

MPC for power systems dispatch based on stochastic optimization

Ion Necoara* Dragos N. Clipici* Panagiotis Patrinos**
Alberto Bemporad**

* *University Politehnica Bucharest, Automatic Control and Systems
Engineering Department, 060042, Bucharest, Romania,*
{ion.necoara,dragos.clipici}@acse.pub.ro

** *IMT Institute for Advanced Studies Lucca, Piazza San Ponziano 6,
55100 Lucca, Italy,* {panagiotis.patrinos, alberto.bemporad}@imtlucca.it

Abstract: In this paper we investigate the problem of optimal real-time power dispatch of an interconnection of conventional power generation plants, renewable resources and energy storage systems. The objective is to minimize imbalance costs and maximize profits whilst satisfying user demand. The managing company is able to trade energy on an electricity market. Energy prices, demand and renewable generation are considered stochastic processes. We show that under certain assumptions, the stochastic power dispatch problem over a finite horizon can be recast into a stochastic optimization formulation but with deterministic constraints. We carry out a systematic study of stochastic optimization methods to solve this problem. We also show that this problem can be approximated by a proper deterministic optimization problem using the sample average approximation method, which can then be solved by standard means.

1. INTRODUCTION

Due to the current trend of deregularizing electricity markets, there is a ever increasing need for efficiency in power production and distribution. The current solution to these problems, from the energy supplier's side, is the integration in the current power systems of renewable energy generation sources and power storage systems. Integrating renewables is a major bottleneck due to high variability and low predictability of renewable resources.

In the majority of electricity markets, market participants (called Balance Responsible Partners (BRPs)) place their bids on the *Day-Ahead* (DA) market and the *Ancillary Services* (AS) market regarding energy delivery and capacity availability for each *Program Time Unit* (PTU) of the following day. At the end of the day-ahead auction, the *Transmission System Operator* (TSO) selects the accepted and rejected bids according to some clearing mechanism [Shahidehpour et al., 2002] and publishes the future prices and volumes, for each PTU of the following day. Subsequently, each BRP determines its *Energy Program* (E-Program) and sends it to the TSO for approval. The latter describes the amount of energy supplied or consumed by the BRP at every *Program Time Unit* (PTU) of the next day. After receiving approval by the TSO, each BRP executes a unit-commitment module to determine the on/off status of its generators for each PTU of the following day. Due to uncertainties in power demand and generation, the existence of an *imbalance market* (or *real-time market*) operated by the TSO is mandatory in order to counteract real-time global energy imbalances, [Jokic, 2007]. Unlike day-ahead prices, imbalance prices are characterized by large volatility, sudden spikes that are hard to predict, and counterintuitive phenomena like negative values.

A BRP must fulfill its E-Program and concurrently cope with uncertainties induced by renewables intermittent generation, time-varying loads and imbalance prices, as well as perturbations of its E-Program due to ancillary services (AS) bids activated by the TSO. If the TSO calls for a specific AS bid, the BRP responsible for this bid is asked to deliver the requested energy, by adjusting its E-Program accordingly. The real-time operation of a BRP is a complicated optimal control problem, due to the coupling between PTUs, uncertainties and complex dynamics of the controllable generators. The controller must determine power set-points for the controllable generators which minimize generation costs (utilize intermittent resources as much as possible) and economically track the deviated E-Program, meaning that it may be profitable sometimes to deviate either upwards or downwards depending on the imbalance price, [Patrinos et al., 2012].

In this paper we formulate the model predictive problem for power systems dispatch into a stochastic optimization problem where the power balance constraint needs to hold almost surely. We show that the stochastic constraints of the original optimization problem, through a linear input feedback form and under certain assumptions regarding the disturbance sets, can be replaced by properly structured deterministic ones. With properly formulated constraints, we show that the MPC problem with the objective function in expectation can be approximated by a deterministic optimization problem using sample average approximation method, which is solvable by standard means. We also show that certain stochastic optimization strategies are applicable to the problem and permit us to solve it more directly, e.g. stochastic gradient algorithm.

The outline of the paper is as follows. In Section 2 we analyze the power dispatch problem. In Section 3 we formulate

the corresponding stochastic MPC problem. In Section 4 we show, under certain assumptions, that the stochastic constraints of the MPC problem can be reformulated into deterministic ones. In Section 5 we address the problem of the expected value objective function for the MPC.

2. OPTIMAL POWER DISPATCH PROBLEM

We consider a system formed by n_p conventional power plants (e.g. coal, oil or gas), n_r renewable plants (e.g. wind, solar), and n_s storage sites used to store excess energy (e.g. reservoirs with upstream pumping). We denote by $p_i(k)$ the average power output over one PTU of the conventional generator i , where $i = 1, \dots, n_p$. Each power plant will of course have its own physical limitations on the average power produced over one PTU, which are expressed as the constraints:

$$p_i^{lb} \leq p_i(k) \leq p_i^{ub}. \quad (1)$$

The conventional power plants are limited on how quickly production can be changed (ramp rate limitations):

$$\Delta_{p_i}^{lb} \leq p_i(k+1) - p_i(k) \leq \Delta_{p_i}^{ub}. \quad (2)$$

The n_s storage sites of the power system are the dynamic components of the system. We denote by x_i the state of the storage site i , which represents the amount of energy currently stored on site (state-of-charge, (SOC)). We consider that a storage site i has the following dynamics:

$$x_i(k+1) = \alpha_i x_i(k) + T(\alpha_i^c u_i^c(k) - \alpha_i^d u_i^d(k)), \quad (3)$$

where T is the length of the PTU, $\alpha_i \leq 1$ is a loss parameter for the stored energy, $u_i^c(k)$ and $u_i^d(k)$ are inputs, which represent the amount of power to be stored and the amount of energy to be redistributed back into the grid, while $\alpha_i^c \leq 1$ and $\alpha_i^d \leq 1$ are conversion loss parameters. The storage units will of course have their own physical limitations as follows:

$$u_i^{c,min} \leq u_i^c(k) \leq u_i^{c,max}, u_i^{d,min} \leq u_i^d(k) \leq u_i^{d,max} \quad (4)$$

$$x_i^{lb} \leq x_i(k) \leq x_i^{ub}, \Delta_{x_i}^{lb} \leq x_i(k) - x_i(k-1) \leq \Delta_{x_i}^{ub}. \quad (5)$$

