_{t_0}'s, and the flexible ramping requirements for $t = t_0 + 1$, i.e., $r^{\text{down}}_{t_0+1}$ and $r^{\text{up}}_{t_0+1}$, will be applied.

By the envelop theorem, the energy LMP vector π_t at time t is given by

$$\pi_t = \mathbf{1}\lambda_t^* + S^\top \mu_t^*. \tag{1.20}$$

The clearing prices for upward ramping and downward ramping at time t are α_t^* and β_t^* respectively.

1.3.4 Energy and Reserve Co-Optimization

In the joint energy and reserve market, dispatch and reserve are jointly determined via a linear program that minimizes the overall cost subject to operating constraints. In the co-optimized energy and reserve market, system-wide and locational reserve constraints are enforced by the market operator to procure enough reserves to cover the first and the second contingency events. We adopt the co-optimization model in [26] as follows:

$$\begin{aligned}
\underset{g,r,s}{\text{minimize}} \quad & \sum_i \left(c_i^g g_i + \sum_j c_{ij}^r r_{ij} \right) + \sum_u c_u^p s_u^l + \sum_v c_v^p s_v^s && && (1.21)\\
\text{subject to} \quad & \mathbf{1}^\top (g - d) = 0 && (\lambda) && (1.22)\\
& S(g-d) \le F^{\max} && (\mu) && (1.23)\\
& \sum_{ij} \delta_{iju}^l r_{ij} + (I_u^{\max} - S_u^{\text{int}}(g-d)) + s_u^l \ge Q_u^l,\ \forall u && (\alpha_u) && (1.24)\\
& \sum_{ij} \delta_{ijv}^s r_{ij} + s_v^s \ge Q_v^s,\ \forall v && (\beta_v) && (1.25)\\
& G_i^{\min} \le g_i + \sum_j r_{ij} \le G_i^{\max},\ \forall i && && (1.26)\\
& R^{\text{down}} \le g - \hat{g}_{t-1} \le R^{\text{up}} && && (1.27)\\
& 0 \le r \le R^{\max} && && (1.28)\\
& s_u^l, s_v^s \ge 0,\ \forall u, v && && (1.29)
\end{aligned}$$

where

i	index of buses;
j	index of reserve types, 10-min spinning, 10-min non spinning, or 30-min operating reserve;
u/v	index of locational/system-wide reserve constraints;
k	index of transmission constraints;
d_i	net load at bus i;
g_i	dispatch of online generator at bus i;
r_{ij}	generation reserve of type j on bus i;
s^l/s^s	local/system reserve deficit of constraint;
c_i^g	cost for generation at bus i;
c_{ij}^r	cost for reserve type j at bus i;
$c_{u/v}^p$	penalty for reserve deficit of constraint u/v;
$I_u^{\max}$	interface flow limit for locational reserve constraint u;
$F^{\max}$	vector of transmission capacities;
Q^l/Q^s	locational/system-wide reserve requirement of constraint;
S	shift factor matrix for transmission lines;
S^{int}	shift factor matrix for interface flows;
δ_{iju}^l	binary value that is 1 when reserve j on bus i belongs to locational reserve constraint u;
δ_{ijv}^s	binary value that is 1 when reserve j on bus i belongs to system-wide reserve constraint v;
$G_i^{\max/\min}$	max/min generation capacity for generator at bus i;
$R^{\text{up/down}}$	vector of upward/downward ramp limits;
$R^{\max}$	vector of reserve capacities.

In this model, the real-time dispatch problem is formulated as a linear programming problem with the objective to maximize the social surplus, subject to real-time operating constraints and the physical characteristics of resources. Energy balance constraint (1.22) and transmission constraint (1.23) are the same as in the economic dispatch for the energy-only market. The system-wide reserve requirement constraint (1.25) is enforced for the market operator to procure enough reserves to cover the first contingency events. A locational reserve constraint (1.24) is used to cover the second contingency event caused by the loss of a generator or a second line in a local area. Therefore, the unloaded tie-line capacity ($I_u^{\max} - S_u^{\text{int}}(g - d)$), as well as the reserve provided by units located in the local reserve zone, can be utilized to cover the local contingency or reserve requirement within 30 min. Note that the interface flow is calculated the same way as transmission line flow using the shift factor matrix S^{int}. Constraints (1.26)-(1.27) correspond to the generation capacity and ramping limits respectively.

Based on the envelop theorem, the energy price vector π for all buses and the reserve clearing price ρ_{ij} of each reserve product j at bus i are defined as

$$\pi = \lambda^* \mathbf{1} - S^\top \mu^* + (S^{\text{int}})^\top \alpha^* \tag{1.30}$$

$$\rho_{ij} = \sum_u \alpha_u^* \delta_{iju}^l + \sum_v \beta_v^* \delta_{ijv}^s \tag{1.31}$$

where $\lambda^*, \mu^*, \alpha^*, \beta^*$ are the optimal values of Lagrangian multipliers.

Note again that the energy-reserve co-optimization model is also of the form of parametric program of (1.1) with parameter $\theta = (d_t, \hat{g}_{t-1})$ that is realized prior to the co-optimization.

