Accelerated Distributed Solutions for Power System State Estimation

Samal Kubentayeva† Elena Gryazina**,*** Sergei Parsegov**,*** Alexander Gasnikov**,****,†

* Moscow Institute of Physics and Technology, Moscow, Russia (e-mail: samal.kubentayeva@phystech.edu)
** Skolkovo Institute of Science and Technology, Moscow, Russia (e-mail: e.gryazina@skoltech.ru, s.parsegov@skoltech.ru)
*** V.A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia
**** A.A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia
† Caucasus Mathematical Center, Adyghe State University, Maykop, Russia (e-mail: gasnikov.av@mipt.ru).

Abstract: Recently, distributed algorithms for power system state estimation (PSSE) have attracted significant attention. Along with such advantages as decomposition, parallelization of the original problem and absence of a central computation unit, distributed state estimation may also serve for local information privacy reasons since the only information to be transmitted are the boundary states of neighboring areas. In this paper, we propose a novel approach for speeding up the existing PSSE algorithms, utilizing recent results in optimization theory, obtain corresponding convergence rate estimates and verify our theoretical results by the experiments on a synthetic 140 bus system.

Keywords: Distributed control and estimation, distributed optimisation for large-scale systems, control and estimation with data loss, power system state estimation, convex optimization, ADMM, accelerated optimization methods.

1. INTRODUCTION

During the past two decades, a significant attention has been paid to the distributed cooperation and coordination in networks. Many examples of collective cooperative behavior can be found in nature (flocking in mammalian herds, bird flocks, fish schools, etc.). The growth of interest to problems of such kind resulted in appearance and wide application of theoretically proven distributed algorithms derived to mimic the instinctive behavior of groups of animals Ren and Beard (2008). The research on distributed cooperation and coordination in networks is performed in different areas starting from social psychology and economics to estimation theory, from physics and robotics to biology, etc. Despite the topics are studied by overlapping research communities using different definitions and approaches, they are forming a new theory of interconnected dynamic systems Antonelli (2013). A basic concept behind any decentralized and distributed approach is to solve some global problem related to a network of entities called agents by their local interactions where a) no centralized unit is needed, b) only small amounts of information are transmitted, and c) the actions (computations, controls) can be taken by low cost local units.

Power systems exhibit an example where both control and optimization distributed algorithms can provide better system monitoring and operation along with reduction in computation and communication. Power system state estimation is one of the most important problems here, since it enables online monitoring of the grid and is crucial for normal operation of related tasks as economic dispatch, optimal power flow, e.g. see Gomez-Exposito et al. (2008), Conejo and Baringo (2018) and references therein.

Distributed decentralized PSSE approaches have some advantages over traditional centralized ones as reduction in computation, communication for each local controller that has to process only a portion of the global state of the grid.

Power grids naturally consist of smaller networks referred to as areas. Having local computation units within each area the implementation of distributed state estimation can replace a centralized coordinator: instead of collecting, storing and processing of huge measurement data for the state estimation the same goal can be achieved by dealing with local measurements and their processing with the exchange of boundary state variables solely. Besides preventing the appearance of communication bottlenecks this may also look fascinating in terms of information...

* The research of E. Gryazina and S. Parsegov is partially supported by Russian Science Foundation (RSF) project 16-11-10015. The research of A. Gasnikov is partially supported by RFBR project 18-71-00048 mol-a-ved and by Yahoo! Research Faculty Engagement Program.
privacy if the interacting areas do not want to indicate their internal states (e.g., if areas are countries).

The use of fast distributed algorithms in microgrid applications, that continue to gain popularity, may also have some extra benefits. First, the microgrid can operate being connected to the main grid or in island mode. In case of disconnection, the monitoring by the central control unit will be no more available. Second, the presence of renewable energy sources causes high oscillations in isolated microgrid due to low inertia of its components. Such a behavior can have a big impact on the grid, e.g., blackouts and grid instabilities in general. Thus, for ensuring microgrid secure operation both in terms of state estimation and control actions, the algorithms must be designed guaranteeing predefined duration of estimation (optimization) and transients (control). For example, the design of fast distributed controllers utilizing the concept of fixed-time convergence for microgrids was studied in Zhang and Hredzak (2019). The recent results in distributed state estimation based on different optimization techniques were presented in Minot et al. (2016), Frasca et al. (2015), Kekatos et al. (2017). The most promising in terms of computation burdens, simplicity and converge rate is a distributed state estimation method proposed in Kekatos and Giannakis (2013), Kekatos et al. (2017). The method is based on the alternating direction method of multipliers (ADMM). Some other novel results in this field are related to secure operation Cintuglu and Ishchenko (2019) and distributed state estimation method under hybrid cyber attacks Du et al. (2019). Nevertheless, as compared to the existing results, the problem of computational speedup in distributed PSSE setup is still lacking.

