# A Class of Distributed Optimization Methods with Event-Triggered Communication

**Martin C. Meinel · Michael Ulbrich · Sebastian Albrecht**

the date of receipt and acceptance should be inserted later

**Abstract** We present a class of methods for distributed optimization with event-triggered communication. To this end, we extend Nesterov's first order scheme to use event-triggered communication in a networked environment. We then apply this approach to generalize the proximal center algorithm (PCA) for separable convex programs by Necoara and Suykens. Our method uses dual decomposition and applies the developed event-triggered version of Nesterov's scheme to update the dual multipliers. The approach is shown to be well suited for solving the active optimal power flow (DC-OPF) problem in parallel with event-triggered and local communication. Numerical results for the IEEE 57 bus and IEEE 118 bus test cases confirm that approximate solutions can be obtained with significantly less communication while satisfying the same accuracy estimates as solutions computed without event-triggered communication.

**Keywords** distributed optimization · convex optimization · non-differentiable optimization · dual decomposition · distributed regularization · local communication · event-triggered communication · DC-OPF problem · IEEE test cases

## 1 Introduction

**Motivation.** Distributed optimization gained a lot of attention in recent years to face the need of fast and efficient solutions for problems arising in the context of large-scale networks such as utility maximization problems (NUM) [13, 14, 23], distributed estimation [17, 18], multi-robot coordination [3, 15], and the optimal power flow (OPF) problem [1, 6, 7, 9]. In this paper we focus on the OPF problem, but other applications are possible. The goal of distributed optimization is to solve these problems in parallel by multiple agents that *jointly* minimize (or maximize) a separable objective function, usually subject to

Martin Meinel
Lehrstuhl für Mathematische Optimierung, Technische Universität München, Fakultät für Mathematik
Boltzmannstr. 3, 85748 Garching b. München, Germany
E-mail: meinel@ma.tum.de

Michael Ulbrich
Lehrstuhl für Mathematische Optimierung, Technische Universität München, Fakultät für Mathematik
Boltzmannstr. 3, 85748 Garching b. München, Germany
E-mail: mulbrich@ma.tum.de

Sebastian Albrecht
Lehrstuhl für Mathematische Optimierung, Technische Universität München, Fakultät für Mathematik
Boltzmannstr. 3, 85748 Garching b. München, Germany
E-mail: albrecht@ma.tum.de

coupling constraints that force them to exchange information, such as their assigned components of the decision variable, during the optimization process. For details we refer to [2, 11, 24].
Often it is advantageous if the communication of cooperating agents is minimal with regards to the frequency and amount of transmitted information. This is necessary for example when ad hoc wireless communication is used where the amount of information that can be sent is limited [22]. Moreover, it is preferable to keep the communication local, i.e., between agents that are in close proximity (often indicated by a connecting edge in the graph representation of the network) to avoid long communication times and data package losses. With these issues in mind, we consider the proximal center algorithm (PCA) by Necoara and Suykens [10] which is a dual decomposition scheme for convex programs with separable structure and where the objective functions are not required to be strictly convex or smooth. The two key ingredients of the algorithm are basically a smoothing technique proposed by Nesterov in [12] to obtain a smooth dual augmented function that is separable in the primal variables, and the usage of Nesterov's optimal first order scheme for smooth optimization [12] to update dual multipliers. Generalizing the approaches of [12] and [10], we equip Nesterov's algorithm with an extension that allows event-triggered communication where the agents exchange information in a non-periodic manner only when it is crucially required in order to maintain convergence, and show that the algorithm is implementable in parallel under suitable conditions. Furthermore, we prove convergence of the developed distributed Nesterov algorithm with event-triggered communication (DNA-EC) for the minimization of a continuously differentiable convex objective function with Lipschitz continuous gradient, maintaining the convergence rate of the original scheme. Finally, the application of the DNA-EC allows us to implement the PCA in an entirely parallel manner yielding the distributed proximal center algorithm with event-triggered communication (DPCA-EC). Moreover, we manage here as well to maintain the convergence rate of the PCA which is of the order $\mathcal{O}(1/\epsilon)$, where $\epsilon$ is the desired accuracy of the objective function value at the approximate solution.
We apply the DPCA-EC to solve the DC-OPF problem [16], which comprises the optimal active power generation dispatch in a power system, and obtain approximate solutions under local and event-triggered communication, significantly reducing communication between agents.
As the notion of event-triggered communication finds its origin in the context of networked control systems [8] it appears only lately in direct connection with distributed optimization and therefore a limited amount of literature exists concerning distributed optimization with event-triggered communication: Zhong and Cassandras extend in [25] a partially asynchronous distributed optimization framework from [2] by event-triggered communication. Their algorithm is applicable for an unconstrained problem with continuously differentiable objective function that has a Lipschitz continuous gradient and follows a general state update scheme, where the current iterate of an agent's state is updated by a scaled update direction (e.g., a gradient step with a certain step size). They show convergence to a stationary point of the objective function.
Wan and Lemmon proposed a distributed algorithm in [21] with event-triggered communication based on a barrier approach. Recognizing drawbacks of the algorithm such as ill-conditioning, they formulate in [20] an augmented Lagrangian function associated with the NUM problem that is implementable distributively with event-triggered communication between users and links, and where the user rates are updated in a steepest ascent fashion. Assuming the utility functions to be twice differentiable, strictly increasing and strictly concave, they can show asymptotic convergence of the algorithm. In [22] Wan and Lemmon use the same approach to solve the DC-OPF problem in a distributed manner with event-triggered communication and show asymptotic convergence assuming strictly increasing, strictly convex and differentiable power generation cost functions.
In contrast to the existing approaches on the one hand the DNA-EC is applicable to problems with convex and continuously differentiable objective function constrained by a block-separable set. On the other hand the DPCA-EC is applicable to problems with separable and convex but not necessarily continuously differentiable objective function constrained by coupled linear constraints. Moreover, as a heritage of their

origins (cf. [12] and [10]), the complexities of the DNA-EC and the DPCA-EC are superior to the complexities of standard gradient projection methods and dual gradient schemes, respectively, which can be implemented in parallel as well. Finally, accuracy estimates for both algorithms are given in this work.