The constraints in (4) represent limitations on the amount of energy which can be stored or redistributed from the site i in one PTU, while those in (5) represent limitations on the total amount of energy that can be stored and on how much the stored energy can fluctuate from one PTU to the other. At every step k , the amount of energy produced, minus the amount of energy stored and traded with other markets has to meet local customer demand. This requirement is known as a power balance requirement:

$$\sum_{i=1}^{n_p} p_i(k) + \sum_{i=1}^{n_s} (u_i^d(k) - u_i^c(k)) + \sum_{i=1}^{n_r} r_i(k) - p^{ex}(k) = d(k), \quad (6)$$

where $r_i(k)$ represents the power output of the renewable generators, $p^{ex}(k)$ is the power imbalance, i.e. the amount of power traded to electricity markets and $d(k)$ is the local energy demand. Here we assume that the local demand includes a deterministic part that is the energy required by the E-Program determined on the day-ahead, plus a stochastic part that is induced by the Ancillary Services requests activated by the TSO and the discrepancy between predicted and actual load due to imperfect demand forecasting. Note that this constraint can always be deemed feasible, due to the possibility of $p^{ex}(k)$ being negative, i.e. we can always purchase more energy from the TSO if the current production does not meet the demand. A similar

modeling approach for the power system can be found in [Patrinos et al., 2011, Xie and Ilic, 2008, Elaiw et al., 2012].

3. STOCHASTIC MPC PROBLEM FORMULATION

The objective of the MPC problem is to obtain a schedule of $p_i(k)$ production rates (set points) for the conventional power plant in order to maximize profits and meet customer demand. The non-renewable n_p conventional power plants are the most expensive to operate. We consider a quadratic operation cost for a conventional power plant i , $l_i(p_i(k)) = Q_i p_i(k)^2 + q_i p_i(k)$, with $Q_i > 0$. Thus, the objective of the problem at each moment k is to minimize:

$$\sum_{i=1}^{n_p} l_i(p_i(k)) - T\lambda(k)p^{ex}(k), \quad (7)$$

where $\lambda(k)$ is the imbalance energy price at time k . We denote by $p(k)$ the vector of $p_i(k)$, and we also denote $u_{cd}(k) = [u_1^c(k) u_1^d(k) \dots u_{n_s}^c(k) u_{n_s}^d(k)]^T$, such that the input vector for the power system at time k is $u(k) = [p(k)^T u_{cd}^T(k) p^{ex}(k)]^T$, with $u(k) \in \mathbb{R}^{n_u}$, $n_u = n_p + 2n_s + 1$. Given that l_i are quadratic costs in the scalars $p_i(k)$, we can express their sum as $l(u(k)) = u(k)^T Q u(k) + q^T u(k)$, with $Q \in \mathbb{R}^{n_u \times n_u}$, $q \in \mathbb{R}^{n_u}$, $Q = \text{diag}(Q_i, 0_{n_u - n_p \times n_u - n_p})$, Q positive semidefinite and $q = [q_1 \dots q_{n_s} 0 \dots 0]^T$. We now outline the fact that the outputs of the renewable plants $r_i(k)$, the customer demand $d(k)$ and the electricity price $\lambda(k)$ are stochastic variables and, from a control perspective, are considered stochastic disturbances for the power system. We thus gather them into a disturbance vector $w(k)$, $w(k) = [r(k)^T d(k) \lambda(k)]^T$, where $r(k) = [r_1(k) \dots r_{n_r}(k)]^T$ and $w(k) \in \mathbb{R}^{n_w}$, with $n_w = n_r + 2$. Given the vectors $u(k)$ and $w(k)$, the term representing traded energy revenue in the cost (7) is:

$$T\lambda(k)p^{ex}(k) = w(k)^T H_{wu} u(k),$$

where $H_{wu} = T e_{n_w} e_{n_u}^T$. Given the individual dynamics (3), the dynamics for the entire state $x(k) = [x_1(k) \dots x_{n_s}(k)]^T$ can be written as $x(k+1) = Ax(k) + Bu(k)$, with $A \in \mathbb{R}^{n_s \times n_s}$, $A = \text{diag}(\alpha_i)$, and $B \in \mathbb{R}^{n_s \times n_u}$ formed from T , α_i^c and α_i^d . In the following, we will consider an MPC problem with a prediction horizon of length N . Now, if we denote the input vector over a prediction horizon of length N as $\bar{u} = [u(1)^T \dots u(N)^T]^T$, where $\bar{u} \in \mathbb{R}^{n_{\bar{u}}}$ with $n_{\bar{u}} = N(n_p + 2n_s + 1)$, we can write the production cost for the conventional generators as:

$$\bar{V}(\bar{u}) = \sum_{k=0}^{N-1} l(u(k)) = \bar{u}^T \bar{Q} \bar{u} + \bar{q}^T \bar{u},$$

where \bar{Q} is block-diagonal with block components Q and $\bar{q} = \mathbf{1}_N \otimes q$. If we also denote the disturbance vector for the entire prediction horizon as $\bar{w} = [w(1)^T \dots w(N)^T]^T$, then the traded energy term for the entire prediction horizon will be $\bar{w}^T H_{\bar{w}\bar{u}} \bar{u}$, where $H_{\bar{w}\bar{u}} = I_N \otimes H_{wu}$. Given that $Q_i > 0$, we have $\bar{Q} \succeq 0$ and $V(\bar{u}, \bar{w}) = \bar{V}(\bar{u}) + \bar{w}^T H_{\bar{w}\bar{u}} \bar{u}$, is convex in \bar{u} , but not strongly convex. We now turn our attention to formulating the constraints for the problem over the entire prediction horizon. The power balance equation (6) can be written as $c_{uw}^T u(k) - d_{uw}^T w(k) = 0$, where c_{uw} and d_{uw} are vectors of appropriate form. Thus, for the

¹ In this paper, by \otimes we denote the standard Kronecker product.

entire prediction horizon, the power balance constraint (6) is expressed as:

$$C_{uw}\bar{u} - D_{uw}\bar{w} = 0, \quad (8)$$

where $C_{uw} = I_N \otimes c_{uw}$ and $D_{uw} = I_N \otimes d_{uw}$. Note that this constraint has to be satisfied almost surely.