The performance of ADMM is limited due to the presence of just one tunable parameter Lan (2019).

In this paper, we use a novel distributed PSSE method adapting accelerated ADMM algorithm (AADMM) studied in Goldstein et al. (2014) and analyze it practical performance.

For our simulations, we use the data from the MATPOWER database for a system with 14 buses. In an electric network with 14 buses, we use the value from the sensors that are connected to some of the lines or buses. And we check the values of the state of the power system on these sensors. We create a manually simulated power system from blocks of 14 buses. Furthermore, we make the connection of blocks into a single energy network the worst possible type for distributed optimization algorithms. In addition, we provide a convergence rate for ADMM for our energy problem. And we check the correctness and reliability of the methods for PSSE problem. And we show the applicability of our approach in numerical experiments.

The paper is organized as follows. In Section 2 we describe the power grid and problem formulation. In Section 3 we present the algorithms (ADMM and accelerated ADMM) for our problem and provide convergence rate theorem. In Section 4 we describe the example constructed based on benchmark IEEE-14 bus grid and illustrate the effectiveness of these methods with some examples.

2. PROBLEM FORMULATION

We consider electrical network as a connected graph $G = \{V, E\}$, where the set $V$ comprises its $N = |V|$ buses, the set $E$ comprises its $L = |E|$ lines. The set of buses contains three types of network nodes: generators, load buses and one slack bus. The most common way to describe power system work is to use the alternating current (AC) power flow model as in Roald and Andersson (2017). This model works according to Kirchhoff’s Laws. Transmission lines are subject to external thermal constraints and phase and voltage limits.

In this paper we use DC power flow model that is well-known simplification (linearization) of full AC model, because it is sufficiently accurate for high voltage modes in large-scale power grids. The other advantages of the model and its applicability are described in Stott et al. (2009). DC approximation assumes that the voltage amplitudes are the same across the network and the phase difference between connecting nodes are sufficiently small. Besides, for any edges of the grid the resistivity-to-reactance ratio is small. The system is modeled within the so-called DC approximation Chertkov et al. (2011) by the following set of linear equations:

$$\forall i \in V : P_i = \sum_{j \in V} P_{ij}$$

$$\forall i \in V : \theta_i = \sum_{j \in V} P_{ij}$$

Equation (2) can be written as $P = H\theta$, where $P = (P_{ij}) \in V$ is the vector of power injections (generation/consumption), $\theta = (\theta_i) \in V$, is the vector of voltage phases and $H$ is the weighed graph Laplacian matrix with respective inductances $-h$. Further in the paper we denote $v_i = \theta_i$, emphasizing that we deal with variables $v_i$.

We consider two types of sensors present in the grid: line flow and bus power injections measurements. When line connects different areas we assume that the measurements belong to one specified area. Thus every sensor measures $x_k$ being one of the entries in the left hand side of equations (1)–(2) and it is associated with a certain area.

Since we use distributed methods to solve the PSSE problem on the network we divide the power system into $K$ regions with the following properties: a) one area can share vertices with no more than two areas, b) each bus (node) can belong to no more than two areas. Every region can be considered as an independently controlled part of the grid, it may be a substationiation or a city.

We model active power flows and injections measurements as linear function of the system states $H_kv_k$ plus a zero mean Gaussian fluctuation $\epsilon$, therefore:

$$z_k = H_kv_k + \epsilon_k, \ \forall k \in K,$$

where the vector $v_k$ collects the system states related to area $k$, $z_k$ depicts the measurements from the sensors located in area $k$. Therefore, we solve the next optimization problem for power system state estimation in local area

$$\min_{v_k \in \mathcal{X}_k} f_k(v_k) = \min_{v_k \in \mathcal{X}_k} \frac{1}{2} \left\| z_k - H_kv_k \right\|^2,$$
where the convex set \( A_k \) collects prior information about the buses in the grid \( G \). For detailed exposition see in Monticelli (2000).