**Contents.** In section 2 we review Nesterov's optimal first order scheme [12] for minimizing a convex function with Lipschitz continuous gradient over a convex and closed set. We show for a suitable setting that a slightly modified version of the algorithm is implementable in parallel with event-triggered communication, yielding our algorithm DNA-EC. We prove its convergence maintaining the complexity $O(\sqrt{L/\epsilon})$ of the original scheme, where $L$ is the Lipschitz constant of the gradient and $\epsilon$ the desired accuracy of the objective function value at the approximate solution. In [12] it is mentioned that this complexity is superior compared to the complexity of the standard gradient projection method with $O(1/\epsilon)$.
In section 3 we review the PCA [10] of Necoara and Suykens and propose a scaling technique to improve the accuracy estimates for the algorithm. Moreover, we modify the PCA by applying the developed DNA-EC to update dual multipliers in parallel and with event-triggered communication, yielding the algorithm DPCA-EC. We show convergence of the DPCA-EC preserving the complexity $O(1/\epsilon)$ of the PCA which is better than the complexity of classical dual gradient schemes with $O(1/\epsilon^2)$ [10]. Moreover, we show how to optimally choose the convexity parameters for a certain choice of strongly convex prox-functions that are used to smoothen the dual function of the considered convex problem. Additionally, we show how to optimally choose the smoothing parameters that are used to scale these prox-functions.
Finally, in section 4 we describe in detail the DC-OPF problem and present numerical results of the application of the DPCA-EC to the IEEE 57 bus and IEEE 118 bus test cases [4], showing that the communication exchange can be reduced significantly without trading off accuracy of the approximate solution.

## 2 Optimal first order scheme with event-triggered communication

### 2.1 Nesterov's optimal scheme for smooth minimization

In [12] Nesterov provides a first order method for optimization problems of the form

$$\min_{x\in Q} f(x) \tag{1}$$

where $f\colon Q \to \mathbb{R}$ is a convex and continuously differentiable function on a closed and convex set $Q \subseteq \mathbb{R}^m$. Further, it is assumed that the gradient of $f$ is Lipschitz continuous:

$$||\nabla f(x) - \nabla f(y)|| \leq L\,||x-y|| \quad \forall x, y \in Q, \tag{2}$$

where $||\cdot||$ denotes the Euclidean norm. Choosing a prox-function $d(x)$, which is a continuous and strongly convex function on Q with convexity parameter $\sigma > 0$ (i.e., it satisfies $d(y) \geq d(x) + \nabla d(x)^T(y-x) + \frac{\sigma}{2}||x-y||^2$ for all $x, y \in Q$), the corresponding center

$$x^0 = \arg\min_{x\in Q} d(x) \tag{3}$$

of the set Q is determined as the starting value for Nesterov's optimization scheme that may be stated as follows:

**Algorithm 2.1.1** *For $k \geq 0$ do*

1. *Compute $\nabla f(x^k)$.*
2. *Find $y^k = \arg\min_{y \in Q} \left\{ \left\langle \nabla f(x^k), y - x^k \right\rangle + \frac{L}{2} \left\| y - x^k \right\|^2 \right\}$.*
3. *Find $z^k = \arg\min_{z \in Q} \left\{ \frac{L}{\sigma} d(z) + \sum_{j=0}^{k} \alpha_j \left\langle \nabla f(x^j), z - x^j \right\rangle \right\}$.*
4. *Set $x^{k+1} = \tau_k z^k + (1 - \tau_k) y^k$.*

Here $\{\alpha_k\}_{k \geq 0}$ are a priori chosen positive step size parameters and $\tau_k = \alpha_{k+1}/A_{k+1}$ with $A_k = \sum_{i=0}^{k} \alpha_i$. Note, that we omitted some constant terms in the argmin-problems of the original algorithm in [12]. We did so to reveal the parallelizable nature that is obvious in Algorithm 2.1.1 if the prox-function $d(x)$ is chosen appropriately, e.g., $d(x) = (\sigma/2)\|x\|^2$ and if $Q$ has a block-separable structure, i.e., if $Q = Q_1 \times \ldots \times Q_s$ with $Q_l \subseteq \mathbb{R}^{m_l}$ and $\sum_{l=1}^{s} m_l = m$.
The following convergence result holds:

**Theorem 2.1.2** [Lemma 1 of [12]] *Let the sequence $\{\alpha_k\}_{k \geq 0}$ satisfy the condition:*

$$\alpha_0 \in (0,\ 1],\ {\alpha_{k+1}}^2 \leq A_{k+1},\ \alpha_k > 0,\ k \geq 0. \tag{4}$$

*Then the relation*

$$A_k f(y^k) \leq \Psi^k = \min_{z \in Q} \left\{ \frac{L}{\sigma} d(z) + \sum_{j=0}^{k} \alpha_j \left( f(x^j) + \left\langle \nabla f(x^j), z - x^j \right\rangle \right) \right\}$$

*holds for $k \geq 0$ and therefore $f(y^k) - f(x^*) \leq L d(x^*)/A_k$, where $x^*$ is an optimal solution to problem (1).*

The following lemma gives a possible choice for $\{\alpha_k\}_{k \geq 0}$ that provides the complexity estimate of $O\left(\sqrt{L/\epsilon}\right)$, where $\epsilon$ is the desired accuracy of the objective function value at the approximate solution:

**Lemma 2.1.3** [Lemma 2 of [12] *For $k \geq 0$ define $\alpha_k = (k+1)/2$. Then*

$$\tau_k = \frac{2}{k+3}, \quad A_k = \frac{(k+1)(k+2)}{4},$$

*and conditions (4) are satisfied.*

**Remark 2.1.4** *Let us note that the following equality for $y^k$ computed in step 2 of Algorithm 2.1.1 holds:*

$$\begin{aligned} y^k &= \arg\min_{y \in Q} \left\{ \frac{2}{L} \left\langle \nabla f(x^k), y - x^k \right\rangle + \left\| y - x^k \right\|^2 + \frac{1}{L^2} \left\| \nabla f(x^k) \right\|^2 \right\} \\ &= \arg\min_{y \in Q} \left\{ \left\| y - x^k + \frac{1}{L} \nabla f(x^k) \right\|^2 \right\}. \end{aligned}$$

*In other words, the second step of Algorithm 2.1.1 is a projected gradient step with step size $1/L$. In a sum, the iterate $x^{k+1}$ in step 4 is a convex combination of $z^k$ and $y^k$ with growing impact of the projected gradient step if the sequence $\{z^k\}_{k \geq 0} \subseteq Q$ is bounded.*

## 2.2 Distributed Nesterov Algorithm with event-triggered communication

To approach a multi-agent framework that allows a parallel implementation of Nesterov's optimal first order scheme 2.1.1, we make the following assumptions:

**Assumptions 2.2.1**

1. *Besides being convex and closed, the set $Q \in \mathbb{R}^m$ is bounded and block-separable:*

$$Q = Q_1 \times \ldots \times Q_s \;\; \textit{with} \;\; Q_l \subset \mathbb{R}^{m_l} \;\; \textit{and} \;\; \sum_{l=1}^{s} m_l = m.$$

2. *The a priori chosen prox-function $d(x)$ is separable in the same way as $Q$:*

$$d(x) = \sum_{l=1}^{s} d_l(x_l) \;\; \textit{with} \;\; x_l \in Q_l \;\; \textit{for} \;\; l = 1, \ldots, s,$$