If we now denote $\bar{x} = [x(1)^T \dots x(N)^T]^T$, $\bar{x} \in \mathbb{R}^{N_s}$ as the state vector for the entire prediction horizon, the dynamics for the entire prediction horizon would imply a constraint:

$$\bar{x} = \bar{A}\bar{B}\bar{u} + c_{x(0)}, \quad (9)$$

where $\bar{A}\bar{B}$ and $c_{x(0)}$ are of appropriate form. Note that constraints (5) for the entire prediction horizon are written as $\bar{x} \in \mathbf{X} = \{\bar{x} : C_{\bar{x}}\bar{x} \leq d_{\bar{x}}\}$, where $C_{\bar{x}} \in \mathbb{R}^{m_{\bar{x}} \times n_{\bar{x}}}$, with $m_{\bar{x}} = 4n_{\bar{x}} = 4Nn_s$, and $d_{\bar{x}} \in \mathbb{R}^{m_{\bar{x}}}$. Constraints for the input over the entire prediction horizon are also written in the same manner, $\bar{u} \in \mathbf{U}$, $\mathbf{U} = \{C_{\bar{u}}\bar{u} \leq d_{\bar{u}}\}$, where $C_{\bar{u}} \in \mathbb{R}^{m_{\bar{u}} \times n_{\bar{u}}}$, with $m_{\bar{u}} = (4N-2)n_p + 4Nn_s$, and $d_{\bar{u}} \in \mathbb{R}^{m_{\bar{u}}}$. Now, given the fact that the cost is not a function of \bar{x} , we can eliminate the states through (9) and arrive at the compact concatenated form of the constraints:

$$C_{\bar{u}\bar{x}}\bar{u} \leq d_{\bar{u}\bar{x}}, C_{\bar{u}\bar{x}} = \begin{bmatrix} C_{\bar{u}} \\ C_{\bar{x}}\bar{A}\bar{B} \end{bmatrix}, d_{\bar{u}\bar{x}} = \begin{bmatrix} d_{\bar{u}} \\ d_{\bar{x}} - C_{\bar{x}}c_{x(0)} \end{bmatrix},$$

and $C_{\bar{u}\bar{x}} \in \mathbb{R}^{(m_{\bar{x}}+m_{\bar{u}}) \times n_{\bar{u}}}$, with $m_{\bar{x}} + m_{\bar{u}} = 8Nn_s + (4N-2)n_p$. We consider that the stochastic disturbance $w(k)$ has a finite polyhedral support set W , such that over the entire prediction horizon we have $\bar{w} \in \mathbf{W} = W \times W \times \dots \times W$. We now formulate the stochastic MPC problem over the entire prediction horizon:

$$\min_{\bar{u}} \mathbf{E}[V(\bar{u}, \bar{w})]$$

$$\text{s.t.: } C_{\bar{u}\bar{x}}\bar{u} \leq d_{\bar{u}\bar{x}}, C_{uw}\bar{u} - D_{uw}\bar{w} = 0 \quad \forall \bar{w} \in \mathbf{W}.$$

Given the disturbance $w(k)$, the standard approach would be to find causal disturbance feedback policies $u(k) = \Phi_k(x(0), w(0), \dots, w(k-1))$ that would minimize the cost. However, this would imply an infinite dimensional problem since the variables would be the functions $\Phi_k(\cdot)$. This motivates us to restrict the policies Φ_k to a finite dimensional subspace, i.e. we choose an input of linear disturbance feedback form $u(k) = Kw(k)$, which is encouraged by the fact that the constraint (8) needs to be satisfied a.s. and implies that \bar{u} is linearly dependent on \bar{w} .

Remark 1. We may also consider a time-varying linear disturbance feedback of the form $u(k) = K(k)w(k)$.

Given that the constraint $C\bar{u} - D\bar{w} = 0$ is equivalent to $(K^T c_{uw} - d_{uw})^T w(k) = 0 \quad \forall w(k) \in W, k \geq 0$, and under the assumption that W has a nonempty interior, it implies that $K^T c_{uw} - d_{uw} = 0$. For compactness of exposition, we denote $\bar{K} = I_N \otimes K$ such that $\bar{u} = \bar{K}\bar{w}$, and we arrive at the MPC problem:

$$\begin{aligned} \min_K \quad & \mathbf{E}[\bar{w}^T \bar{K}^T \bar{Q} \bar{K} \bar{w} + (\bar{q} + H_{\bar{w}\bar{u}}^T \bar{w})^T \bar{K} \bar{w}] \\ \text{s.t.:} \quad & K^T c_{uw} = d_{uw}, C_{\bar{u}\bar{x}} \bar{K} \bar{w} \leq d_{\bar{u}\bar{x}} \quad \forall \bar{w} \in \mathbf{W}. \end{aligned} \quad (10)$$

The objective function of (10) can be rewritten such that the resulting optimization problem is:

$$\begin{aligned} \min_K \quad & \text{Tr}(\bar{K}^T \bar{Q} \bar{K} \Sigma_{\bar{w}} + H_{\bar{w}\bar{u}} \bar{K} \Sigma_{\bar{w}}) + m_{\bar{w}}^T \bar{K}^T \bar{Q} \bar{K} m_{\bar{w}} \\ & + (H_{\bar{w}\bar{u}}^T m_{\bar{w}} + \bar{q})^T \bar{K} m_{\bar{w}} \\ \text{s.t.:} \quad & K^T c_{uw} = d_{uw}, C_{\bar{u}\bar{x}} \bar{K} \bar{w} \leq d_{\bar{u}\bar{x}} \quad \forall \bar{w} \in \mathbf{W}. \end{aligned} \quad (11)$$

where $m_{\bar{w}} = \mathbf{E}[\bar{w}]$ and $\Sigma_{\bar{w}} = \text{Cov}[\bar{w}] = \mathbf{E}[(\bar{w} - m_{\bar{w}})(\bar{w} - m_{\bar{w}})^T]$. There are two main issues regarding this problem.