According with the property of splitting the network, each area has common node (bus) with another one. Two areas that contain a common bus restore the value of this bus, also they take into account the value of each other and adjust shared state. Therefore, in order to solve the PSSE task on full buses grid, we need to introduce conditions for the connections between areas. Denote the voltage of common bus of the regions \( k \) and \( l \) by \( v_k[l] \) and \( v_l[k] \). Thus, we add consensus condition \( v_k[l] = v_l[k] = u_{kl} \), where \( u_{kl} \) is the new variable of distributed methods. Thus, the consensus condition helps to adjust the value of the common bus.

We obtained the following optimization problem, which we solve using ADMM and accelerated ADMM.

\[
\begin{align*}
\min_{(v_k \in A_k)} & \sum_{k=1}^{K} f_k(v_k) \\
\text{s.t.} & \quad v_k[l] = u_{kl} \quad \forall l \in B_k, \\
& \quad v_l[k] = u_{kl} \quad \forall k \in B_l,
\end{align*}
\]

where \( B_k \) is the set of areas sharing states with area \( k \), \( B_l \), respectively, with the same way with area \( l \).

3. METHODS FOR SOLVE PSSE TASK

In this part of paper, we use two efficient methods for solving problem (5), which are also actively used in Power systems, see works by Kekatos et al. (2017) and Sulc et al. (2014). The first method is the classical ADMM, which is described in Boyd et al. (2011) article. The second is the accelerated ADMM from Goldstein et al. (2014). This method is also covered in Ahmadi et al. (2018), Franca et al. (2018) and others. This method uses the idea of accelerated alternating procedures, in which presented a convergence rate theorem for the ADMM method as applied to the PSSE problem, the proof of which uses a similar way as in Theorem 3.10 in Lan (2019).

**Theorem 1.** Let \( v_k^i, u_{kl}^i, y_{kl}^i, \ i = 1, \ldots, N \) is a sequence generated by algorithm 1 with \( \mu > 0 \). Define \( \bar{v}_k = \sum_{i=1}^{N} v_k^i / N \) and the same way \( \bar{u}_{kl}, \bar{y}_{kl} \). Then we have

\[
\begin{align*}
& f_k(\bar{v}_k) - f_k(v_k^i) \leq \frac{\mu}{2N} \sum_{l \in B_k} ||y_k[l] - u_{kl}^i||_2^2, \\
& ||\bar{v}_k[l] - \bar{u}_{kl}||_2 \leq \frac{1}{N} \left( \frac{2}{\mu} ||y_{kl}^i||_2 + ||u_{kl}^0 - u_{kl}||_2 \right). \quad (7)
\end{align*}
\]

The proof is provided in the appendix. We consider the ergodic ADMM sequence \( (\bar{v}_k, \bar{u}_{kl}, \bar{y}_{kl}) \) in Theorem 1, which has a convergence rate higher than in the non-ergodic ADMM system. In addition, a convergence comparison for ergodic and non-ergodic methods can be found in Davis and Yin (2016) paper.

3.2 Accelerated Alternating Direction method of Multipliers

We suggest to apply for our problem Accelerated variant of the ADMM from Goldstein et al. (2014). This method is also covered in Ahmadi et al. (2018), Franca et al. (2018) and others. This method uses the idea of accelerated method from Nesterov (1983) (see Guminov et al. (2019) for accelerated alternating procedures), in which presented a first-order minimization scheme with global convergence \( O(1/N^2) \) for the class of smooth target functions. Nesterov’s accelerated method is a modification of gradient descent, which is accelerated by the step of overrelaxation.
In the work of Goldstein et al. (2014) authors assume that both functions from general problem (6) are strongly convex. But in our case, $g(y)$ is a zero function, so not strongly convex. Therefore, in theory, the method have been applied for considered problem may not yield any acceleration, as was written in Theorem 2 of Goldstein et al. (2014). However, with the right selection of parameters, the method can be significantly accelerated. Thus, we get an effective solution to the PSSE problem. Note, that in the method can be significantly accelerated. Thus, we get a similar arrangement of sensors on a 14-bus power grid. Further, this system is called a block or a group, which is a practical one of the main reason to use ADMM approach and its accelerated variants instead of not augmented and not alternating gradient type procedures and its accelerated variants.