*where $d_l : \mathbb{R}^{m_l} \to \mathbb{R}$ is a strongly convex function with convexity parameter $\sigma$.*

Assumptions 2.2.1 suggests to divide the decision variables into $s$ sub-blocks, i.e., $x = (x_1, \ldots, x_s) \in Q \subseteq \mathbb{R}^m$ with $x_l \in \mathbb{Q}_l \subseteq \mathbb{R}^{m_l}$. Moreover, we denote in the following by $agent_l$ the agent that is responsible for updating sub-block $x_l$. In many cases it can be observed that that sub-block $l$ of the gradient $\nabla f(x) \in \mathbb{R}^m$, which is denoted by $\nabla_l f(x) \in \mathbb{R}^{m_l}$, does not depend on the whole vector $x \in Q$, i.e., $agent_l$ needs to communicate only with the agents that control the sub-blocks of $x$ necessary to compute $\nabla_l f(x)$. We describe this relation by a information dependency graph (IDG) which defines the communication between the agents that correspond to the nodes of this undirected graph. A more formal description of this observation may be the following for $l = 1, \ldots, s$:

**Assumption 2.2.2**

$$\nabla_l f(x) = \nabla_l f(y) \Leftrightarrow x_j = y_j \;\; \textit{for all} \;\; j \in N_{IDG}(l) \cup \{l\}, \;\; \textit{and} \;\; x, y \in Q.$$

Here $N_{IDG}(l) = \{j_1, \ldots, j_{\eta_l}\}$ denotes the set of indices of $agent_l$'s $\eta_l$ neighbors in the IDG. The crucial point is that the maximal degree of the IDG, which will appear in the convergence result of the DNA-EC, can be assumed to be independent of the network size for problems arising in large scale networks as the structure of the objective function is usually independent of the network size.

Moreover, especially in the case of spaciously distributed multi-agent networks, it is advantageous to reduce the communication traffic to the minimum by introducing event-triggered communication. Here information is transmitted by an agent in iteration $k$ only if the measured deviation of its content from earlier sent information during the optimization process exceeds a certain threshold $\Delta_k$. This threshold has to be chosen appropriately to ensure overall convergence. To introduce event-triggered communication in a distributed version of Algorithm 2.1.1, we assume in the following without restriction that $0 \in Q_l$ and define similar to [22] and [25] for $l = 1, \ldots, s$ and $k \geq 0$ by

$$x^{l,k} = \left(x_1^{l,k}, \ldots, x_s^{l,k}\right) \in Q \subset \mathbb{R}^m, \;\; \text{where} \;\; \begin{cases} \left\|x_j^{l,k} - x_j^k\right\|_1 \leq \Delta_k \in \mathbb{R}_+, & \text{if } j \in N_{IDG}(l), \\ x_j^{l,k} = 0, & \text{if } j \notin N_{IDG}(l) \cup \{l\}, \\ x_l^{l,k} = x_l^k, & \end{cases} \tag{5}$$

the iterates containing the possibly outdated versions of the sub-blocks $x_{j_1}^k, \ldots, x_{j_{\eta_l}}^k$ for $k \geq 0$ that are available to $agent_l$ who updates sub-block $x_l^k$. The threshold $\Delta_k$ in (5) defines how far the outdated information $x_j^{l,k}$ is allowed to deviate from the latest information $x_j^k$ that $agent_j$ with $j \in N_{IDG}(l)$ holds. If this threshold is exceeded, $agent_j$ transmits the data to $agent_l$ in order to ensure convergence of the distributed algorithm. In the following we will set $\Delta_0 = 0$ which is natural as there is no outdated information at the beginning of the optimization process, i.e., every agent starts with the same information.

Finally under the above assumptions the DNA-EC can be stated as follows:

**Algorithm 2.2.3 (DNA-EC)** *For $l = 1,\dots,s$ and $k \geq 0$ do in parallel*

1. *Compute* $\nabla_l f(x^{l,k})$.
2. *Find* $\tilde{y}_l^k = \underset{y_l \in Q_l}{\arg\min} \left\{ \left\langle \nabla_l f(x^{l,k}), y_l - x_l^k \right\rangle + L\varDelta_k \eta_l \left\| y_l - x_l^k \right\|_1 + \frac{L}{2} \left\| y_l - x_l^k \right\|^2 \right\}$.
3. *Find* $\tilde{z}_l^k = \underset{z_l \in Q_l}{\arg\min} \left\{ \frac{L}{\sigma} d_l(z_l) + \sum_{j=0}^{k} \alpha_j \left\langle \nabla_l f(x^{l,j}), z_l - x_l^j \right\rangle \right\}$.
4. *Set* $x_l^{k+1} = \tau_k \tilde{z}_l^k + (1-\tau_k)\tilde{y}_l^k$.
5. *Exchange information: For* $j = 1,\dots,s$

   ***if*** $l \in N_{IDG}(j)$ ***and*** $\left\| x_l^{j,k} - x_l^{k+1} \right\|_1 > \varDelta_{k+1}$ ***then***

   $x_l^{j,k+1} = x_l^{k+1}$,

   ***else***

   $x_l^{j,k+1} = x_l^{j,k}$.

Note that we added the term $L\varDelta_k \eta_l \left\| y_l - x_l^k \right\|_1$ in step 2 to be able to prove convergence of the DNA-EC which coincides with Algorithm 2.1.1 if $\varDelta_k = 0$ for $k \geq 0$. Furthermore, note that $\nabla_l f(x^{l,k})$ has to be computed in the first step only if $x^{l,k} \neq x^{l,k-1}$ for $l \in \{1,\dots,s\}$ and $k \geq 1$, i.e., event-triggered communication additionally yields a reduction of computational effort. Also, in a lot of cases the minimization problems in step 2 and 3 can be solved easily. If for instance the feasible set Q is component-wisely separable and the prox-function $d(z)$ is chosen according to $d(z) = (\sigma/2)\|z\|^2$, it is straightforward to determine the closed form solutions of these problems:

**Example 2.2.4** *Let $Q = Q_1 \times Q_2 \times \dots \times Q_n$ with compact and convex sets $Q_i \subseteq \mathbb{R}$ and $d(z) = (\sigma/2)\|z\|^2$. Then the problems in step 2 of Algorithm 2.2.3 can be written as:*

$$\tilde{y}_l^k = \underset{y_l \in Q_l \subseteq \mathbb{R}}{\arg\min} \left\{ \nabla_l f(x^{l,k})(y_l - x_l^k) + L\varDelta_k \eta_l |y_l - x_l^k| + \frac{L}{2}(y_l - x_l^k)^2 \right\}.$$