First of all, the cost is in expectation, and we investigate under which conditions applied to \bar{w} we can transform it into a convex deterministic optimization problem. Secondly, if the variable \bar{w} would have an infinite support set \mathbf{W} , then we would get an infinite number of constraints. In the following two sections we will address these issues.

4. CONSTRAINTS REFORMULATION

We now examine the structure of the constraints of problem (11). After a few simple calculations, it can be observed that the constraint $C_{\bar{x}}\bar{A}\bar{B}\bar{K}\bar{w} \leq d_{\bar{x}} - C_{\bar{x}}c_{x(0)}$ consists of the following:

$$\begin{aligned} \sum_{t=0}^k A^{k-t} BKw(t) &\leq x^{ub} - A^k x(0) \\ \sum_{t=0}^k -A^{k-t} BKw(t) &\leq -x^{lb} + A^k x(0) \\ \sum_{t=1}^k A^{k-t}(A-I)BKw(t-1) + BKw(k) &\leq \Delta_x^{ub} - \delta_{x(0)}^k \\ \sum_{t=1}^k A^{k-t}(I-A)BKw(t-1) - BKw(k) &\leq \delta_{x(0)}^k - \Delta_x^{lb}, \end{aligned} \quad (12)$$

where $k = 0, \dots, N-1$, $\delta_{x(0)}^k = A^k(A-I)x(0)$. These categories of constraints are obtained from the constraints in (5) together with constraint elimination via (9). These add up to $4N$ linear constraints in K . Now, given constraints on \bar{u} , we note that $C_{\bar{u}}$ is formed from matrices:

$$C_{np} = [I_{np} \ 0_{np \times 2ns+1}], C_{cd} = [0_{2ns \times np} \ I_{2ns} \ 0_{2ns \times 1}].$$

Thus, the constraint $C_{\bar{u}}\bar{K}\bar{w} \leq d_{\bar{u}}$ comprises the following constraints:

$$C_{np}Kw(k) \leq p^{ub}, -C_{np}Kw(k) \leq -p^{lb}, \quad (13)$$

$$C_{cd}Kw(k) \leq u_{cd}^{ub}, -C_{cd}Kw(k) \leq -u_{cd}^{lb}, \quad (14)$$

for $k = 0, \dots, N-1$ and

$$\begin{cases} C_{np}K(w(k) - w(k-1)) \leq \Delta p^{ub} \\ C_{np}K(w(k-1) - w(k)) \leq -\Delta p^{lb}, \end{cases} \quad (15)$$

for $k = 1, \dots, N-1$. Constraints of type (13)-(14) are obtained from (1) and (4), while constraints (15) are obtained from (2). These constraints add up to $6N-2$ constraints in the variable K , with a total of $10N-2$ constraints in K for problem (11). Now, given that the constraint sets \mathbf{X} and \mathbf{U} are formed by box constraints, they are compact. If we define the constraint set

$$\mathcal{K} = \{\bar{K} : (12) - (15) \text{ satisfied } \forall \bar{w} \in \mathbf{W}\},$$

we find that it is also compact. In the following we will examine two options regarding the form of the set W , either expressed as a convex hull of vertices, or as a multivariate linear inequality.

4.1 Reformulation for convex hull description of W

We first consider W expressed as the convex hull of a number of n_v vertices w_i , i.e. $W = \text{conv}\{w_1, \dots, w_{n_v}\}$, such that any $w(t) \in W$ can be expressed as $w(t) = \sum_{i=1}^{n_v} \alpha_i(t)w_i$, with $0 \leq \alpha_i(t) \leq 1$ and $\sum_{i=1}^{n_v} \alpha_i(t) = 1$. Consequently, the set \mathbf{W} is also a convex hull of vertices \bar{w}^i , with $i = 1, \dots, n_v^N$ formed from all possible combinations of vertices from W over the prediction horizon N .

Thus, the compact form of the stochastic MPC problem over the prediction horizon N is:

$$\begin{aligned} \min_K \quad & \text{Tr}(\bar{K}^T \bar{Q} \bar{K} \Sigma_{\bar{w}} + H_{\bar{w}\bar{u}} \bar{K} \Sigma_{\bar{w}}) + m_{\bar{w}}^T \bar{K}^T \bar{Q} \bar{K} m_{\bar{w}} \\ & + (H_{\bar{w}\bar{u}}^T m_{\bar{w}} + \bar{q})^T \bar{K} m_{\bar{w}} \\ \text{s.t.}: \quad & K^T c_{uw} = d_{uw}, \quad C_{\bar{u}\bar{x}} \bar{K} \bar{w}_i \leq d_{\bar{u}\bar{x}}, \quad i = 1, \dots, n_{\bar{v}}^N. \end{aligned} \quad (16)$$

Now, given that \bar{w} is in a convex hull of vertices, this implies that constraints (12) to (15) need to be satisfied in all the vertices of \mathbf{W} . We thus get a total number of $(10N - 2)n_{\bar{v}}^N$ constraints in K . However, given the structure of matrices $C_{\bar{u}}$ and $C_{\bar{x}}$, we get $4 \frac{n_{\bar{v}}^{N+1} - n_{\bar{v}}}{n_{\bar{v}} - 1} + 2(N-1)n_{\bar{v}}^2 + 4Nn_{\bar{v}}$ constraints in K . Note that in this case we obtain a number of constraints which is exponential in the length of the prediction horizon N .