1.4 Structural Solutions of Multiparametric Program

We have seen in the section that many real-time market operations can be modeled as a general form of the multi-parametric linear or quadratic program defined in (1.1). Here we present several key structural results on multiparametric linear/quadratic programming that we use to develop our approach. See [27, 28, 29] for multiparametric programming for more comprehensive expositions.

1.4.1 Critical Region and its Geometric Structure

Recall the multiparametric (linear/quadratic) programm defined earlier

$$\underset{x}{\text{minimize}}\ z(x) \text{ subject to } Ax \le b + E\theta \qquad (y) \tag{1.32}$$

where the decision variable x is in $\mathcal{R}^n$, the cost function $z(x)$ linear or quadratic, and the parameter θ is in a bounded subspace $\Theta \subset \mathcal{R}^m$.

Let the optimal primal solution be $x^*(\theta)$, the associated dual solution $y^*(\theta)$, and the value of optimization $z^*(\theta)$. We will assume that the MLP/MQP is not (primal or dual) degenerate[2] for all parameter values. Approaches for the degeneracy cases are presented in [29].

The solution structure of (1.32) is built upon the notion of *critical region partition*. There are several definitions for critical region. Here we adopt the definition from [29] under primal/dual non-degeneracy assumption.

DEFINITION 1.1 A *critical region* $\mathcal{C}$ is defined as the set of all parameters such that for every pair of parameters $\theta, \theta' \in \mathcal{C}$, their respective solutions $x^*(\theta)$ and $x^*(\theta')$ of (1.1) have the same active/inactive constraints.

To gain a geometric insight into this definition and its implication, suppose that, for some θ_o, the constraint of (1.32) is not binding at the optimal solution, *i.e.,* $Ax^*(\theta_o) < b + E\theta_o$. Then the constraint (1.32) is none binding for all θ's in a small enough neighborhood of θ_o. Thus this small neighborhood belongs to

[2] For a given θ, (1.32) is said to be primal degenerate if there exists an optimal solution $x^*(\theta)$ such that the number of active constraints is greater than the dimension of x. By dual degeneracy we mean that the dual problem of (1.32) is primal degenerate.

the critical region $\mathcal{C}_o$ in which all linear inequality constraints of (1.32) are none binding at their own optima. We can then expand this neighborhood until one of the linear inequalities becomes binding. That particular linear equality defines a boundary of $\mathcal{C}_0$ in the form of a hyperplane.

Conceptually, the above process defines the critical region $\mathcal{C}_0$ of the form of a polyhedron. For a quadratic objective function, because all linear inequality constraints are none binding for all $\theta \in \mathcal{C}_0$, the solution of (1.32) must all be of the form $x^*(\theta) = f_0$.

Similarly, if we take θ_1 for which only one linear inequality constraint is binding at $x^*(\theta_1)$. We can then obtain a different polyhedral critical region $\mathcal{C}_1$ containing θ_1, for which the same constraint binding condition holds. By ignoring all none binding constraints and eliminating one of the variables, say x_1 in the equality constraint, we can solve an unconstrained quadratic optimization with θ only appears in the linear term. Thus, for all $\theta \in \mathcal{C}_1$, the optimal solution is of a parametric affine form

$$x^*(\theta) = F_1\theta + f_1 \quad \text{for all } \theta \in \mathcal{C}_1,$$

where F_1 and f_1 are independent of $\theta \in \mathcal{C}_1$.

The significance of the above parametric form of the solution is that, once we have (F_1, f_1), we no longer need to solve the optimization whenever $\theta \in \mathcal{C}_1$.

It turns out that every θ is in one and only one critical region. Because there is an only finite number of binding constraint patterns, the parameter space Θ is partitioned into a finite number of critical regions. Within critical region $\mathcal{C}_1$, the optimal solution is of the form $x^*(\theta) = F_i\theta + f_i$ for all $\theta \in \mathcal{C}_1$.

1.4.2 A Dictionary Structure of MLP/MQP Solutions

We now make the intuitive arguments given above precise mathematically and computationally. In particular, we are interested in not only the existence of a set of a finite number critical regions $\{\mathcal{C}_i\}$ that partitions the parameter space, but also the computation procedure to obtain these critical regions and their associated functional forms of the primal and dual solutions.

The following Theorem summarizes the theoretical and computational basis for the proposed probabilistic forecasting approach.

THEOREM 1.2 ([16, 17]) *Consider (1.32) with cost function $z(x) = c^\top x$ for MPLP and $z(x) = \frac{1}{2}x^\top Qx$ for MPQP where Q is positive definite. Given a parameter θ_0 and the solution of the parametric program $x^*(\theta_0)$, let $\tilde{A}, \tilde{E}$ and $\tilde{b}$ be, respectively, the submatrices of A, E and subvector of b corresponding to the active constraints. Let $\bar{A}, \bar{E}$ and $\bar{b}$ be similarly defined for the inactive constraints. Assume that (1.32) is neither primal nor dual degenerate.*

(1). For the MPLP, the critical region $\mathcal{C}_0$ that contains θ_0 is given by

$$\mathcal{C}_0 = \left\{\theta \middle| (\bar{A}\tilde{A}^{-1}\tilde{E} - \bar{E})\theta < \bar{b} - \bar{A}\tilde{A}^{-1}\tilde{b}\right\}. \tag{1.33}$$