*With $Q_l = [\underline{Q}_l, \overline{Q}_l]$ and $\varDelta_L = L\varDelta_k \eta_l$, the minimum $\tilde{y}_l^k$ over $Q_l$ for $l = 1,\dots,m$ is:*

$$\tilde{y}_l^k = \max\left\{ \min\left\{ y_l^*, \overline{Q}_l \right\}, \underline{Q}_l \right\} \quad \text{where} \quad y_l^* = \begin{cases} -\frac{\nabla_l f(x^{l,k}) + \varDelta_L}{L} + x_l^k, & \text{if } -\frac{\nabla_l f(x^{l,k}) + \varDelta_L}{L} \geq 0, \\ -\frac{\nabla_l f(x^{l,k}) - \varDelta_L}{L} + x_l^k, & \text{if } -\frac{\nabla_l f(x^{l,k}) - \varDelta_L}{L} \leq 0, \\ x_l^k, & \text{else.} \end{cases}$$

*Accordingly, the solutions of the problems in step 3 are:*

$$\tilde{z}_l^k = \max\left\{ \min\left\{ z_l^*, \overline{Q}_l \right\}, \underline{Q}_l \right\} \quad \text{for} \quad l = 1,\dots,m,$$

*with* $z_l^* = -\frac{\sum_{j=0}^{k} \alpha_j \nabla_l f(x^{l,j})}{L}$.

Before we can prove the convergence of the DNA-EC we have to provide three lemmas starting with the first where we make use of the Lipschitz continuity assumption of the gradient $\nabla f(x)$ similarly to [12]:

**Lemma 2.2.5** *For $y$, $x^k$, $x^{l,k} \in Q$ and $k \geq 0$ the following inequality holds:*

$$f(y) \leq f(x^k) + \sum_{l=1}^{s} \left\langle \nabla_l f(x^{l,k}), y_l - x_l^k \right\rangle + L\varDelta_k \sum_{l=1}^{s} \eta_l \left\| y_l - x_l^k \right\|_1 + \frac{L}{2} \left\| y - x^k \right\|^2.$$

*Proof* The Lipschitz continuity assumption of $\nabla f(x)$ is equivalent to:

$$f(y) \le f(x) + \langle \nabla f(x),\ y - x\rangle + \frac{L}{2}\|y-x\|^2 \quad \forall x, y \in Q.$$

With the definition of $x^{l,k}$ in (5) and Assumption 2.2.2 we have for $x^k, y \in Q$:

$$\begin{aligned}
f(y) &\le f(x^k) + \left\langle \nabla f(x^k),\ y - x^k \right\rangle + \frac{L}{2}\left\|y - x^k\right\|^2 \\
&= f(x^k) + \sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,k}), y_l - x_l^k \right\rangle + \sum_{l=1}^{s}\left\langle \nabla_l f(x^k) - \nabla_l f(x^{l,k}), y_l - x_l^k \right\rangle + \frac{L}{2}\left\|y - x^k\right\|^2 \\
&\le f(x^k) + \sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,k}), y_l - x_l^k \right\rangle + \\
&\quad L\sum_{l=1}^{s}\left\|(x_{j_1}^k,\dots,x_{j_{\eta_l}}^k) - (x_{j_1}^{l,k},\dots,x_{j_{\eta_l}}^{l,k})\right\|\left\|y_l - x_l^k\right\| + \frac{L}{2}\left\|y - x^k\right\|^2 \qquad (6)\\
&\le f(x^k) + \sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,k}), y_l - x_l^k \right\rangle + L\Delta_k\sum_{l=1}^{s}\eta_l\left\|y_l - x_l^k\right\|_1 + \frac{L}{2}\left\|y - x^k\right\|^2,
\end{aligned}$$

where we use the Lipschitz continuity of the gradient of $f$ similar to [2] to obtain (6). □

**Lemma 2.2.6** *Applying Lemma 2.2.5 to $\tilde{y}^k$ computed in step 2 of the DNA-EC we obtain for $k \ge 0$:*

$$f(\tilde{y}^k) \le f(x^k) + \min_{y\in Q}\left\{\sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,k}), y_l - x_l^k \right\rangle + L\Delta_k\sum_{l=1}^{s}\eta_l\left\|y_l - x_l^k\right\|_1 + \frac{L}{2}\left\|y - x^k\right\|^2\right\}.$$

Let us define the constant

$$\rho_k = \eta_{l^*} L C \sum_{j=0}^{k}\alpha_j \Delta_j \quad \text{for} \quad k \ge 0, \qquad (7)$$

where $\eta_{l^*} = \max_{l=1,\dots,s}\eta_l$ is the maximum number of neighbors of each agent in the IDG and $C$ is the diameter of the closed, convex and bounded set Q with respect to the 1-norm:

$$C = \max_{y,x\in Q}\|x-y\|_1. \qquad (8)$$

Obviously, one has $\rho_0 = 0$ as we assumed $\Delta_0 = 0$, i.e., the agents start with the same information. For $k \ge 0$ we set

$$\widetilde{\Psi}^k = \min_{z\in Q}\left\{\rho_k + \frac{L}{\sigma}d(z) + \sum_{j=0}^{k}\alpha_j\left(f(x^j) + \sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,j}), z_l - x_l^j \right\rangle\right)\right\}, \qquad (9)$$

$$\tilde{z}^k = \arg\min_{z\in Q}\left\{\rho_k + \frac{L}{\sigma}d(z) + \sum_{j=0}^{k}\alpha_j\left(f(x^j) + \sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,j}), z_l - x_l^j \right\rangle\right)\right\}, \qquad (10)$$

where $\tilde{z}^k$ coincides with the iterate that has to be computed in step 3 of the DNA-EC. Finally, we will describe the error that occurs in the convergence proof of the DNA-EC due to the usage of event-triggered communication with

$$E_k = A_{k-1}\tau_{k-1}\sum_{l=1}^{s}\left\langle \nabla_l f(x^k) - \nabla_l f(x^{l,k}), \tilde{y}_l^{k-1} - \tilde{z}_l^{k-1} \right\rangle \quad \text{for} \quad k \ge 1, \qquad (11)$$

and set $E_0 = 0$. Let the starting point $x^0$ be chosen according to (3). Without loss of generality it can be assumed that $d(x^0) = 0$ (otherwise choose $\hat{d}(x) = d(x) - d(x^0)$ instead). In the following lemma we prove a relation between the sequences of points $\{x^k\}_{k\geq 0}$, $\{\tilde{y}^k\}_{k\geq 0}$, and $\{\tilde{z}^k\}_{k\geq 0}$ generated by the DNA-EC, building the basis of our convergence result:

**Lemma 2.2.7** *Let the sequence* $\{\alpha_k\}_{k\geq 0}$ *satisfy the condition:*

$$\alpha_0 \in (0,\ 1],\ {\alpha_{k+1}}^2 \leq A_{k+1},\ k \geq 0 \tag{12}$$

*and set*

$$x^{k+1} = \tau_k \tilde{z}^k + (1-\tau_k)\tilde{y}^k. \tag{13}$$

*Then the relation*

$$\widetilde{\Psi}^k \geq A_k f(\tilde{y}^k) + \sum_{j=0}^{k} E_j \tag{14}$$

*holds for* $k \geq 0$.

