Implementation of the AC-based economic dispatch in the Russian electricity market

Mikhail Davidson
ConsultNext, SkolTech, MSU
Moscow, Russia
m.davidson@consultnext.ru

Gleb Labutin
Markets Development Department
JSC “System Operator of United Power System”
Moscow, Russia
labutin@so-ups.ru

Abstract—The Russian wholesale electricity market started to operate in its present form since 2006. Its unique specificity is the AC-based economic dispatch model underlying the market applications. Solution method for the day-ahead and balancing market applications was developed and implemented. Now it is employed by the commercial operator and the system operator of the Russian energy system. It is proved to be an adequate and reliable scheduling market technology.

Index Terms—Day ahead market, balancing market, AC OPF, sequential quadratic programming.

I. INTRODUCTION

In the period 2001-2008 the Russian electric power sector had undergone a substantial reform. As a result a competitive electricity market was established and a number of independent generation and distribution companies emerged. An overview of the market model is given in [3].

This required a new technology of market based short-term scheduling which was developed and implemented in 2006-2007. Since then it is exploited by the commercial market operator (Administrator of the Trading System – ATS) that runs the Day-Ahead market and the System Operator (SO UPS) that runs the Balancing market. The aim of the latter is to reflect the most recent system conditions and to trade deviations from the Day-Ahead market. Balancing market schedule is computed every one hour ahead of real time and the resulting generation station schedules have a status of dispatch instructions.

Most part of the energy system of Russia is interconnected and operates at a single frequency. The system is highly constrained and is distinctive for transferring significant amounts of power over HV backbone grid, and pattern of these transfers is changing throughout a day. A DC model widely used in many energy markets might lead to significant scheduling errors and non-transparent deviations. The AC model yields the required accuracy in modelling losses and power flows to correctly account for the constraints set up by the System Operator. For these reasons the mathematical model underlying the market applications is AC-based economic dispatch problem.

The network model used in the market applications has around 9000 nodes and 15000 buses. The time horizon is broke down into hourly intervals for which intertemporal constraints are present: ramp rate constraints and energy volume constraints (especially for hydros). Hence the resulting optimization problem is a large scale nonlinear and nonconvex program. Due to the size and nonlinearity the issues of timing and convergence reliability emerge that make developing solution algorithm a challenging problem.

The algorithm is based on sequential quadratic programming with an intensive use of parallel computations. The method is efficient for both day-ahead market and balancing market where the timing restrictions are tighter. In this paper we focus on the model setup for the balancing market for which the key difference with the day-ahead is that the load forecast is provided by the system operator while at the day-ahead consumer bids are used.

II. MODEL SETUP AND PROBLEM FORMULATION

In this section we describe constraint types and control variables used in the model. The model includes both intertemporal and hourly constraints. The latter are referred to by the index \( \tau = 1, ..., T \), where \( T \) is the number of hourly intervals in the session. Also a subset of variables relating to hour \( \tau \) is denoted by the corresponding index.

A. Power balance constraints

Trading in the RF energy market concerns only active power and not reactive power. While reactive power balance equations are explicitly present in the problem formulation, optimization of reactive power is not performed in order to avoid potentially negative effects of using active power controls to enforce VQ constraints and vice-versa [9].

The values of such controls as transformer taps and generator voltage setpoints and other devices that essentially impact the reactive power balance are accepted as inputs to the problem and not varied. At the same time regulation of reactive power by generation sources is modeled through switching PV and PQ bus types depending on the maximum and minimum reactive power capacity of a source and
required reactive power consumption/generation at the corresponding bus. The maximum and minimum reactive power capacity is assumed constant not depending on the active power output of a generator. In other words accounting for reactive power is similar to the load flow model.

The active power controls that are subject to optimization at the balancing market are the volume bids submitted by the generators. The load forecast is provided by the System Operator. Thus, the power balance equations connect the state variables \( y \) (voltage magnitudes and angles), the active power volume bids \( x \), and the fixed net bus injections \( d \) representing the demand:

\[
[\lambda] \quad F_r(y_r) + B x_r = d_r, \tag{1}
\]

(here the quantity in square brackets denotes the corresponding Lagrange multipliers, \( B \) is the buses - bids incidence matrix. The number of equations in \( 1 \) is twice the number of buses (excluding swing bus) reflecting active and reactive nodal power balance. We use separate notation for the swing bus balance:

\[
[\lambda_0] \quad F_{or}(y_r) = 0. \tag{2}
\]

Since we don’t optimize for reactive power controls the reactive power swing bus equation is slack and \( 2 \) represents only active power balance (the right hand side is zero to simplify the notation).