Algorithm 2 Accelerated ADMM

Require: matrix $H$, system state vector $z$, parameter $\rho$, tolerances $\varepsilon_{\text{primal}} = 10^{-3}$, $\varepsilon_{\text{dual}} = 10^{-4}$, initial points $\bar{u}_i^0 = 0, \bar{u}_{kl}^0 = 0, y_i^0 = 0$, $y_{kl}^0 = 0$ for each $k, l \in K$

1: $i \leftarrow 0$
2: repeat
3:   \[
    v_k^i = \min_{v_k \in \mathbb{K}} f(v_k) + \sum_{l \in B_k} \beta_l^{-1} (v_k[l] - u_{kl}^{i-1})
    + \sum_{l \in B_k} \frac{\rho}{2} \|v_k[l] - u_{kl}^{i-1}\|_2^2
    \]
4:   $u_{kl}^i = \frac{1}{2}(v_k^i[l] + v_l^i[k])$
5:   $y_{kl}^i = \frac{1}{2}v_k^i[l] + \rho(v_k^i[l] - u_{kl}^i)$
6:   $\alpha_i = \alpha_i + 1 + \sqrt{1 + \alpha_i^2}$
7:   $\bar{u}_{kl}^i = \frac{1}{2}(u_{kl}^i - u_{kl}^{i-1})$
8:   $\bar{y}_{kl}^i = y_{kl}^i + \frac{\alpha_i - 1}{\alpha_i} (y_{kl}^i - y_{kl}^{i-1})$
9: until $\|x^i\|_2^2 < \varepsilon_{\text{primal}}$ and $\|s^i\|_2^2 < \varepsilon_{\text{dual}}$
10: $i \leftarrow i + 1$

4. NUMERICAL EXPERIMENTS

In this section, we simulate power grids with two different topologies for solving distributed methods. Also, we consider three examples in the modeled network. We test ADMM and ADMM on this system and compare the convergence rate on graphs. We implemented both methods in PYTHON 3 using the CVXPY tools from Diamond S. (2016) to solve the PSSE problem.

4.1 Example 1

We take the 14-bus system (case14) from the MAT-Power database. And we install 22 sensors for the network, where one of them collects values on the nodes and the other takes values on the lines. For more precision, see the Kekatos et al. (2017) article section 4, figure 2, we took a similar arrangement of sensors on a 14-bus power grid. Further, this system is called a block or a group, which is a practical one of the main reason to use ADMM approach and its accelerated variants instead of not augmented and not alternating gradient type procedures and its accelerated variants.

Fig. 1. Grid topology

from which we synthetically build an electrical network with challenging topology for distributed methods. The modeled network consists of 14-bus blocks. We take 10 copies of such blocks and bring them together into a chain that forms a ring. Thus, we get a closed chain with 10 "beads" as a 14-bus groups. In addition, every two consecutive groups are communicated by only one connecting line. Also, on this line, there is a sensor that takes readings. Thus, we solve the PSSE problem on a network of 10 groups of 14 buses, therefore the power grid comprises $10 \times 14 = 140$ nodes. To apply distributed methods, we divide the received network into areas. To do this, we separate each 14-bus group into 4 areas in the same way. Thus, the total network will be divided into 40 areas. Recall that the partition into regions occurs according to the following property. Each region overlaps with no more than two other regions and each bus can belong to no more than two areas. Without loss of generality, we can consider the obtained domains to be independent and solve the problem by distributed methods, the operation of which is discussed in Section 3.

The power grid with proposed topology is challenging for distributed methods as ADMM and accelerated ADMM because these methods effectively solve the problem of recovering values on a well-connected network. But in the example the nodes from the first group are poorly connected to the ones from 3d and further groups. The importance of the graph topology is written by Blondel and Tsitsiklis (1997). In addition, the assumption that a more connected graph is more suitable than a less connected graph is confirmed on the second example.