And for any $\theta \in \mathcal{C}_0$, the primal and dual solutions are given by, respectively,

$$x^*(\theta) = \tilde{A}^{-1}(\tilde{b} + \tilde{E}\theta), \quad y^*(\theta) = y^*(\theta_0).$$

(2). For the MPQP, the critical region $\mathcal{C}_0$ that contains θ_0 is given by

$$\mathcal{C}_0 = \{\theta | \theta \in \mathcal{P}_p \bigcap \mathcal{P}_d\}, \tag{1.34}$$

where $\mathcal{P}_p$ and $\mathcal{P}_d$ are polyhedra defined by

$$\begin{aligned}\mathcal{P}_p &= \{\theta | \bar{A}H^{-1}\tilde{A}^\top (\tilde{A}Q^{-1}\tilde{A}^\top)^{-1}(\tilde{b} + \tilde{E}\theta) - \bar{b} - \bar{E}\theta < \mathbf{0}\},\\ \mathcal{P}_d &= \{\theta | (\tilde{A}Q^{-1}\tilde{A}^\top)^{-1}(\tilde{b} + \tilde{E}\theta) \le \mathbf{0}\}.\end{aligned}$$

And for any $\theta \in \mathcal{C}_0$, the primal and dual solutions are given by

$$x^*(\theta) = H^{-1}\tilde{A}^\top (\tilde{A}H^{-1}\tilde{A}^\top)^{-1}(\tilde{b} + \tilde{E}\theta), \tag{1.35}$$

$$y_i^*(\theta) = \begin{cases} 0 & \textit{the ith constraint is inactive} \\ -e_i^\top (\tilde{A}H^{-1}\tilde{A}^\top)^{-1}(\tilde{b} + \tilde{E}\theta) & \textit{otherwise} \end{cases} \tag{1.36}$$

where e_i is the unit vector with ith entry equal to one and zero otherwise.

Figure 1.4.2 illustrates the geometric structure of MLP/MQP on the parameter space Θ partitioned by a finite number of polyhedral critical regions $\{calC_i\}$. We then attach each critical region $\mathcal{C}_i$ a unique signature, which we shall referred to as a *word* $W_i = (F_i, f_i, G_i, g_i, H_i, h_i)$, that defines completely the primal/dual solutions and the critical region. Specifically, the primal and dual solutions for $\theta \in \mathcal{C}_i$ is given by, respectively,

$$x^*(\theta) = F_i\theta + f_i, \quad y^*(\theta) = G_i\theta + g_i, \tag{1.37}$$

and the critical $\mathcal{C}_i$ is defined by $\mathcal{C}_i = \{\theta | H_i\theta + h_i \le 0\}$.

Thus the complete solution of the MLP/MQP can be viewed as a *dictionary* in which each word of the dictionary defines the solution from within a critical region. In the next section, we present an online learning approach that learns the dictionary adaptively.

1.5 Probabilistic Forecasting via Online Dictionary Learning

In this section, we present an online learning approach to probabilistic forecasting aimed at achieving computation scalability for large networks and a high degree of accuracy that requires a large number of Monte Carlo runs. To this end, we first examine the computation costs of probabilistic forecasting, which motivates the development of an online dictionary learning solution.

In developing a tractable forecasting solution, we emulate the process by which a child learn how to speak a language. A natural way is to acquire words as the child encounters them. Along the process, the child accumulates a set of words most useful for her. Even if a word was known to her earlier but is forgotten later, she can relearn the word making the words less likely to forget. Among all the

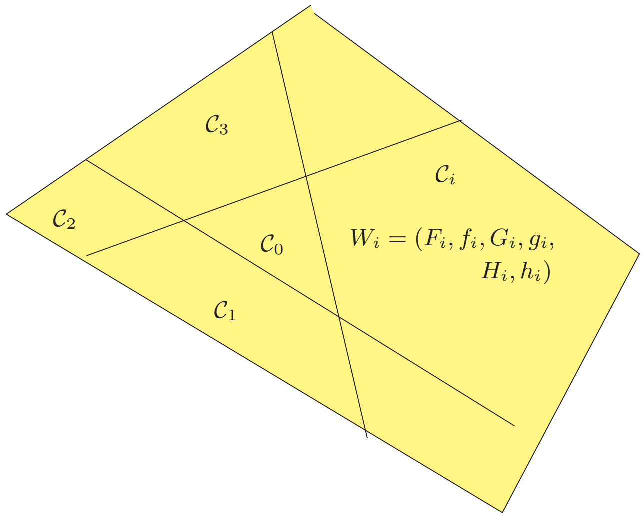

Figure 1.2 Illustration of the geometry of the solution structure on the parameter space Θ.

words in the language dictionary, only a small fraction of them are sufficient for all practical purposes, and the set of words most useful may change dynamically.

1.5.1 Complexity of Probabilistic Forecasting

As outlined in Sec. 1.2, the engine of the probabilistic forecast is the Monte Caro simulations defined by MSMO. Thus the main computation cost comes from repeatedly solving the optimal dispatch problem defined by the multi-parametric program (1.1) for random samples generated from load and generation forecasts.