4.2 Reformulation for halfspace description of W

We now assume that W is expressed as a system of linear inequalities, i.e., $W = \{w : C_w w \leq d_w\}$, with $C_w \in \mathbb{R}^{m_w \times n_w}$, $d_w \in \mathbb{R}^{m_w}$. For the stochastic disturbance vector over the entire prediction horizon we consider the disturbance set as: $\mathbf{W} = \{\bar{w} : C_{\bar{w}} \bar{w} \leq d_{\bar{w}}\}$, with $C_{\bar{w}} = I_N \otimes C_w$, $d_{\bar{w}} = \mathbf{1}_N \otimes d_w$. Each one of the constraints (12) to (15) can be expressed as

$$\sum_{t=l}^k c_j^t K w(t) \leq d_j \quad \forall w(t) \in W,$$

where l and k take values from 0 to $N - 1$, depending on the constraint, while c_j^t are the rows of the matrices $A^{k-t}B$, $A^{k-t}(A - I)B$, C_{np} or C_{cd} and d_j are components of $d_{\bar{u}\bar{x}}$. These constraints could be equivalently written as:

$$\begin{aligned} \max_{w(t)} \quad & \sum_{t=l}^k w(t)^T K^T (c_j^t)^T \leq d_j \\ \text{s.t.} \quad & C_w w(t) \leq d_w. \end{aligned}$$

If we denote $\bar{w}^k = [w(l)^T, w(l+1), \dots, w(k)^T]^T$, then the previous constraint can be written as:

$$\begin{aligned} \max_{\bar{w}^k} \quad & c_j^l \dots c_j^k (I_k \otimes K) \bar{w}^k \leq d_j \\ \text{s.t.}: \quad & (I_k \otimes C_w) \bar{w}^k \leq \mathbf{1}_k \otimes d_w. \end{aligned}$$

If we denote $c_j = [c_j^l \dots c_j^k]^T$, we form the dual:

$$\min_{\lambda_j \geq 0} \max_{\bar{w}^k} c_j^T (I_k \otimes K) \bar{w}^k + \lambda_j^T ((I_k \otimes C_w) \bar{w}^k - \mathbf{1}_k \otimes d_w) \leq d_j.$$

Given the fact that this problem is unconstrained in the variable \bar{w}^k , it is unbounded if the linear term is nonzero. We thus get the following equivalent constraints:

$\lambda_j \geq 0$, $-\lambda_j^T (\mathbf{1}_k \otimes d_w) \leq d_j$, $(I_k \otimes K^T) c_j = -(I_k \otimes C_w^T) \lambda_j$, where λ_j are Lagrange multipliers. The multipliers λ_j can be further divided into $k - l$ subcomponents λ_j^t and the previous constraints are rewritten as:

$$\lambda_j^t \geq 0, \quad K^T c_j^t = -C_w^T \lambda_j^t, \quad t = l, \dots, k, \quad \sum_{t=l}^k (\lambda_j^t)^T d_w \leq d_j.$$

Given that $C_{\bar{x}} \bar{A} \bar{B}$ has $4N$ block rows where each of the blocks has n_s rows, the number of additional variables λ_j^t of dimension m_w is $4n_s \frac{N(N+1)}{2}$, while from $C_{\bar{u}}$ we get an additional $6Nn_p + 4Nn_s$ variables. We thus get a total number of $4 \frac{N(N+3)}{2} n_s + 6n_p N$ constraints of type

$K^T c_j^i = -C_w^T \lambda_j^i$, the same number for $\lambda_j^T \geq 0$, while for the constraints $\sum_{t=l}^k (\lambda_j^t)^T d_w \leq d_j$ we get a total number of $8Nn_s + (4N - 2)n_p$. The value that l and k take depends on the type of constraints, i.e. on which $w(k)$ of \bar{w} are included in the constraint. Thus, if row j belongs to constraints of type (12) we have $l = 0$ and $k = 0, \dots, N - 1$. For constraints of type (13) we have $k = 0, \dots, N - 1$ and $l = k$ since we have only one $w(k)$ in the constraints. Lastly, for constraints (15) we have $k = 1, \dots, N - 1$ and $l = k - 1$ since we have $w(k)$ and $w(k - 1)$ in the constraint. Thus, by this form of W , we get the following optimization problem:

$$\begin{aligned} \min_{K, \lambda_j^t} \quad & \text{Tr}(\bar{K}^T \bar{Q} \bar{K} \Sigma_{\bar{w}} + H_{\bar{w}\bar{u}} \bar{K} \Sigma_{\bar{w}}) + m_{\bar{w}}^T \bar{K}^T \bar{Q} \bar{K} m_{\bar{w}} \\ & + (H_{\bar{w}\bar{u}}^T m_{\bar{w}} + \bar{q})^T \bar{K} m_{\bar{w}} \\ \text{s.t.}: \quad & K^T c_{uw} = d_{uw}, \quad \lambda_j^t \geq 0, \quad K^T c_j^t = -C_w^T \lambda_j^t, \end{aligned} \quad (17)$$

$$\sum_{t=l}^k (\lambda_j^t)^T d_w \leq d_j, \quad t = l, \dots, k, \quad j \in (12), (13), (15)$$

5. OBJECTIVE FUNCTION REFORMULATION

In the previous section we established problems (16) and (17) with a finite number of constraints and we now address their objective functions. The most straightforward approach would be to estimate $m_{\bar{w}}$ and $\Sigma_{\bar{w}}$, e.g. based on the history of \bar{w} or approximate them via empirical averages. Then, problems (16) and (17) are transformed into convex deterministic problems and can be solved efficiently via standard optimization software. On the other hand, if these terms cannot be estimated, then we proceed to analyze different methods of solving problems (16) and (17). It can be observed that problems (16) and (17) can be cast as follows:

$$f^* = \min_{x \in X} f(x) \quad (= \mathbf{E}[F(x, \xi)]), \quad (18)$$

where $x = \bar{K}$, $X = \mathcal{K}$, $\xi = \bar{w}$ and $F(x, \xi) = V(\bar{K}, \bar{w})$.