B. Power flow constraints

Power flow constraints represent a set of most likely contingency scenarios under given operating conditions modelled and precomputed by the System Operator. For every such constraint SO defines a power flow limit that guarantees system security after any single contingency, and these are the inputs to the model:

\[
[\sigma] \quad G_r(y_r) \leq \tilde{f}_r. \tag{3}
\]

C. Intertemporal constraints

These include ramp rate and fuel generator constraints. The latter model the daily water supply of hydros to be optimally allocated among hourly intervals. Both ramp rate and fuel constraints are linear and link only volume variables. We denote these constraints by a single inequality (including upper and lower bounds):

\[
IT x \leq b_{IT} \tag{4}
\]

where \( x = (x_1, ..., x_T)^T \).

D. Bounds for variables

For generation volumes \( x \), there are upper and lower bounds submitted by the generators in their bids

\[
lb_x \leq x \leq ub_x. \tag{5}
\]

We don’t apply bounds on voltage magnitudes since as said above the controls affecting reactive power and voltage magnitudes are not optimized.

E. Problem formulation

The objective function used at the RF balancing market is minimum of the cost of energy produced by the generators according to their price bids. Thus, the problem formulation is

\[
\sum_{\tau} c_{\tau}^T x_{\tau} \to \min \tag{6}
\]

subject to constraints \( 1 \) – \( 5 \). This is a nonlinear, nonconvex and large-scale optimization problem. The size of the problem is of order 5e+5 variables and constraints. Modeling PV-PQ bus switching brings a discrete discontinuous nature to the problem.

III. THE ALGORITHM

SQP methods proved to be efficient in application to general nonlinear problems [5]. In particular, successful application of SNOPT [4] to power flow optimization is reported in [2]. Here we describe a variant of SQP algorithm used in the Russian energy market.

For a given pair \( (y, x) \) and the corresponding vector of Lagrange multipliers \( (\lambda, \sigma) \) generic form SQP sub-problem as applied to \( 5 \) appears as follows

\[
\sum_{\tau} \left( c_{\tau}^T \Delta x_{\tau} + 1/2 \Delta y_{\tau}^T \mathbf{H}_r(\lambda_{\tau}, \sigma_{\tau}, y_{\tau}) \Delta y_{\tau} \right) \to \min \tag{7}
\]

subject to

\[
\mathbf{D}_r(y_{\tau}) \Delta y_{\tau} + \mathbf{B}_r \Delta x_{\tau} = r_{\tau}^F \tag{8}
\]

\[
\mathbf{D}_{or}(y_{\tau}) \Delta y_{\tau} = r_{\tau}^{or} \tag{9}
\]

\[
\mathbf{D}_G(y_{\tau}) \Delta y_{\tau} \leq r_{\tau}^G \tag{10}
\]

\[
IT \Delta x \leq u^{IT} \tag{11}
\]

\[
lb_x \leq x + \Delta x_{\tau} \leq ub_x \tag{12}
\]

where \( 8 \) – \( 10 \) are linearization of constraints \( 1 \) – \( 3 \) around \( y \), and

\[
\mathbf{H}_r(\lambda_{\tau}, \sigma_{\tau}, y_{\tau}) = \sum_i \left( \lambda_{\tau}^T \nabla^2_{y_{\tau}} F_{\tau}(y_{\tau}) \right) + \sum_j \left( \sigma_{\tau}^T \nabla^2_{y_{\tau}} G_{\tau}(y_{\tau}) \right)
\]

is the element of the Hessian of the Lagrange function of \( 6 \) corresponding to hour \( \tau \). Only Hessians of constraints \( 1 \) – \( 3 \) enter the formula for \( H_r \) (let the first sum contain Hessian terms of both equations \( 1 \) and \( 2 \)).

Despite the size of the problem is large, the number of degrees of freedom is much smaller. As long as the state variables \( y \) stay away from the boundary of the solvability region the Jacobian \( \mathbf{D}_r(y_{\tau}) \) is nonsingular and the states \( y \) could be seen as implicit functions of controls \( x \), and the state increments could be expressed via the control increments:

\[
\Delta y_{\tau} = \mathbf{D}_r^{-1}(y_{\tau}) r_{\tau}^F - \mathbf{D}_r^{-1}(y_{\tau}) \mathbf{B}_r \Delta x_{\tau}. \tag{13}
\]