*Proof* The proof extends that of Lemma 1 in [12] and uses induction. For $k = 0$ we have

$$d(z) \geq \underbrace{d(x^0)}_{=0} + \underbrace{\nabla d(x^0)^T (z - x^0)}_{\geq 0} + \frac{\sigma}{2}\left\|z - x^0\right\|^2 \geq \frac{\sigma}{2}\left\|z - x^0\right\|^2,$$

due to the strongly convexity of $d(z)$. It follows that

$$\begin{aligned}
\widetilde{\Psi}^0 &= \min_{z\in Q}\left\{\underbrace{\rho_0}_{=0} + \frac{L}{\sigma}d(z) + \alpha_0\left(f(x^0) + \sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,0}), z_l - x_l^0\right\rangle\right)\right\} \\
&\geq \alpha_0 \min_{z\in Q}\left\{\frac{L}{2\alpha_0}\left\|z - x^0\right\|^2 + f(x^0) + \sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,0}), z_l - x_l^0\right\rangle\right\} \\
&\geq \alpha_0 f(\tilde{y}^0) = A_0 f(\tilde{y}^0) + \underbrace{E_0}_{=0},
\end{aligned}$$

where the last inequality follows with Lemma 2.2.6. Now assume that the relation $\widetilde{\Psi}^k \geq A_k f(\tilde{y}^k) + \sum_{j=0}^{k} E_j$ holds for some $k \in \mathbb{N}_0$. As the function

$$h_k(z) := \rho_k + \frac{L}{\sigma}d(z) + \sum_{j=0}^{k}\alpha_j\left(f(x^j) + \sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,j}), z_l - x_l^j\right\rangle\right)$$

is strongly convex with convexity parameter $L$ for $k \geq 0$, we have

$$\begin{aligned}
\widetilde{\Psi}^{k+1} &= \min_{z\in Q}\left\{\rho_k + \eta_{l^*} LC\alpha_{k+1}\Delta_{k+1} + \frac{L}{\sigma}d(z) + \sum_{j=0}^{k+1}\alpha_j\left(f(x^j) + \sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,j}), z_l - x_l^j\right\rangle\right)\right\} \\
&\geq \min_{z\in Q}\left\{\widetilde{\Psi}^k + \frac{L}{2}\left\|z - \tilde{z}^k\right\|^2 + \eta_{l^*} LC\alpha_{k+1}\Delta_{k+1} + \alpha_{k+1}\left(f(x^{k+1}) + \sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,k+1}), z_l - x_l^{k+1}\right\rangle\right)\right\} \\
&\geq \min_{z\in Q}\left\{\widetilde{\Psi}^k + \frac{L}{2}\left\|z - \tilde{z}^k\right\|^2 + \eta_{l^*} L\alpha_{k+1}\Delta_{k+1}\left\|z - \tilde{z}^k\right\|_1 + \right. \\
&\qquad \left.\alpha_{k+1}\left(f(x^{k+1}) + \sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,k+1}), z_l - x_l^{k+1}\right\rangle\right)\right\}.
\end{aligned}$$

Due to the convexity of $f$, the definition of $x^{k+1}$ in (13), and the induction hypothesis, we have

$$
\begin{aligned}
&\widetilde{\Psi}^k + \alpha_{k+1}\left(f(x^{k+1}) + \sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,k+1}), z_l - x_l^{k+1}\right\rangle\right) \\
&\geq A_k f(\tilde{y}^k) + \alpha_{k+1}\left(f(x^{k+1}) + \sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,k+1}), z_l - x_l^{k+1}\right\rangle\right) + \sum_{j=0}^{k} E_j \\
&\geq A_k\left(f(x^{k+1}) + \left\langle \nabla f(x^{k+1}),\ \tilde{y}^k - x^{k+1}\right\rangle\right) + \alpha_{k+1}\left(f(x^{k+1}) + \sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,k+1}), z_l - x_l^{k+1}\right\rangle\right) + \sum_{j=0}^{k} E_j \\
&= A_k\left(f(x^{k+1}) + \sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,k+1}), \tilde{y}_l^k - x_l^{k+1}\right\rangle\right) + A_k\left(\sum_{l=1}^{s}\left\langle \nabla_l f(x^{k+1}) - \nabla_l f(x^{l,k+1}), \tilde{y}_l^k - x_l^{k+1}\right\rangle\right) + \\
&\alpha_{k+1}\left(f(x^{k+1}) + \sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,k+1}), z_l - x_l^{k+1}\right\rangle\right) + \sum_{j=0}^{k} E_j \\
&= A_{k+1} f(x^{k+1}) + \alpha_{k+1}\left(\sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,k+1}), z_l - \tilde{z}_l^k\right\rangle\right) + \\
&\quad \underbrace{A_k\left(\sum_{l=1}^{s}\left\langle \nabla_l f(x^{k+1}) - \nabla_l f(x^{l,k+1}), \tau_k(\tilde{y}_l^k - \tilde{z}_l^k)\right\rangle\right)}_{=\,E_{k+1}} + \sum_{j=0}^{k} E_j,
\end{aligned}
$$

where the last equality follows with (13) and the fact that $\tau_k = \alpha_{k+1}/A_{k+1}$.
From condition (12) it follows that ${A_{k+1}}^{-1} \geq {\tau_k}^2$ and we obtain:

$$
\begin{aligned}
\widetilde{\Psi}^{k+1} &\geq A_{k+1} f(x^{k+1}) + \min_{z\in Q}\Bigg\{\eta_{l^*} L\alpha_{k+1}\Delta_{k+1}\left\|z - \tilde{z}^k\right\|_1 + \frac{L}{2}\left\|z - \tilde{z}^k\right\|^2 + \\
&\quad \alpha_{k+1}\sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,k+1}), z_l - \tilde{z}_l^k\right\rangle\Bigg\} + \sum_{j=0}^{k+1} E_j \\
&= A_{k+1} f(x^{k+1}) + A_{k+1}\min_{z\in Q}\Bigg\{\eta_{l^*} L\Delta_{k+1}\tau_k\left\|z - \tilde{z}^k\right\|_1 + \frac{L}{2A_{k+1}}\left\|z - \tilde{z}^k\right\|^2 + \\
&\quad \tau_k\sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,k+1}), z_l - \tilde{z}_l^k\right\rangle\Bigg\} + \sum_{j=0}^{k+1} E_j \\
&\geq A_{k+1} f(x^{k+1}) + A_{k+1}\min_{z\in Q}\Bigg\{\eta_{l^*} L\Delta_{k+1}\tau_k\left\|z - \tilde{z}^k\right\|_1 + \frac{L}{2}{\tau_k}^2\left\|z - \tilde{z}^k\right\|^2 + \\
&\quad \tau_k\sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,k+1}), z_l - \tilde{z}_l^k\right\rangle\Bigg\} + \sum_{j=0}^{k+1} E_j
\end{aligned}
\tag{15}
$$