5.1 Sample average approximation (SAA) method

A well known strategy for solving problem (18) is the Sample Average Approximation (SAA) strategy [Kleywegt et al., 2002, Linderoth et al., 2006], from which we obtain deterministic convex problems of the form:

$$\min_{x \in X} \hat{f}_{N_{SAA}}(x) \quad \left(= \frac{1}{N_{SAA}} \sum_{i=1}^{N_{SAA}} F(x, \xi_i) \right), \quad (19)$$

where $\hat{f}_{N_{SAA}}$ is an approximate objective function to f by taking N_{SAA} samples ξ_i . In our case, given that $F(x, \xi)$ is convex quadratic in x and that X is a convex set, then problem (19) is a convex quadratic optimization problem and can be solved by standard means. By taking an appropriate number of N_{SAA} samples, a suboptimal solution to problem (19) is proven to be, with an acceptable degree of probability, a suboptimal solution to problem (18). To this purpose, we denote the corresponding ϵ -suboptimal solution and δ -suboptimal solution sets for these problems:

$$S^\epsilon = \{x \in X : f(x) \leq f^* + \epsilon\},$$

$$\hat{S}_{N_{SAA}}^\delta = \left\{ x \in X : \hat{f}_N(x) \leq f_N^* + \delta \right\}.$$

In the following we will establish a bound on the sample number N_{SAA} required such that for a precision $\delta \in [0, \epsilon)$

we will have $\mathbb{P}(\hat{S}_{N_{SAA}}^\delta \subset S^\epsilon) \geq 1 - \alpha$, i.e. a δ -suboptimal solution for the approximate problem (19) will be a ϵ -suboptimal solution to the original problem (18) with probability at least $1 - \alpha$. To obtain the bound on the number of samples N_{SAA} , we first need to introduce the following assumptions [Shapiro et al., 2009]:

- A.1 The expected value of $f(x)$ is well defined and finite for all $x \in X$.
- A.2 The set X is bounded and not necessarily finite. For any $x', x \in X$ there exists a constant $\sigma > 0$ such that the moment-generating function $M_{x',x}(t) = \mathbf{E}[e^{tY_{x',x}}]$ of the random variable $Y_{x',x} = [F(x', \xi) - f(x')] - [F(x, \xi) - f(x)]$ satisfies:

$$M_{x',x}(t) \leq \exp\left(\frac{\sigma^2 \|x' - x\|^2 t^2}{2}\right) \quad \forall t \in \mathbb{R}.$$

- A.3 There exists a constant L , independent of ξ , such that
- $$|F(x', \xi) - F(x, \xi)| \leq L \|x' - x\|,$$
- for a.e. $\xi \in \Xi$ and all $x', x \in X$.

Given these assumptions, we have the following theorem for selecting the sample size N_{SAA} .

Theorem 1. ([Shapiro et al., 2009]). Suppose that assumptions A.1-A.3 hold, $f(x)$ is convex in x , and the set of optimal solutions S is nonempty, closed, and convex. Given constants $\epsilon, \delta \in [0, \epsilon)$ and $\epsilon' > 3/4\epsilon + \delta/4$, we have that:

$$\mathbb{P}(\hat{S}_N^\delta \subset S^\epsilon) \geq 1 - \alpha, \quad (20)$$

for any $a > \epsilon$ and a sample size N_{SAA} that satisfies

$$N_{SAA} \geq \frac{\mathcal{O}(1)\sigma^2 \bar{D}_{a,\epsilon}^2}{(\epsilon - \delta)^2} \left[n \ln\left(\frac{\mathcal{O}(1)LD_a^*}{\epsilon - \delta}\right) + \ln\left(\frac{1}{\alpha}\right) \right],$$

where $\mathcal{O}(1)$ is a generic constant, D_a^* is the diameter of S^a and $\bar{D}_{a,\epsilon} = \sup_{x \in S^a/S^{\epsilon'}} \text{dist}(x, S)$.

If additionally $f(x)$ is strongly convex with a parameter σ_f and we take $\delta \in [0, \epsilon/2]$, then (20) can be attained with sample size that satisfies

$$N_{SAA} \geq \frac{\mathcal{O}(1)\sigma^2}{\sigma_f \epsilon} \left[n \ln\left(\frac{\mathcal{O}(1)LD_a^*}{\epsilon}\right) + \ln\left(\frac{1}{\alpha}\right) \right]. \quad (21)$$

Now, given the power dispatch problem, we have that the sets X and W are both bounded polyhedra and assumption A.1 is satisfied. Regarding assumption A.3, in our case we have $F(x, \xi) = F(\bar{K}, \bar{w}) = V(\bar{K}, \bar{w})$. Thus, for any \bar{K} and \bar{K}' we would have that:

$$\begin{aligned} & |V(\bar{K}, \bar{w}) - V(\bar{K}', \bar{w})| \\ & \leq |\bar{V}(\bar{K}\bar{w}) - \bar{V}(\bar{K}'\bar{w})| + |\bar{w}^T H_{\bar{w}\bar{u}}(\bar{K} - \bar{K}')\bar{w}| \\ & \leq |\bar{w}^T \bar{K}^T \bar{Q} \bar{K} \bar{w} - \bar{w}^T \bar{K}'^T \bar{Q} \bar{K}' \bar{w}| \\ & \quad + (\|\bar{q}\| \|\bar{w}\| + \|\bar{w}\|^2 \|H_{\bar{w}\bar{u}}\|) \|\bar{K} - \bar{K}'\| \leq \kappa(\bar{w}) \|\bar{K} - \bar{K}'\| \end{aligned}$$

where $\kappa(\bar{w}) = (\|\bar{w}\|^2 (\lambda_{\max}(Q) + \|H_{\bar{w}\bar{u}}\|) + \|\bar{q}\| \|\bar{w}\|)$ is now the Lipschitz constant depending on \bar{w} . Given that W is a bounded polyhedron and implicitly \mathbf{W} is bounded, we can take the Lipschitz constant $L = \max_{\bar{w}} \kappa(\bar{w})$ and assumption A.3 is satisfied.

Regarding assumption A.2, we have $\mathbb{E}[Y_{x',x}] = 0$. By Hoeffding's inequality [Shapiro et al., 2009], if we have that $|Y_{x',x}| \leq \sigma$, where $\sigma > 0$, then $M_{x',x}(t) \leq \exp(\sigma^2 t^2 / 2)$. For our problem we have through Lipschitz continuity that:

$$\begin{aligned} |Y_{\bar{K}', \bar{K}}| & \leq |V(\bar{K}', \bar{w}) - V(\bar{K}, \bar{w})| + |\mathbb{E}[V(\bar{K}, \bar{w}) - V(\bar{K}', \bar{w})]| \\ & \leq (\kappa(\bar{w}) + \mathbb{E}[\kappa(\bar{w})]) \|\bar{K} - \bar{K}'\| \\ & \leq \left(\max_{\bar{w}} \kappa(\bar{w}) + \mathbb{E}[\kappa(\bar{w})] \right) \|\bar{K} - \bar{K}'\| \\ & \leq 2L \|\bar{K} - \bar{K}'\|, \end{aligned}$$

where the last inequality follows from the fact that $\mathbb{E}[\kappa(\bar{w})] \leq \max_{\bar{w}} \kappa(\bar{w})$. Thus, we can take $\sigma = 2L$ and assumption A.2 is satisfied.