Consider a system of N buses. The structural results in Theorem 1.2 allows us to solve the LP/QP defined in (1.32) no more than $K(N)$ times where $K(N)$ is the number of critical regions. For fixed N, the complexity in terms of the number of LP/QP calls to achieve arbitrarily accurate probabilistic forecasting is finite. Let M be the number of Monte Carlo simulations needed to generate probabilistic forecasts. To achieve a consistent estimate of the conditional probability distributions, we require $M \to \infty$. The computation costs per Monte Carlo run, in the limit as $M \to \infty$, is given by

$$\lim_{M\to\infty} \frac{M \times O(N^2) + O(K(N)N^{\alpha}}{M} = O(N^2)$$

where the first term in the numerator corresponds to cases when solving LP/QP are unnecessary and the second term for the cases when LP/QP has to be solved (only once per critical region) by a polynomial algorithm. This means that, in high accuracy regime, the computation complexity per Monte Carlo run is roughly that of a matrix-vector multiplication.

On the other hand, in the large system asymptotic regime when $N \to \infty$, the computation cost is dominated by $O(K(N)N^{\alpha})$ where $K(N)$, the number of critical regions for an N bus network. Unfortunately, $K(N)$ may grow exponentially with N, which means that direct Monte Carlo simulations for a large network remain intractable in the worst case.

1.5.2 An Online Dictionary Learning Approach to Probabilistic Forecasting

We leverage the dictionary structure of the MLP/MQP to develop an online learning approach[3] in which critical regions are learned sequentially to overcome the curse of dimensionality. To overcome the memory explosion, we also allow words previously learned but rarely used to be forgotten. It is the combination of sequential learning and dynamic management of remembered words that makes the simulation-based probabilistic forecasting scalable for large systems.

Given that, in the worst case, there may be exponentially (in N) a large number of critical regions for an N-bus system, obtaining analytical characterizations of all critical regions and the associated solution functions can be prohibitive. If, however, we are interested not in the worst case, but in the likely realizations of the stochastic load and generation parameter θ_t, not all critical regions need to be characterized. In fact, since θ_t represents the physical processes of load and generation, it is more likely that θ_t concentrates around its mean. As a result, each realization of θ_t may encounter a few critical regions.

A skeleton algorithm of the online dictionary approach is given in Algorithm 1. We assume that at time t, we have acquired a dictionary $\mathcal{D}_t = \{W_i, i = 1, \cdots, K_t\}$ with K_t entries, each corresponds to a critical region that has been learned from the past. Upon receiving a realization of random load and generation θ_t, the algorithm checks if θ_t belongs to a critical region whose word representation is already in the dictionary. This mounts to search for a word $W_i = (F_i, f_i, G_i, g_i, H_i, h_i)$ such that $H_i\theta_t + h_i \leq 0$. If yes, the primal solution is given by the affine mapping defined by F_i and f_i and dual solution by G_i and g_i. Otherwise, we need to solve (1.32) and obtain a new word W according to Theorem 1.2.

The main intuition of the dictionary learning approach is that the parameter process θ_t represents the physical processes of aggregated load and generation. Such processes tend concentrates around its mean. As a result, each realization of θ_t may encounter a few critical regions. This intuition is confirmed in our experiments discussed in Sec 1.6 where for a three thousand bus system with one million Monte Carlo runs, less than 20 critical regions cover more than 99%

[3] Widely used in the signal processing community, dictionary learning refers to acquiring a dictionary of signal bases to represent a rich class of signals using words (atoms) in the dictionary [30, 31].

Algorithm 1 Online Dictionary Learning

Input: the mean trajectory $\{\bar{d}_1, \bar{d}_2, \cdots, \bar{d}_T\}$ of load and associated (forecast) distributions $\{\mathcal{F}_1, \mathcal{F}_2, \cdots, \mathcal{F}_T\}$.
Initialization: compute the initial critical region dictionary $\mathcal{C}_0$ from the mean load trajectory.
for $m = 1, \cdots, M$ **do**
  for $t = 1, \cdots, T$ **do**
    Generate sample d_t^m and let $\theta_t^m = (d_t^m, g_{t-1}^m)$.
    Search $\mathcal{C}_{t-1}^m$ for critical region $C(\theta_t^m)$.
    if $C(\theta_t^m) \in \mathcal{C}_{t-1}^m$ **then**
      Compute g_t^m from the affine mapping $g^*_{C(\theta_t^m)}(\theta)$.
    else
      Solve g_t^m from DC-OPF (1.2-1.6) using θ_m^t, compute $C(\theta_t^m)$, and update $\mathcal{C}_t^m = \mathcal{C}_{t-1}^m \cup \{C(\theta_t^m)\}$.
    end if
  end for
end for
Output: the critical region dictionary $\mathcal{C}_T^M$.

of cases. Thus a dictionary of 20 words is mostly adequate for representing the level of randomness in the system.

1.6 Numerical Simulations

We present in this section two sets of simulation results. The first compares the computation cost of the proposed method with that of direct Monte Carlo simulations. To this end, we used the 3210 "Polish network" [32]. The second set of simulations focuses on probabilistic forecasting. With this example, we aim to demonstrate the capability of the proposed method in providing joint and marginal distributions of LMPs and power flows, a useful feature not available in existing forecasting methods.