For the first example, we set the following parameters for the methods: $\mu = 1$, $\varepsilon_{\text{dual}} = 10^{-4}$, $\varepsilon_{\text{primal}} = 10^{-3}$, $\rho = 2$, where $\varepsilon_{\text{dual}}$ is a dual tolerance to achieve condition $u_{kl}^i = u_{kl}^{i+1}$ and $\varepsilon_{\text{primal}}$ is a primal tolerance to attain $v_k^i[l] = u_{kl}$ in each iteration.

For each method, we show in Figure 2 the logarithmic norm of primal residual for primal feasibility, which is
defined as \( r^i_k = v^k_l - u^k_l \). And we show the logarithmic norm of the dual residual for the dual feasibility condition, which is defined as \( s^i_k = u^k_l - u^k_{l-1} \) in Figure 2. We remark that the accelerated ADMM method significantly improves the convergence rate for primal and dual residuals in contrast with standard ADMM. In no more than 60 iterations, the accelerated method achieves convergence within the \( \varepsilon_{\text{primal}} = 10^{-3} \), which is not accessible by the ADMM method after 100 iterations.

4.2 Example 2

In this example, we change the network connection structure. Namely, we remove one edge connecting the first and last groups. Accordingly, we get no longer a ring of blocks, but a chain of 14 bus groups. Therefore, the power system is a network of 10 blocks of 14-bus groups connected in series with each other. Thus, the network relations between the blocks deteriorated. And setting the parameters of one block to communicate with another block became twice as weak as in Example 1.

As seen in Figure 1, the convergence of both methods sharply worsened due to a break in communication. Which clearly shows the importance of network topology.

4.3 Example 3

And now, for the power network from Example 1, we change the parameter \( \rho \) and see how it affects the convergence of the AADMM algorithm. As can be seen from Figure 3, the parameter \( \rho \) can be selected for better convergence, but after a certain value of \( \rho = 7 \) it again worsens the convergence. Thus, we can conclude that AADMM is sensitive to the choice of parameter \( \rho \).

5. CONCLUSION

In our paper, a novel approach to distributed state estimation was proposed. For considered DC linear measurement model we adapted accelerated ADMM in order to speed up the convergence of estimates to the exact minimizer. Numerical experiments exhibit higher convergence rate as compared to conventional ADMM, that confirms our theoretical results.

The future plans are to extend the proposed distributed approach for state estimation of multi-area interconnected power systems to 1) the case of asynchronous communications and 2) the AC case.

REFERENCES

Ren, W., and Beard, R. Distributed Consensus in Multi-Vehicle Cooperative Control. Berlin: Springer, ISBN 978-1-84800-014-8, 2008.

Antonelli, G. Interconnected dynamic systems: An overview on distributed control. IEEE Control Systems Magazine, volume 33, pages 76–88. 2013.

Gomez-Exposito, Antonio and Conejo, Antonio and Caizares, Claudio. Electric Energy Systems: Analysis and Operation. doi: 10.1201/9781420007275, 2008.
Goldstein, T., ODonoghue, B., Setzer, S., Du, Dajun and Li, Xue and Li, Wenting and Chen, Rui, Zhang, Runfan and Hredzak, Branislav. Nonlinear Sliding Mode and Distributed Control of Battery Energy Storage and Photovoltaic Systems in AC Microgrids with Time Delays. *IEEE Transactions on Industrial Informatics*, doi: 10.1109/TII.2019.2896032, 2018.

Minot, A., and Lu, Y. M. and Li, N. A distributed Gauss-Newton method for power system state estimation. *IEEE Transactions on Power Systems*, volume 31, pages 3804–3815, 2016.

Frasc, Paolo and Ishii, Hideaki and Ravazzi, Chiara and Tempo, Roberto. Distributed randomized algorithms for opinion formation, centrality computation and power systems estimation: A tutorial overview. *European journal of control*, volume 24, pages 2–13. 2015.

Kekatos, V., Wang, G., and Zhu, H., and Giannakis, G. B. PSSE Redux: Convex Relaxation, Decentralized, Robust, and Dynamic Approaches. *arXiv: 1708.03981*, pages 16–20, 2017.

Kekatos, V., and Giannakis, G. B. Distributed robust power system state estimation. *IEEE Transactions on Power Systems*, volume 28, pages 1617–1626. 2013.

Du, Dajun and Li, Xue and Li, Wenting and Chen, Rui and Fei, Minrui and Wu, Lei. ADMM-Based Distributed State Estimation of Smart Grid Under Data Deception and Denial of Service Attacks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2019.