Given that the number of constraints in (16) is exponential in the prediction horizon N , we prefer to tackle problem (17), where the number of constraints grows quadratically with N . We thus obtain a number of N_{SAA} random samples \bar{w}_i of the vector \bar{w} such that the optimization problem for the entire prediction horizon is:

$$\begin{aligned} \min_{\bar{K}} \quad & \frac{1}{N_{SAA}} \sum_{i=1}^{N_{SAA}} V(\bar{K}, \bar{w}_i) \quad (22) \\ \text{s.t.} \quad & K^T c_{uw} = d_{uw} \\ & \lambda_j^t \geq 0, K^T c_j^t = -C_w^T \lambda_j^t, \quad t = l, \dots, k \\ & \sum_{t=l}^k (\lambda_j^t)^T d_w \leq d_j, \quad t = l, \dots, k, j \in (12), (13), (15) \end{aligned}$$

Now, given the fact that the cost $V(\bar{w})$ is actually quadratic in terms of $\bar{p} = [p(1) \dots p(N)]^T$, we can write $p(k) = E_{up} u(k)$ and $\bar{p} = \bar{E}_{up} \bar{u}$, where $\bar{E}_{up} = I_N \otimes E_{up}$. Thus, considering that $\bar{u} = \bar{K} \bar{w}$ we have:

$$V(\bar{K}, \bar{w}_i) = \bar{w}_i^T \bar{K}^T \bar{E}_{up}^T \bar{Q} \bar{E}_{up} \bar{K} \bar{w}_i + \bar{q}^T \bar{E}_{up} \bar{K} \bar{w}_i. \quad (23)$$

We can replace the costs $V(\bar{K}, \bar{w}_i)$ by auxiliary variables t_i and some further constraints added to the problem. Thus, problem (22) is formulated as an SDP:

$$\begin{aligned} \min_{t, \bar{K}} \quad & \frac{1}{N_{SAA}} \mathbf{1}_{N_{SAA}}^T t_i \\ \text{s.t.} \quad & K^T c_{uw} = d_{uw} \\ & \lambda_j^t \geq 0, K^T c_j^t = -C_w^T \lambda_j^t, \quad t = l, \dots, k \\ & \sum_{t=l}^k (\lambda_j^t)^T d_w \leq d_j, \quad t = l, \dots, k, j \in (12), (13), (15) \\ & \begin{bmatrix} t_i & y_i(\bar{K})^T \\ y_i(\bar{K}) & \bar{Q}^{-1} \end{bmatrix} \succeq 0, \quad i = 1, \dots, N_{SAA}, \end{aligned}$$

where $y_i = \bar{E}_{up} \bar{K} \bar{w}_i + \bar{Q}^{-1} (\frac{1}{2} \bar{q} + \bar{E}_{up} \bar{q}_{wu})$ and $\bar{q}_{wu} = H_{\bar{w}\bar{u}}^T \bar{w}_i$.

5.2 Stochastic gradient method

We can also apply other alternative methods for solving the stochastic optimization problems defined previously, such as the stochastic gradient method. In order to have an optimization problem of a more standard form, we denote by $z = \text{vec}(K)$ the vectorization of matrix K . Given this, the term $w^T K^T Q K w + q^T K w$ can be written as:

$$z^T (w w^T \otimes Q) z + (w \otimes q)^T z.$$

Recall that the cost function $V(\bar{K}, \bar{w})$ is actually:

$$F(K, \bar{w}) = \sum_{k=0}^{N-1} w(k)^T K^T Q K w(k) + (q^T + w(k)^T H_{wu}) K w(k)$$

and given that $z = \text{vec}(K)$, it can be rewritten as:

$$F(z, \bar{w}) = z^T Q_w z + q_w^T z, \quad (24)$$

where

$$Q_w = \sum_{k=0}^{N-1} w(k)w(k)^T \otimes Q, q_w = \sum_{k=0}^{N-1} w(k) \otimes (q^T + w(k)^T H_{wu}).$$

The stochastic gradient method solves a problem of type (18), where $F: D \times \Omega \rightarrow \mathbb{R}$ and $\xi: \Omega \rightarrow \mathbb{R}^n$ is a convex vector with a probability distribution on a set $\Omega \in \mathbb{R}^n$. In order to apply the stochastic gradient method, we first require the following assumptions.

Assumption 1. There exists a possibility to generate i.i.d. samples ξ_1, ξ_2, \dots of realizations of the random vector ξ .

Assumption 2. The set $X \subset D$ is convex and closed. The function $F(\cdot, \xi)$ is convex for every $\xi \in \Omega$ and $f(x)$ is finite for every $x \in D$.

Assumption 3. There exists a scalar $M > 0$ such that

$$\mathbb{E}[\|\nabla F(x, \xi)\|_2^2] \leq M^2 \quad \forall x \in X. \quad (25)$$

Under the latter assumption, the function f is convex over X and the following holds [Nemirovski et al., 2009]:

$$\partial f(x) = \mathbb{E}[\partial_x F(x, \xi)] \quad \forall x \in D,$$

i.e. the expectation of the subgradients of F in x are subgradients of $f(x)$. For the moment, we assume that $f(x)$ is differentiable and denote by $\nabla f(x)$ its gradient. The classic stochastic gradient algorithm has an iteration that mimicks the deterministic gradient method. Thus, given an initial point $x_0 \in X$ and stepsizes γ_k , then the iteration of the algorithm is:

$$x_{k+1} = \Pi_X(x_k - \gamma_k \nabla F(x_k, \xi_k)), \quad (26)$$

where $\Pi_X(y)$ denotes the projection operator of the vector y on the set X . For a convex objective function $f(x)$, we have the following theorem for the convergence rate of the stochastic gradient algorithm.