1.6.1 General setup

We selected the "duck curve" [1] as the expected net load profile as shown in Figure 1.3. We were particularly interested in three scenarios: Scenario 1 represented a time ($T = 55$) when the net load was held steady at the mid range. Scenario 2 ($T = 142$) was when the net load was on a downward ramp due to the increase of solar power. Scenario 3 ($T = 240$) was at a time when the netload was at a sharp upward ramp. The three scenarios represented different operating conditions and different levels of randomness.

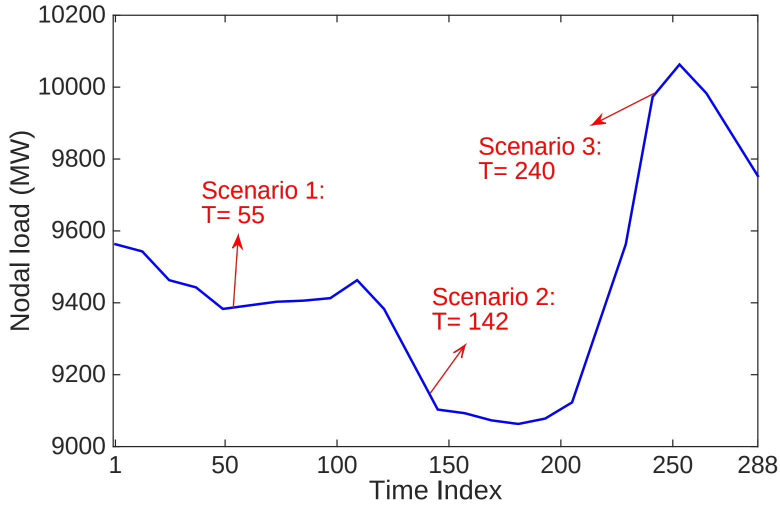

Figure 1.3 The “duck curve” of net load over the different time of the day.

The net load—the load offset by renewable generation—was distributed throughout the network. A renewable generation connected to a bus, say a wind farm, was modeled as a Gaussian random variable $\mathcal{N}(\mu, (\eta\mu)^2)$ with mean μ and standard deviation $\eta\mu$. Similar models were used for load forecasts.

Given a forecasting or simulation horizon T, the real-time economic dispatch model was a sequence of optimizations with one DCOPF in each 5-minute interval. In this model, the benchmark technique solved a series of single period DCOPF models with ramp constraints that coupled the DCOPF solution at time t with that at time $t-1$. Computationally, the simulation was carried out in a Matlab environment with yalmip toolbox and IBM CPLEX on a laptop with an Intel Core i7-4790 at 3.6 GHz and 32 GB memory.

1.6.2 The 3120-bus System

The 3120-bus system (Polish network) defined by MATPOWER [32] was used to compare the computation cost of the proposed method with direct Monte Carlo simulation [13]. The network had 3120 buses, 3693 branches, 505 thermal units, 2277 loads and 30 wind farms. Twenty of the wind farms were located at PV buses and the rest at PQ buses. For the 505 thermal units, each unit had upper and lower generation limits as well as a ramp constraint. Ten transmission lines 1, 2, 5, 6, 7, 8, 9, 21, 36, 37 had capacity limits of 275 MW.

The net load profile used in this simulation was the duck curve over a 24-hour simulation horizon. The total load was at the level of 27,000 MW during morning peak load hours with 10% of renewables distributed across 30 wind farms. One large wind farm had rated capacity of 200 MW, 20 midsize wind farms at the

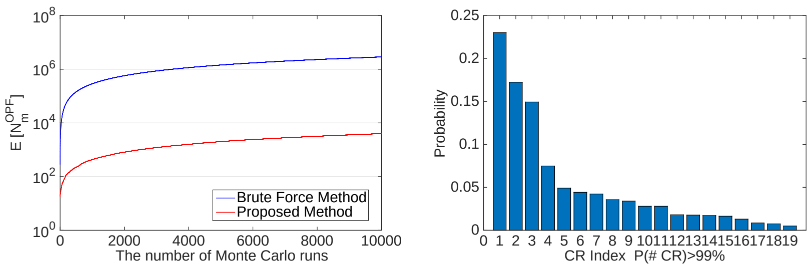

Figure 1.4 Left: The expected number of OPF computations vs. the total number of Monte Carlo simulations. Right: The distribution of the critical regions observed for the proposed method.

rated capacity of 150 MW, and 9 small wind farms at 50-80 MW. Wind farm i produced Gaussian distributed renewable power $\mathcal{N}(\mu_i, (0.03\mu_i)^2)$.

The left panel of Figure 1.4 shows the comparison of the computation cost between the proposed approach and the benchmark technique [13]. The two methods obtained identical forecasts, but ODL had roughly three orders of magnitude reduction in the number of DCOPF computations required in the simulation. This saving came from the fact that only 3989 critical regions appeared in about 2.88 million random parameter samples. In fact, as shown in the right panel of Figure 1.4, 19 out of 3989 critical regions represented 99% of all the observed critical regions.