Du, Dajun, and Li, Xue, and Li, Wenting, and Chen, Rui, and Fei, Minrui and Wu, Lei. ADMM-Based Distributed State Estimation of Smart Grid Under Data Deception and Denial of Service Attacks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, doi: 10.1109/TSMC.2019.2896292, volume-PP, 2019.

Cintuglu, Mehmet and Ishchenko, Dmitry. Secure Distributed State Estimation for Networked Microgrids. *IEEE Internet of Things Journal*, doi: 10.1109/JIOT.2019.2902793, volume-PP, pages 1–1, 2019.

Yuan, G. Lectures on Optimization Methods for Machine Learning. *H. Milton Stewart School of Industrial and Systems Engineering*, Georgia Institute of Technology, Atlanta, GA, 2019.

Goldstein, T., and ODonoghue, B., and Setzer, S., and Baraniuk, R. Fast alternating direction optimization methods. *SIAM Journal on Imaging Sciences*, volume-7, pages 1588-1623, 2014.

Roald, L., and Andersson, G. Chance-Constrained AC Optimal Power Flow: Reformulations and Efficient Algorithms. *IEEE Transactions on Power Systems*, volume-33, pages 2906 – 2918, 2017.

Stott, B., and Jardim, J., and Alsac, O. DC Power Flow Revisited. *IEEE Transactions on Power Systems*, volume-24, pages 1290 – 1300, 2009.

Chertkov, M., and Pan, F., and Stepanov, M. G. Predicting Failures in Power Grids: The Case of Static Overloads. *IEEE Transactions on Smart Grid*, volume-2, pages 162 – 172, 2011.

Monticelli, A. Electric power system state estimation. *Proceedings of the IEEE*, volume-88, pages 262 – 282, 2000.

Sule, P., and Backhaus, S., and Chertkov, M. Optimal Distributed Control of Reactive Power via the Alternating Direction Method of Multipliers. *IEEE Transactions on Energy Conversion*, volume-29, pages 968 – 977, 2014.

Boyd, S., and Parikh, N., and Chu, E., and Peleato, B., and Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, volume-3, pages 1–122, 2011.

Davis, D. and Yin, W. Convergence Rate Analysis of Several Splitting Schemes. *Splitting Methods in Communication, Imaging, Science, and Engineering*, Springer, Cham, pages 115–163, 2016.

Ahmadi, M., and Cubuktepe, M., and Topcu, U., and Tanaka, T. Distributed Synthesis Using Accelerated ADMM. *Annual American Control Conference (ACC)*, 2018.

Frau, G., and Robinson, D. P., and Vidal, R. ADMM and Accelerated ADMM as Continuous Dynamical Systems. *arXiv preprint, arXiv:1805.06579*, 2018.

Nesterov, Y. A method of solving a convex programming problem with convergence rate O(1/k^2). *Soviet Mathematics Doklady*, volume-27, pages 372–376, 1983.

Guminov, S., and Dvurechensky, P., and Gasnikov, A. Accelerated Alternating Minimization. *arXiv preprint arXiv:1906.03622*, 2019.

Ouyang, Y., and Chen, Y., and Lan, G., and Pasiliao, E. An accelerated linearized alternating direction method of multipliers. *SIAM Journal on Imaging Sciences*, volume-8, pages 644–681, 2015.

Diamond S., Boyd S. CVXPY: A Python-Embedded Modeling Language for Convex Optimization. *Journal of Machine Learning Research*, 2016.

Blondel, V., and Tsitsiklis, J. N. NP-hardness of some linear control design problems. *SIAM Journal on Control and Optimization*, volume-35, pages 2118–2127, 1997.

6. APPENDIX

We prove the Theorem 1 in a similar way as in Lan (2019). Let the primal-dual gap function Q be

\[ Q(t_k, v_k) = f_k(v_k) + \sum_{l \in B_k} \langle y_{kl}, v_k[l] - u_{kl} \rangle - f_k(v_k) \]

\[ - \sum_{l \in B_k} \langle y_{kl}, v_k[l] - u_{kl} \rangle, \]

where \( t_k \equiv (v_k, u_{kl}, y_{kl}) \) and \( v_k \equiv (v_k, u_{kl}, y_{kl}) \).