For arbitrary $z \in Q$ define

$$y = \tau_k z + (1-\tau_k)\tilde{y}^k.$$

As $\tau_k \in [0,1]$, we have $y \in Q$ and with the definition of $x^{k+1}$ in (13) we can write:

$$y - x^{k+1} = \tau_k(z - \tilde{z}^k).$$

It follows that

$$\begin{aligned}
&\min_{z\in Q}\left[\eta_{l^*} L\Delta_{k+1}\tau_k \left\|z-\tilde{z}^k\right\|_1 + \frac{L}{2}{\tau_k}^2 \left\|z-\tilde{z}^k\right\|^2 + \tau_k\left(\sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,k+1}), z_l - \tilde{z}_l^k\right\rangle\right)\right\} \\
&= \min_{y\in\tau_k Q+(1-\tau_k)\tilde{y}^k}\left\{\eta_{l^*} L\Delta_{k+1}\left\|y-x^{k+1}\right\|_1 + \frac{L}{2}\left\|y-x^{k+1}\right\|^2 + \sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,k+1}), y_l - x_l^{k+1}\right\rangle\right\} \\
&\geq \min_{y\in Q}\left\{\eta_{l^*} L\Delta_{k+1}\left\|y-x^{k+1}\right\|_1 + \frac{L}{2}\left\|y-x^{k+1}\right\|^2 + \sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,k+1}), y_l - x_l^{k+1}\right\rangle\right\} \\
&\geq \min_{y\in Q}\left\{L\Delta_{k+1}\sum_{l=1}^{s}\eta_l\left\|y_l-x_l^{k+1}\right\|_1 + \frac{L}{2}\left\|y-x^{k+1}\right\|^2 + \sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,k+1}), y_l - x_l^{k+1}\right\rangle\right\} \\
&\geq f(\tilde{y}^{k+1}) - f(x^{k+1}).
\end{aligned} \tag{16}$$

Substituting (16) in (15) yields $\widetilde{\Psi}^{k+1} \geq A_{k+1} f(\tilde{y}^{k+1}) + \sum_{j=0}^{k+1} E_j$. □

Finally, we use the result of Lemma 2.2.7 to prove convergence of the DNA-EC:

**Theorem 2.2.8** *Let the sequence $\{x^k\}_{k\geq 0}$ and $\{\tilde{y}^k\}_{k\geq 0}$ be generated by the DNA-EC with $\alpha_k$ as in Lemma 2.1.3 and $\Delta_k = \beta\delta^k$ where $\delta \in (0,1)$ and $\beta \in \mathbb{R}_+$. Then for $k \geq 0$ the inequality*

$$f(\tilde{y}^k) - f(x^*) < \frac{\sigma 6\eta_{l^*}\beta L C g'(\delta) + 4Ld(x^*)}{\sigma(k+1)(k+2)} \tag{17}$$

*holds, where $x^*$ is an optimal solution to problem (1), $C$ is defined as in (8), and $g(\delta) = \sum_{j=0}^{\infty} \Delta_j = 1/(1-\delta)$ for $\delta \in (0,1)$.*

*Proof* To prove the theorem we have to derive an upper bound for the left-hand side $\widetilde{\Psi}^k$ in inequality (14) and a lower bound for the error $\sum_{j=0}^{k} E_j$ occurring in the right-hand side. We start with

$$\begin{aligned}
\widetilde{\Psi}^k &= \min_{z\in Q}\left\{\rho_k + \frac{L}{\sigma}d(z) + \sum_{j=0}^{k}\alpha_j\left(f(x^j) + \sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,j}), z_l - x_l^j\right\rangle\right)\right\}\\
&= \min_{z\in Q}\left\{\rho_k + \frac{L}{\sigma}d(z) + \sum_{j=0}^{k}\alpha_j\left(f(x^j) + \left\langle \nabla f(x^j),\ z - x^j\right\rangle\right) + \right.\\
&\quad \left.\sum_{j=0}^{k}\alpha_j\left(\sum_{l=1}^{s}\left\langle \nabla_l f(x^{l,j}) - \nabla_l f(x^j), z_l - x_l^j\right\rangle\right)\right\}\\
&\le \min_{z\in Q}\left\{\rho_k + \frac{L}{\sigma}d(z) + \sum_{j=0}^{k}\alpha_j\left(f(x^j) + \left\langle \nabla f(x^j),\ z - x^j\right\rangle\right) + \right.\\
&\quad \left.\sum_{j=0}^{k}\alpha_j\left(\sum_{l=1}^{s} L\left\|(x_{j_1}^{l,j},\ldots,x_{j_{\eta_l}}^{l,j}) - (x_{j_1}^{j},\ldots,x_{j_{\eta_l}}^{j})\right\|\left\|z_l - x_l^j\right\|\right)\right\} \qquad (18)\\
&\le \rho_k + \frac{L}{\sigma}d(x^*) + A_k f(x^*) + L\eta_{l^*}\sum_{j=0}^{k}\alpha_j\Delta_j\left\|x^* - x^j\right\|_1 \qquad (19)\\
&\le 2\eta_{l^*}LC\sum_{j=0}^{k}\alpha_j\Delta_j + \frac{L}{\sigma}d(x^*) + A_k f(x^*) = \eta_{l^*}\beta LC\sum_{j=1}^{k}(j+1)\delta^j + \frac{L}{\sigma}d(x^*) + A_k f(x^*)\\
&< \eta_{l^*}\beta LC g'(\delta) + \frac{L}{\sigma}d(x^*) + A_k f(x^*),
\end{aligned}$$

where we used the Lipschitz continuity assumption of the gradient of $f$ to obtain (18) and the fact that $f$ is convex (as it was done in Nesterov's proof of Theorem 2 in [12]) to obtain (19). Similarly, we derive a lower bound for the accumulated error $\sum_{j=0}^{k} E_j$:

$$\begin{aligned}
\sum_{j=0}^{k} E_j &\ge -\left|\sum_{j=1}^{k} A_{j-1}\tau_{j-1}\left(\sum_{l=1}^{s}\left\langle \nabla_l f(x^j) - \nabla_l f(x^{l,j}), \tilde{y}_l^{j-1} - \tilde{z}_l^{j-1}\right\rangle\right)\right|\\
&\ge -\sum_{j=1}^{k} A_{j-1}\tau_{j-1}\left(\sum_{l=1}^{s} L\left\|(x_{j_1}^{j},\ldots,x_{j_{\eta_l}}^{j}) - (x_{j_1}^{l,j},\ldots,x_{j_{\eta_l}}^{l,j})\right\|\left\|\tilde{y}_l^{j-1} - \tilde{z}_l^{j-1}\right\|\right)\\
&\ge -\eta_{l^*}L\sum_{j=1}^{k} A_{j-1}\tau_{j-1}\Delta_j\left\|\tilde{y}_l^{j-1} - \tilde{z}_l^{j-1}\right\|_1\\
&\ge -\eta_{l^*}LC\sum_{j=1}^{k} A_{j-1}\tau_{j-1}\Delta_j = -\eta_{l^*}\beta LC\sum_{j=1}^{k}\frac{j(j+1)}{4}\frac{2}{j+2}\delta^j\\
&> -\frac{\eta_{l^*}\beta LC}{2}g'(\delta).
\end{aligned}$$