Theorem 2. [Nemirovski et al., 2009] If we denote by $D_X = \max_{x \in X} \|x - x_1\|$, $\nu_k = \gamma_k / \sum_{t=1}^k \gamma_t$, and the approximate solution $\tilde{x}_1^k = \sum_{t=1}^k \nu_t x_t$, then for the iterates (26) we have that $\mathbb{E}[f(\tilde{x}_1^k) - f(x^*)] \leq \frac{D_X M}{\sqrt{k}}$ for a constant step size $\gamma_k = \frac{D_X}{M\sqrt{k}}$. Additionally, if we assume that $f(x)$ is strongly convex with a constant σ_f , then for a stepsize $\gamma_k = \frac{\Theta}{k}$ with a constant $\Theta > \frac{1}{2\sigma_f}$ we have that $\mathbb{E}[\|x_k - x^*\|^2] \leq \frac{R(\Theta)}{k}$, where $R(\Theta) = \max\left\{\frac{\Theta^2 M^2}{2\sigma_f \Theta - 1}, \|x_1 - x^*\|^2\right\}$.

For the proof of this theorem see [Nemirovski et al., 2009]. Given the first result in the theorem above and a precision ϵ for the suboptimality of \tilde{x}_1^k , by Markov's inequality we have $\mathbb{P}(f(\tilde{x}_1^k) - f(x^*) \geq \epsilon) \leq \frac{D_X M}{\epsilon\sqrt{k}}$. Thus, if we impose that $\mathbb{P}(f(\tilde{x}_1^{N_{SG}}) - f(x^*) < \epsilon) > 1 - \alpha$, i.e. $\tilde{x}_1^{N_{SG}}$ is a ϵ -suboptimal solution to problem (18) with probability at least $1 - \alpha$, $\alpha \in (0, 1)$, we require that the number of iterations N_{SG} satisfy $N_{SG} \geq \frac{D_X^2 M^2}{\epsilon^2 \alpha^2}$. Under the light-tail assumption [Nemirovski et al., 2009]:

$$\mathbb{E}\left[\exp\left(\frac{\|\nabla F(x, w)\|^2}{M^2}\right)\right] \leq \exp(1) \quad \forall x \in X, w \in W, \quad (27)$$

and for $\epsilon > 0$ and $\alpha \in (0, 1/2)$, an iterate $\tilde{x}_1^{N_{SG}}$ of (26) satisfies $\mathbb{P}\left[f(\tilde{x}_1^{N_{SG}}) - f(x^*) > \epsilon\right] \leq \alpha$, after a number of iterations N_{SG} which satisfy:

$$N_{SG} \geq \frac{\mathcal{O}(1) D_X^2 M^2 \ln^2(1/\alpha)}{\epsilon^2}.$$

Now, given the form (24) of the objective function, we get $\nabla F(z, w) = Q_w z + q_w$ and:

$$\begin{aligned} \|\nabla F(z, \bar{w})\| &\leq \|Q_w\| \|z\| + \|q_w\| \\ &= \sum_{k=0}^{N-1} \|w(k)\|^2 (\|Q\| \|z\| + \|H_{wu}\|) + \|w(k)\| \|q\|. \end{aligned}$$

Now, given that \mathcal{K} and W are bounded sets, we denote:

$$V_z = \max_{z \in \mathcal{K}} \|z\|, V_w = \max_{w \in W} \|w\|,$$

which are both finite. We have $\|\nabla F(z, \bar{w})\| \leq M$, where:

$$M = N (V_w^2 (\|Q\| V_z + \|H_{wu}\|) + V_w \|q\|),$$

which is finite and thus (25) and (27) are implicitly satisfied. For a strongly convex f with constant σ_f , the second result in Theorem 2 for a number of iterations N_{SG}^{sc} :

$$N_{SG}^{sc} \geq \max\left\{\frac{\Theta^2 M^2}{\epsilon \alpha (2\sigma_f \Theta - 1)}, \frac{\|x_1 - x^*\|^2}{\epsilon \alpha}\right\}$$

we would obtain $\mathbb{P}(\|x_{N_{SG}^{sc}} - x^*\|^2 \geq \epsilon) \leq \delta$. Note that the stochastic gradient method is essentially an iterative method in which requires samplings at each iteration. On the other hand, the SAA method first of all requires samplings in order to approximate the stochastic optimization problem by a deterministic one. This still leaves us with a deterministic optimization problem of considerable dimensions which still needs to be solved.

REFERENCES

- A.M. Elaiw, X. Xia, and A.M. Shehata. Application of model predictive control to optimal dynamic dispatch of generation with emission limitations. *Electric power systems research*, 84(1):31–44, 2012.
- A. Jokic. *Price-based optimal control of electrical power systems*. PhD thesis, T.U. Eindhoven, 2007.
- A.J. Kleywegt, A. Shapiro, and T. Homem-de Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502, 2002.
- J. Linderoth, A. Shapiro, and S. Wright. The empirical behavior of sampling methods for stochastic programming. *Annals Op. Research*, 142(1):215–241, 2006.
- A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19(4):1574–1609, 2009.
- P. Patrinos, S. Trimoli, and A. Bemporad. Stochastic MPC for real-time market-based optimal power dispatch. In *Conf. on Decision and Control*, 2011.
- P. Patrinos, D. Bernardini, A. Maffei, A. Jokic, and A. Bemporad. Two-time-scale MPC for economically optimal real-time operation of balance responsible parties. In *Proc. IFAC 8th Power Plant and Power Systems Control Symposium*, pages 741–746, Toulouse, France, 2012.
- M. Shahidehpour, H. Yamin, and Z.Y. Li. Market operations in electric power systems, 2002.
- A. Shapiro, D. Dentcheva, and A.P. Ruszczyński. *Lectures on stochastic programming: modeling and theory*, volume 9. SIAM, 2009.
- L. Xie and M.D. Ilic. Model predictive dispatch in electric energy systems with intermittent resources. In *IEEE International Conference on Systems, Man and Cybernetics*, pages 42–47, 2008.