Clearly dimension of the controls (the number of generator bids) is much smaller than the number of buses. Note that
$DF_r(y_r)$ is sparse and sparse LU factorization could be used to efficiently perform the required computations. The reduced gradients of $F_0r$ and $G_r$ could be readily computed. Despite the reduced gradients does not keep sparse structure, the size of the corresponding matrices is generally small. Dimension of $F_0r$ is the number of swing buses (islands) in the system. Most part of the system operates as a single whole, however temporarily a part may operate separately due to grid maintenance. In practice the number of islands does not exceed 3. Flow constraints (3) are managed using an extended active set strategy: inequality (10) contains linearized constraints (3) binding at the current iterate $y_r$ or at any of the previous iterates. At a cost of slightly expanding the active set we avoid oscillations and stabilize the convergence. In our experience the number of the elements in the extended active set rarely exceeds 20 per hourly interval. Hence computation of the reduced gradients for (2) and (3) is cheap.

Additional computational savings could be achieved by employing the fact that the structure of the Jacobian does not change from iteration to iteration: the symbolic analysis of the matrix structure required for the sparse LU can be performed only once. However, this holds only until there is PV-PQ bus type switching.

Computation of the reduced Hessian is the most time consuming. The reduced Hessian has the form

$$\tilde{H}_r(\lambda_r, \sigma_r, y_r) = B_r^T(DF_r^{-1})^TH_rDF_r^{-1}B_r.$$ 

The $L$, $U$ factors of $DF_r$ are not sparse and so is $\tilde{H}_r$. While dimension of $\tilde{H}_r$ is much smaller than that of $H_r$, we in addition take into account that the columns of $B_r$ are identical for generators located at the same bus. For computing the reduced Hessian it is sufficient to keep only one column per bus which reduces the time for computing $\tilde{H}_r$ up to three times.

In order that the solution to the SQP sub-problem be well defined $\tilde{H}_r$ has to be positive semidefinite. In general this does not hold. In the algorithm we use a positive semidefinite approximation to the reduced Hessian denoted by $\tilde{H}_r^+$.

In reduced form the SQP sub-problem (7) will appear as follows

$$\sum_r (\tilde{c}_r^T \Delta x_r + 1/2 \Delta x_r^T \tilde{H}_r^+ \Delta x_r) \rightarrow \min$$

subject to constraints

$$L_{0r} \Delta x_r = r_{0r}^r \quad (15)$$
$$S_r \Delta x_r \leq r_s^r \quad (16)$$

and (11), (12) that remain unchanged. Here $L_{0r}$ and $S_r$ are the reduced gradients of $F_0r$ and $G_r$, respectively, $\tilde{c}_r$ is the combination of $c_r$ and terms that appear after substitution (13):

$$\tilde{c}_r = c_r - (r_{0r}^F)^T(DF_r^{-1})^TH_rDF_r^{-1}$$  \hspace{1cm} (17)$$

Note that $L_{0r}$ is loss coefficient vector and $S_r$ is power transfer distribution factor matrix representing active loss and power flow constraint sensitivities to bus injections. In particular

$$L_{0r} = 1 + \partial L_r / \partial x_d^d$$  \hspace{1cm} (18)$$

where on the right hand side there is the gradient of the active power loss function with respect to active power nodal injections. The optimal solution $\Delta x_r^+$ of (14) is used to update the current iterate $x_r$. State variables are updated via (13). To update the dual multipliers $\lambda_r$ note that from the first order necessary conditions for (7) one has:

$$\lambda_r^+ + \lambda_{0r}^rDF_0rDF_r^{-1} + \sigma_r^TDG_rDF_r^{-1} = 0$$  \hspace{1cm} (19)$$

(for the dual multipliers in (7) we use the same notation as for (6)). This yields the well known decomposition of the locational marginal prices (LMPs) through marginal cost of losses and marginal cost of transmission constraints [1].

Thus (14) is solved for $\Delta x_r^+$, $\lambda_{0r}^r, \sigma_r^+$ that are used to update the current iterate and to build a new quadratic sub-problem. The following specifications complete the definition of the SQP algorithm:

- Initial approximation is always chosen to be all zero angles and unit voltages (“flat start”). The multipliers $\lambda_{0r}$ are initialized according to (19) with $\lambda_{0r}$ being some positive proxy of the system price. The multipliers $\sigma_r$ are zero as transmission constraints are not tight at flat start state.

- Since the algorithms starts from an infeasible state, at the first stage the step size is chosen in order to minimize the error of the power balance nodal equations; if at some iteration the feasibility error is within the specified tolerance the step size is chosen to minimize a combination of the objective and the feasibility error.

- The stopping criterion is defined using the error of the first order necessary conditions for (6).