Substituting these bounds in (14) results in:

$$f(\tilde{y}^k) - f(x^*) < 4\left(\frac{3}{2}\eta_{l^*}\beta LC g'(\delta) + \frac{L}{\sigma}d(x^*)\right)/\left((k+1)(k+2)\right).$$

□

## 3 Application of the DNA-EC in the Proximal Center Algorithm

### 3.1 Proximal Center Algorithm

In this section we briefly describe the PCA [10] that is applicable for partially separable convex problems of the following form:

$$\min_{x_i \in X_i (i=1,\ldots,n)} \left\{ \sum_{i=1}^{n} \Phi_i(x_i) \ : \ \sum_{i=1}^{n} A_i x_i = b_A, \ \sum_{i=1}^{n} B_i x_i \leq b_B \right\}, \tag{20}$$

where $\Phi_i \colon \mathbb{R}^{m_i} \to \mathbb{R}$ is a continuous and convex but not necessarily smooth function on a given compact and convex set $X_i$ for $i = 1,\ldots,n$. $A_i$ denotes a $m_A \times m_i$ matrix, $B_i$ denotes a $m_B \times m_i$ matrix ($i = 1,\ldots,n$), $b_A \in \mathbb{R}^{m_A}$, and $b_B \in \mathbb{R}^{m_B}$. For the ease of notation we denote in the following by $(\mu,\lambda)$ the vector $(\lambda^T,\mu^T)^T$. As the problem structure suggests, the authors use a dual decomposition approach resulting in a block-separably constrained dual problem. To obtain a continuously differentiable dual objective function that can be evaluated in parallel, scaled prox-functions $d_{X_i}$ with convexity parameters $\sigma_{X_i} > 0$ ($i = 1,\ldots,n$) are added to the Lagrangian of (20) resulting in the following augmented Lagrangian

$$\mathcal{L}_c(x_1,\ldots,x_n,\mu,\lambda) = \sum_{i=1}^{n} \Phi_i(x_i) + \left\langle \sum_{i=1}^{n} A_i x_i - b_A, \mu \right\rangle + \left\langle \sum_{i=1}^{n} B_i x_i - b_B, \lambda \right\rangle + c \sum_{i=1}^{n} d_{X_i}(x_i),$$

which is separable in the primal variables $x_i$ ($i = 1,\ldots,n$). It follows, that the corresponding augmented dual objective function

$$f_c(\mu,\lambda) = \min_{x_i \in X_i (i=1,\ldots,n)} \left\{ \sum_{i=1}^{n} \Phi_i(x_i) + \left\langle \sum_{i=1}^{n} A_i x_i - b_A, \mu \right\rangle + \left\langle \sum_{i=1}^{n} B_i x_i - b_B, \lambda \right\rangle + c \sum_{i=1}^{n} d_{X_i}(x_i) \right\} \tag{21}$$

can be evaluated in parallel. As mentioned in [10], different smoothing parameters $c_{X_i}$ ($i = 1,\ldots,n$) instead of single parameter $c$ can be considered as well and we show in subsection 3.5 how to optimally choose them for certain prox-functions.

The augmented dual objective $f_c$ is concave and due to the uniqueness of the minimizers $x_i(\mu,\lambda)$, that solve the right-hand side of equation (21), continuously differentiable [10]. Moreover, it can be shown (cf. Theorem 3.4 in [10]) that the gradient

$$\nabla f_c(\mu,\lambda) = \begin{pmatrix} \sum_{i=1}^{n} A_i x_i(\mu,\lambda) - b_A \\ \sum_{i=1}^{n} B_i x_i(\mu,\lambda) - b_B \end{pmatrix}$$

is Lipschitz continuous with the Lipschitz constant

$$L_c = \sum_{i=1}^{n} \frac{\|A_i\|^2 + \|B_i\|^2}{c\sigma_{X_i}}. \tag{22}$$

Let $M^* \times \Lambda^*$ be the set of optimal dual multipliers, assumed to be nonempty in the following. Applying Nesterov's algorithm 2.1.1 to the problem

$$\operatorname*{arg\,max}_{(\mu,\lambda)\in Q_A \times Q_B} f_c(\mu,\lambda) = \operatorname*{arg\,min}_{(\mu,\lambda)\in Q_A \times Q_B} -f_c(\mu,\lambda), \tag{23}$$

where $Q_A \times Q_B \subseteq \mathbb{R}^{m_A} \times \mathbb{R}_+^{m_B}$ is a given closed and convex set that contains at least one optimal dual multiplier $(\mu,\lambda)^* \in M^* \times \Lambda^*$, results in the following PCA [10]:

**Algorithm 3.1.1 (PCA)** *For $k \geq 0$ do*

1. *Given $(u,h)^k \in Q_A \times Q_B$, for $i = 1,\ldots,n$ compute*

$$x_i^{k+1} = \arg\min_{x_i \in X_i} \left\{ \Phi_i(x_i) + \left\langle u^k, A_i x_i \right\rangle + \left\langle h^k, B_i x_i \right\rangle + c d_{X_i}(x_i) \right\}.$$

2. *Compute* $\nabla f_c((u,h)^k) = \begin{pmatrix} \sum_{i=1}^n A_i x_i^{k+1} - b_A \\ \sum_{i=1}^n B_i x_i^{k+1} - b_B \end{pmatrix}$.
3. *Find* $(\mu,\lambda)^k = \arg\max_{(\mu,\lambda) \in Q_A \times Q_B} \left\{ \left\langle \nabla f_c((u,h)^k), (\mu,\lambda) - (u,h)^k \right\rangle - \frac{L_c}{2} \left\| (\mu,\lambda) - (u,h)^k \right\|^2 \right\}$.
4. *Find* $(v,t)^k = \arg\max_{(v,t) \in Q_A \times Q_B} \left\{ -\frac{L_c}{\sigma} d(v,t) + \sum_{j=0}^{k} \frac{j+1}{2} \left\langle \nabla f_c((u,h)^j), (v,t) - (u,h)^j \right\rangle \right\}$.
5. *Set* $(u,h)^{k+1} = \frac{2}{k+3}(v,t)^k + \frac{k+1}{k+3}(\mu,\lambda)^k$.