1.6.3 The IEEE 118-bus System

The performance of the proposed algorithm was tested on the IEEE 118-bus system [32] shown in Figure 1.5. Here the system was partitioned into three subareas. There were 10 capacity constrained transmission lines (labeled blue) at the maximum capacity of 175 MW. The system included 54 thermal units with ramp limits, 186 branches, and 91 load buses. All of which were connected with Gaussian distributed load with standard deviation at the level of $\eta = 0.15\%$ of its mean. The mean trajectory of the net load again followed the "duck curve." Three scenarios were tested, each included 1000 Monte Carlo runs to generate required statistics.

Scenario 1: T=55

The first scenario was $T = 55$ on the duck curve. This was a case when the system operated in a steady load regime where the load did not have a significant change. Figure 1.6 showed some of the distributions obtained by the proposed technique. The top left panel showed the average LMP at all buses where the average LMPs were relatively flat with the largest LMP difference appeared between bus 94 and bus 95. The top right panel showed the joint LMP distribution at bus 95 and 94. It was apparent that the joint distribution of LMP at these two buses was

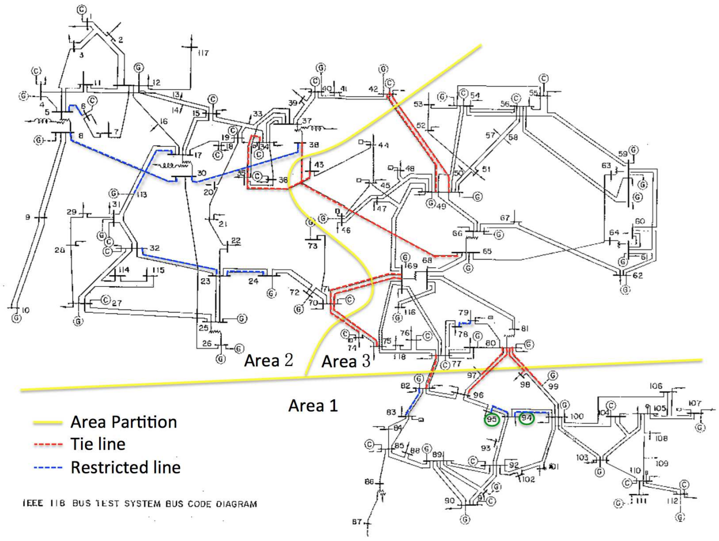

Figure 1.5 The diagram of IEEE 118-bus system. Blue lines are capacity limited. Red lines are tie lines.

concentrated at a single point mass, which corresponded to the case that all realizations of the random demand fell in the same critical region. The bottom left panel showed the power flow distribution at line 147 connecting bus 94-95. As expected, line 147 was congested. The bottom right panel showed the power flow distribution of line 114, which was one of the tie lines connecting areas 2 and 3. The distribution of power flow exhibited a single mode Gaussian-like shape.

Scenario 2: T=142

The second scenario at T=142 involved a downward ramp. This was a case when the load crossed the boundaries of multiple critical regions. In Figure 1.7, the top left panel showed the joint probability distribution of LMP at buses 94-95, indicating that the LMPs at these two buses had two possible realizations, one showing small LMP difference with a high probability, the other a bigger price difference with a low probability. The top right panel showed the power flow distribution on the line connecting bus 94-95. It was apparent that the line was congested with non-zero but relatively small probability, which gave rise to the more significant price difference between these two buses. The bottom panels showed the power flow distributions on tie lines 115 and 153. In both cases, the power flow distribution had three modes, showing little resemblance of Gaussian distributions.

Scenario 3: T=240

The third scenario at T=240 involved a steep up ramp at high load levels. This was also a case when the random load crossed the boundaries of multiple critical

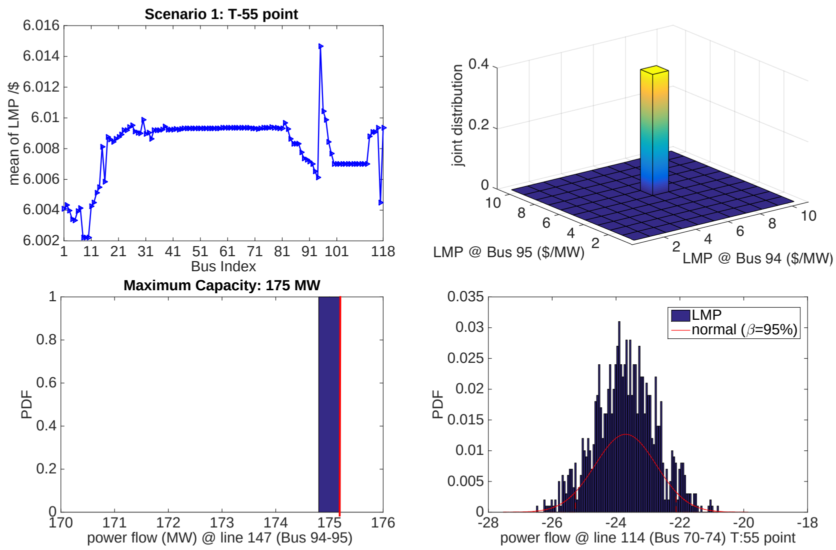

Figure 1.6 Top left: The expected LMPs at all buses. Top right: joint LMP distribution at buses 94-95. Bottom left: power flow distribution on line 147. Bottom right: power flow distribution on line 114.

regions. In Figure 1.8, the top left panel indicated 4 possible LMP realizations at buses 94-95. With probability near half that the LMPs across buses 94-95 had a significant disparity and the other half the LMPs on these two buses roughly the same. The power flow on tie line 152 had a Gaussian-like distribution shown in the top right panel whereas tie line 128 had a power flow distribution spread in four different levels shown in the bottom left panel. It is especially worthy of pointing out, from the bottom right panel, that the power flow on line 66 had opposite directions.