This breaks down into checking that the current state is feasible to a given tolerance and that the norm of $\tilde{H}_r^+ \Delta x_r$ is
within the optimality tolerance. The first order optimality conditions for the SQP sub-problem (14) imply

$$\bar{c}_r + \bar{H}^r \Delta x_r + L^T_{0r} \lambda_{0r} + S^T_{0r} \sigma_r + \pi_r = 0$$  \hspace{0.5cm} (20)

where $\pi_r$ denote the aggregate dual term related to constraints (11) and (12). For every unit $\pi_r$ is a marginal profit/loss of the unit given the system price and the unit price bid. Since the state is feasible up to a given tolerance, the residual $\pi_r$ is small and then from (17) $\bar{c}_r$ is close to $c_r$, and (20) is close to

$$c_r + \pi_r = L^T_{0r} \lambda_{0r} + S^T_{0r} \sigma_r$$ \hspace{0.5cm} (21)

The complementarity conditions in (14) provide that unit’s marginal profit $\pi_r$ is greater or equal zero as long as the unit operates at its maximum capacity and less or equal zero if it is at its minimum. Then (21) yields the equilibrium condition for every unit at given system price at the right hand side of (21).

Since energy pricing in the Russian energy market is based on LMPs, conditions (20) make the resulting schedule and prices clear, transparent and justifiable for the market participants. Conditions (20) have dimension of national currency units per MWh. Hence the choice of the optimality tolerance could be linked to the minimum fraction of the currency unit (for example, one hundredth) up to which the price is stated for accounting purposes.

Finally in this section we note that the computations required to evaluate the constraints and the reduced gradients and the hessian matrix are independent for each time interval which allows using parallelization. This leads to substantial time savings as parallelization covers the most time consuming computations.

The reduced problem has the dimensions an order of magnitude less than the original problem and can be efficiently solved by standard QP-solvers.

IV. CONVERGENCE

Conditions that guarantee local convergence of the SQP algorithm are studied in a number of papers, see [5] for a review. In [7] the local quadratic rate of convergence to a solution $x^*$ is proved under the following assumptions:

a) Second order necessary optimality conditions hold at the solution.

b) The gradients of the active constraints at the solution are linearly independent.

c) Strict complementarity slackness holds.

As shown in [10] last two conditions could be relaxed. Instead of b) it is sufficient that Mangasarian-Fromovitz [6] constraint qualification holds, i.e.

d) the gradients of equalities (in our case (15)) are linearly independent, and there is a point where inequalities (11), (12), and (16) are strict.

While the original condition of linear independence of the gradients in fact implies the uniqueness of optimal dual solution, in the relaxed form it is sufficient to require just an existence of such dual solution for which the strict complementarity holds.

In order to analyze the structural properties of optimization problem arising in the Russian energy market from the point of view of the above sufficient conditions we tried to select a representative set of sessions for an empirical study. The set covers the first half of 2017 and includes two sessions for every month that correspond to a working day and a weekend. The total number of selected sessions is thus twelve each containing 24 hourly intervals. For each session we took a solution satisfying the stopping criterion to which the method converged and analyzed the above conditions.

It turned out that b) is not satisfied for any of the sessions selected while d) holds for all of them. We were not able to verify the relaxed form of c), but for the dual solution obtained as a result of the iteration process the strict complementarity does not hold. However if one takes not all the active constraints but only nonlinear active constraints (whose linearized form is (15) and (16)) then c) is satisfied for them in all cases considered, and moreover, the gradients of these constraints are linear independent. In general this fact by itself does not imply uniqueness of the corresponding dual multipliers. But for the problem in question these multipliers define the nodal prices. Throughout the test set considered the nodal prices were unique for the primal solution obtained. This does not imply, however, the uniqueness of the whole vector of the dual solution. The linear constraints (11) and (12) are box constraints or ramp-rate constraints for units, or daily energy volume constraints for stations. Typical examples of degeneracy here might be bids with the same price at the same bus or units with ramp rate exactly equal to the MWt range between minimum and maximum capacity of the unit. Some of such degeneracies could be preprocessed but in our experience they don’t essentially affect the convergence since they don’t affect the nodal prices. Most important is the uniqueness of the nodal prices as they enter the expression for the Hessian, and are critical for the convergence.