Proof.

1) Due to the optimality conditions of the

\[ v_k^i = \min_{v_k \in K} f_k(v_k) + \sum_{l \in B_k} y_{kl}^{i-1}(v_k[l] - u_{kl}^{i-1}) \]

\[ + \sum_{l \in B_k} \frac{\mu}{2} \| v_k[l] - u_{kl}^{i-1} \|^2, \]

we obtain
which together with the equation \( y_{kl}^i = y_{kl}^{i-1} + \mu(v_{kl}^i - u_{kl}^{i-1}) \) represents that

\[
Q(t_k^i, t_k) \leq \frac{1}{2\mu} \sum_{l \in B_k} (\|y_{kl}^i - y_{kl}^{i-1}\|^2 - \|y_{kl} - y_{kl}^i\|^2)
- \frac{1}{2\mu} \sum_{l \in B_k} (\|y_{kl}^{i-1} - y_{kl}\|^2)
+ \frac{\mu}{2} \sum_{l \in B_k} ||y_{kl}^i - u_{kl}^i||^2 - ||y_{kl} - u_{kl}||^2
+ \frac{\mu}{2} \sum_{l \in B_k} (\|y_{kl}^{i-1} - u_{kl}^{i-1}\|^2 - \|y_{kl} - u_{kl}\|^2).
\]

As a result, we have

\[
f_k(\bar{v}_k) - f_k(v_k) = \frac{t}{2} \sum_{l \in B_k} ||v_{kl}^i - u_{kl}^i||^2 - ||v_{kl} - u_{kl}||^2.
\]
\[ f_k(\bar{v}_k) - f_k(\bar{v}_k^*) + \sum_{l \in B_k} \langle \bar{y}_{kl}, \bar{v}_k[l] - \bar{u}_{kl} + \frac{1}{\mu N}(\bar{y}_{kl}^0 - \bar{y}_{kl}^N) \rangle \]
\begin{align*}
&\leq \frac{1}{2 \mu N} \sum_{l \in B_k} \left( \|\bar{y}_{kl}^0\|_2^2 - \|\bar{y}_{kl}^N\|_2^2 \right) \\
&\quad + \frac{\mu}{2N} \sum_{l \in B_k} \left( \|\bar{u}_{kl}^0 - \bar{u}_{kl}^*\|_2^2 - \|\bar{u}_{kl}^N - \bar{u}_{kl}^*\|_2^2 \right) \\
&\leq \frac{1}{2N} \left( \frac{1}{\mu} \sum_{l \in B_k} \|\bar{y}_{kl}\|_2^2 + \mu \sum_{l \in B_k} \|\bar{u}_{kl}^0 - \bar{u}_{kl}^*\|_2^2 \right) \\
&\quad = \frac{\mu}{2N} \sum_{l \in B_k} \|\bar{u}_{kl}^0 - \bar{u}_{kl}^*\|_2^2,
\end{align*}
for all \( u_{kl}, l \in B_k, k = 1, \ldots, K \). In the last equality, we took into account that we initialized \( y_{kl}^0 = 0 \). Further, from this relation follows, that \( v_k[l] - v_k + \frac{1}{\mu N}(y_{kl}^0 - y_{kl}^N) = 0 \), and hence the first equation in (7) is satisfied. Also, from the last and (12) follow the second equation from (7).

\[ \|\bar{v}_k[l] - \bar{u}_{kl}\|_2 \leq \frac{1}{\mu N} \left[ \frac{2}{\mu} \|\bar{y}_{kl}^0 - y_{kl}^*\|_2 \right. \] 
\begin{align*}
&\quad + \left. \|\bar{u}_{kl}^0 - \bar{u}_{kl}^*\|_2 \right] \\
&\quad = \frac{1}{N} \left[ \frac{2}{\mu} \|\bar{y}_{kl}^0 - y_{kl}^*\|_2 + \|\bar{u}_{kl}^0 - \bar{u}_{kl}^*\|_2 \right] \\
&\quad = \frac{1}{N} \left[ \frac{2}{\mu} \|\bar{y}_{kl}^*\|_2 + \|\bar{u}_{kl}^0 - \bar{u}_{kl}^*\|_2 \right].
\end{align*}