1.7 Conclusion

We present in this paper a new methodology of online probabilistic forecasting and simulation of the electricity market. The main innovation is the use of online dictionary learning to obtain the solution structure of parametric DCOPF sequentially. The resulting benefits are the significant reduction of computation costs and the ability to adapt to changing operating conditions. Numerical simulations show that, although the total number of critical regions associated with the parametric DCOPF is very large, only a tiny fraction of critical regions appear in a large number of Monte Carlo runs. These insights highlight the potential of further reducing both computation costs and storage requirements.

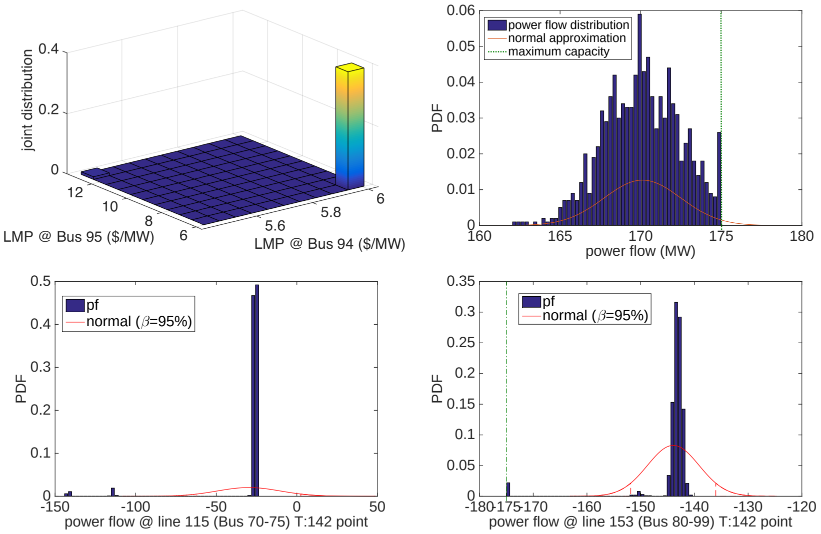

Figure 1.7 Top left: joint LMP distribution at buses 94-95. Top right: power flow distribution on line 147. Bottom left: power flow distribution on line 115. Bottom right: power flow distribution on line 153.

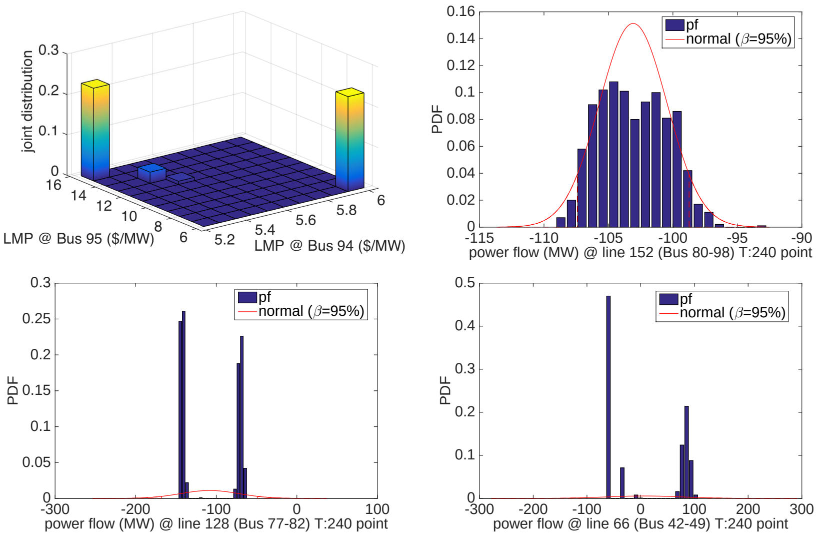

Figure 1.8 Top left: joint LMP distribution at buses 94-95. Top right: power flow distribution on line 152. Bottom left: power flow distribution on line 128. Bottom right: power flow distribution on line 152.

References

[1] California Independent System Operator , “What the duck curve tells us about managing a green grid,” https://www.caiso.com/Documents/FlexibleResourcesHelpRenewables_FastFacts.pdf , accessed: 2019-03-13.

[2] Electric Reliability Council of Texas, “Ercot launches wholesale pricing forecast tool,” http://www.ercot.com/news/press_releases/show/26244, accessed: 2019-03-13.

[3] Alberta Electric System Operator, “Price, supply and demand information,” https://www.aeso.ca/market/market-and-system-reporting/, accessed: 2019-03-13.