The Hessian $H_r(\lambda_r, \sigma_r, y_r)$ is generally not positive definite and condition a) does not hold. However, it is positive definite at “flat start” state. Let $\lambda^a_r$ denote dual multipliers to active power balance constraints, and $\bar{\lambda}_r$ denote dual multipliers to reactive power balance constraints (for a particular $r$). Note that power flow constraints prices $\sigma_r$ are zero at “flat start”. Then, the multipliers $\lambda^a_r$, which are the nodal prices, can be expressed as

$$\lambda^a_r = \lambda_{0r} (1 + \partial L_r / \partial x^a_r) ,$$

(see (18), (19)). Similarly

$$\lambda^r_r = \lambda_{0r} \partial L_r / \partial x^r_r ,$$
where $\frac{\partial L_r}{\partial x'_r}$ is the gradient of the active power loss function with respect to reactive power nodal injections.

At “flat start” the components of the gradient of the active power loss function are zero, and hence nodal prices $\lambda'_r$ are equal for all nodes and vector $\lambda'_r$ is zero. In such a case $H_r$ coincides (up to a multiplier) with the Hessian of the active power loss function (as a functions of state variables) which is positive definite if line reactance in the system is positive. In some neighborhood of the “flat start” state corresponding to “normal” operating conditions (where the components of the loss function gradient are sufficiently small) the hessian $H_r$ is close to convex and the convergence is fast. In abnormal cases when operating point approaches boundary the loss function gradient tends to infinity and convergence is violated.

Summarizing the results of the analysis we conclude that while from a formal point of view the sufficient conditions are not fully satisfied, the violations are not significant for the convergence unless the operating point is close to the boundary of the solvability set. The latter usually shows up in some local fragment of the system which is the source of the problem, and it is critically important to develop mechanisms that help to identify such “weak” locations.

V. CONCLUSIONS

For the period of operation, the model proved to be an adequate short term scheduling market technology. The locational marginal pricing mechanism provides economic signals to participants that are concordant with the dispatch. The accuracy of the model could be illustrated by the amount of dispatch instructions issued beyond the computed schedule. According to the official report of the SO UPS for the year 2016 [8] this is within 1% of the scheduled amounts. Also according to the ATS data (www.rosenergo.ru), the corresponding financial imbalance (or “make-whole” payment) constitutes around 1% of the tressed volume, hence the resulting wholesale price is only slightly blurred by the regulatory component.

The technology has also proved to be reliable. However since the scheduling is time critical, a possibility for the dispatch personnel to analyze the resulting electrical regimes (especially in the case of problems or convergence breakdowns) is limited. Further research and work are aimed at developing algorithms for identification of weak locations in the energy system that may be a cause of divergence and providing the corresponding information to operation personnel.

REFERENCES

[1] R.E. Bohn, M.C. Caramanis, F.C. Schweppe, “Optimal pricing in electrical networks over space and time”, Rand Journal of Economics, vol. 15, pp. 360-376, 1984

[2] A. Castilho, R. P. O’Neill, “Computational performance of solution techniques applied to the ACOPF”. Federal Energy Regulatory Commission Staff Technical Paper, 2013. [Online]. Available: https://www.ferc.gov/industries/electric/indus-plb/mkt-planning/opf-papers/acopf-5-computational-testing.pdf

[3] M.R. Davidson, Yu. V. Dogadushkina, E. M. Kreines, N.M. Novikova, Yu. A. Udaltsov, and L. V. Shiryaeva, “Mathematical Model of the Competitive Wholesale Power Market In Russia”. J. of computer and Systems Sciences International. 2004; 43(3), pp. 72-83

[4] P.E.Gill, W.Murray, M.D.Saunders, “SNOPT: An SQP algorithm for large-scale constrained optimization”; SIAM Review, vol. 47, pp.99-131.

[5] N. I. M. Gould and Ph. L. Toint, “SQP methods for large-scale nonlinear programming”, in System Modelling and Optimization (Cambridge, 1999), Kluwer Academic, Boston, MA, 2000, pp. 149–178.

[6] O.L. Mangasarian and S. Fromovitz, “The Fritz-John necessary optimality conditions in the presence of equality and inequality constraints,” Journal of Mathematical Analysis and Applications, vol. 17, pp. 37–47, 1967.

[7] S. M. Robinson, “Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear-programming algorithms,” Mathematical Programming, vol. 7, pp. 1-16, 1974.

[8] Report on operation of the United Power System in 2016, System Operator of the United Power System, 2017. [Online]. Available: http://so- ups.ru/fileadmin/files/company/reports/disclosure/2017/ups_rep2016.pdf [In Russian]

[9] B. Stott, O. Alsac. “Optimal Power Flow—Basic Requirements for Real-Life Problems and their Solutions,” White Paper, 2012.

[10] S. J. Wright, “Superlinear Convergence of a Stabilized SQP Method to a Degenerate Solution,” Computational Optimization and Applications, vol. 11, pp. 253-275, 1998.