[4] Australian Energy Market Operator, “Data dashboard: Forecast spot price (pre-dispatch),” https://aemo.com.au/Electricity/National-Electricity-Market-NEM/Data-dashboard, accessed: 2019-03-13.

[5] Q. P. Zheng, J. Wang, and A. L. Liu, “Stochastic optimization for unit commitmenta review,” *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 1913–1924, 2015.

[6] Independent Market Operator , “Wholesale electricity market design summary,” https://www.aemo.com.au/-/media/Files/PDF/wem-design-summary-v1-4-24-october-2012.pdf, 2012, accessed: 2019-03-13.

[7] G. Li, J. Lawarree, and C.-C. Liu, “State-of-the-art of electricity price forecasting in a grid environment,” in *Handbook of power systems II*. Springer, 2010, pp. 161–187.

[8] R. Weron, “Electricity price forecasting: A review of the state-of-the-art with a look into the future,” *International Journal of Forecasting*, vol. 30, no. 4, pp. 1030–1081, 2014.

[9] R. Bo and F. Li, “Probabilistic LMP forecasting considering load uncertainty,” *IEEE Transactions on Power Systems*, vol. 24, no. 3, pp. 1279–1289, 2009.

[10] A. Rahimi Nezhad, G. Mokhtari, M. Davari, S. Hosseinian, and G. Gharehpetian, “A new high accuracy method for calculation of LMP as a random variable,” in *International Conference on Electric Power and Energy Conversion Systems*, 2009, pp. 1–5.

[11] M. Davari, F. Toorani, H. Nafisi, M. Abedi, and G. Gharepatian, “Determination of mean and variance of LMP using probabilistic DCOPF and T-PEM,” *PECon08*, pp. 1280–1283, 2008.

[12] J. Bastian, J. Zhu, V. Banunarayanan, and R. Mukerji, "Forecasting energy prices in a competitive market," *IEEE computer applications in power*, vol. 12, no. 3, pp. 40–45, 1999.

[13] L. Min, S. T. Lee, P. Zhang, V. Rose, and J. Cole, "Short-term probabilistic transmission congestion forecasting," in *Proc. of the 3rd International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*, 2008, pp. 764–770.

[14] Y. Ji, J. Kim, R. J. Thomas, and L. Tong, "Forecasting real-time locational marginal price: A state space approach," in *Proc. of the 47th Asilomar Conference on Signals, Systems, and Computers*, 2013, pp. 379–383.

[15] Y. Ji, R. J. Thomas, and L. Tong, "Probabilistic forecast of real-time LMP via multiparametric programming," in *Proc. of the 48th Hawaii International Conference on System Sciences*, 2015, pp. 2549 – 2556.

[16] ——, "Probabilistic forecasting of real-time lmp and network congestion," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 831–841, 2017.

[17] W. Deng, Y. Ji, and L. Tong, "Probabilistic forecasting and simulation of electricity markets via online dictionary learning," in *HICSS*, 2017.

[18] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *International Journal of Forecasting*, vol. 32, no. 3, pp. 914–938, 2016.

[19] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renewable and Sustainable Energy Reviews*, vol. 32, pp. 255–270, 2014.

[20] D. W. Van der Meer, J. Widén, and J. Munkhammar, "Review on probabilistic forecasting of photovoltaic power production and electricity consumption," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 1484–1512, 2018.

[21] Y. Guo and L. Tong, "Pricing multi-period dispatch under uncertainty," in *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2018, pp. 341–345.

[22] California Independent System Operator , "Flexible ramping product," http://www.caiso.com/informed/Pages/StakeholderProcesses/CompletedClosedStakeholderInitiatives/FlexibleRampingProduct.aspx, accessed: 2019-03-13.

[23] N. Navid and G. Rosenwald, "Ramp capability product design for miso markets," *Market Development and Analysis*, 2013.

[24] C. Wu, G. Hug, and S. Kar, "Risk-limiting economic dispatch for electricity markets with flexible ramping products," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 1990–2003, 2016.

[25] L. Xu and D. Tretheway, "Flexible ramping products," *CAISO Proposal*, 2012.

[26] T. Zheng and E. Litvinov, "Ex post pricing in the co-optimized energy and reserve market," *IEEE Transactions on Power Systems*, vol. 21, no. 4, pp. 1528–1538, 2006.

[27] T. Gal and J. Nedoma, "Multiparametric linear programming," *Management Science*, vol. 18, pp. 406–442, 1972.

[28] A. Bemporad, M. Morari, V. Dua, and E. N. Pistikopoulos, "The explicit linear quadratic regulator for constrained systems," *Journal of Automatica*, vol. 38, no. 1, pp. 3–20, 2002.

[29] F. Borrelli, A. Bemporad, and M. Morari, *Predictive control for linear and hybrid systems*. Cambridge University Press, 2017.

[30] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T.-W. Lee, and T. J. Sejnowski, "Dictionary learning algorithms for sparse representation," *Neural computation*, vol. 15, no. 2, pp. 349–396, 2003.

[31] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, 2010.

[32] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, 2011, http://www.pserc.cornell.edu//matpower/.