Note that we omitted again some constant terms in the original version of the argmax-problems as done in subsection 2.1 to reveal the parallelizable character of the PCA. 3.1.1. Finally, $d(v,t)$ in step 4 denotes the prox-function for the set $Q_A \times Q_B$ with convexity parameter $\sigma$ and center $(u,h)^0 = \arg\min_{(u,h) \in Q_A \times Q_B} d(u,h)$. To close this section we state a convergence result of the PCA that can be proved using the following Lemma which gives a lower and an upper bound on the primal gap:

**Lemma 3.1.2** *For every $(\mu,\lambda)^* \in M^* \times \Lambda^*$, $(\mu,\lambda) \in Q_A \times Q_B$, and $\hat{x}_i \in X_i$ $(i = 1,\ldots,n)$ the following inequalities hold:*

$$-\left\| (\mu,\lambda)^* \right\| \left\| \begin{matrix} \sum_{i=1}^n A_i \hat{x}_i - b_A \\ \left[ \sum_{i=1}^n B_i \hat{x}_i - b_B \right]^+ \end{matrix} \right\| \leq \sum_{i=1}^n \Phi_i(\hat{x}_i) - f^* \leq \sum_{i=1}^n \Phi_i(\hat{x}_i) - f_0(\mu,\lambda), \tag{24}$$

*where $f^* = f_0((\mu,\lambda)^*)$ and $[\cdot]^+$ denotes the projection onto $\mathbb{R}_+^{m_B}$.*

*Proof* The proof is similar to the proof of Lemma 3.3 in [10]. We have

$$\begin{aligned}
f^* &= \min_{x_i \in X_i (i=1,\ldots,n)} \left\{ \sum_{i=1}^n \Phi_i(x_i) + \left\langle \sum_{i=1}^n A_i x_i - b_A, \mu^* \right\rangle + \left\langle \sum_{i=1}^n B_i x_i - b_B, \lambda^* \right\rangle \right\} \\
&\leq \sum_{i=1}^n \Phi_i(\hat{x}_i) + \left\langle \sum_{i=1}^n A_i \hat{x}_i - b_A, \mu^* \right\rangle + \left\langle \sum_{i=1}^n B_i \hat{x}_i - b_B, \lambda^* \right\rangle \\
&\leq \sum_{i=1}^n \Phi_i(\hat{x}_i) + \left\langle \begin{pmatrix} \left| \sum_{i=1}^n A_i \hat{x}_i - b_A \right| \\ \sum_{i=1}^n B_i \hat{x}_i - b_B \end{pmatrix}, \begin{pmatrix} |\mu^*| \\ \lambda \end{pmatrix} \right\rangle \\
&\leq \sum_{i=1}^n \Phi_i(\hat{x}_i) + \left\langle \begin{bmatrix} \left| \sum_{i=1}^n A_i \hat{x}_i - b_A \right| \\ \sum_{i=1}^n B_i \hat{x}_i - b_B \end{bmatrix}^+, \begin{pmatrix} |\mu^*| \\ \lambda^* \end{pmatrix} \right\rangle \\
&\leq \sum_{i=1}^n \Phi_i(\hat{x}_i) + \left\| \begin{matrix} \sum_{i=1}^n A_i \hat{x}_i - b_A \\ \left[ \sum_{i=1}^n B_i \hat{x}_i - b_B \right]^+ \end{matrix} \right\| \left\| (\mu,\lambda)^* \right\|,
\end{aligned}$$

where the last inequality follows by the Cauchy-Schwarz inequality. □

Due to the compactness of the sets $X_i$, positive and finite constants $D_{X_i}$ exist which satisfy:

$$D_{X_i} \geq \max_{x_i \in X_i} d_{X_i}(x_i) \quad \text{for} \quad i = 1,\ldots,n. \tag{25}$$

**Theorem 3.1.3** *Assume that* $Q_A \times Q_B = \mathbb{R}^{m_A} \times \mathbb{R}_+^{m_B}$ *and* $d(\mu,\lambda) = (\sigma/2)\|(\mu,\lambda)\|^2$. *Taking* $c = \epsilon / \sum_{i=1}^n D_{X_i}$ *in (21) and* $k+1 = \left\lceil 2\sqrt{L_c/\epsilon} \right\rceil$, *where* $L_c = \left(\sum_{i=1}^n D_{X_i}/\epsilon\right)\left(\sum_{i=1}^n \left(\|A_i\|^2 + \|B_i\|^2\right)/\sigma_{X_i}\right)$, *then after k iterations of Algorithm 3.1.1 an approximate solution to the problem (20) is given by*

$$\hat{x}_i = \sum_{j=0}^{k} \frac{2(j+1)}{(k+1)(k+2)} x_i^{j+1} \quad \textit{for} \quad i = 1,\ldots,n \tag{26}$$

*and with* $(\hat{\mu},\hat{\lambda}) = (\mu,\lambda)^k$ *the following bounds on the duality gap are satisfied:*

$$-\left\|(\mu,\lambda)^*\right\| \left( \left\|(\mu,\lambda)^*\right\| + \sqrt{\|(\mu,\lambda)^*\|^2 + 2} \right) \epsilon \le \sum_{i=1}^n \Phi_i(\hat{x}_i) - f_0(\hat{\mu},\hat{\lambda}) \le \epsilon, \tag{27}$$

*as well as the following bound on the constraint violation:*

$$\left\| \begin{matrix} \sum_{i=1}^n A_i \hat{x}_i - b_A \\ \left[ \sum_{i=1}^n B_i \hat{x}_i - b_B \right]^+ \end{matrix} \right\| \le \epsilon \left( \left\|(\mu,\lambda)^*\right\| + \sqrt{\|(\mu,\lambda)^*\|^2 + 2} \right).$$

*Proof* Applying Lemma 3.1.2 the proof is almost identical to the proof of Theorem 3.7 in [10]. □

### 3.2 Scaling the separable convex problem

Before we present the DPCA-EC in the following subsection, we propose a scaling technique for problem (20) to balance the bounds (27) on the duality gap in Theorem 3.1.3. In particular, if the norm of the optimal dual multipliers $\|(\mu,\lambda)^*\|$ in the left-hand side of (27) is very large, $\epsilon$ has to be chosen very small to ensure a tight lower bound on the duality gap, which increases the necessary number of iterations $k$ as we have

$$k+1 = \left\lceil 2\sqrt{\frac{L_c}{\epsilon}} \right\rceil = \left\lceil \frac{2}{\epsilon} \sqrt{\sum_{i=1}^n D_{X_i} \sum_{i=1}^n \frac{\|A_i\|^2 + \|B_i\|^2}{\sigma_{X_i}}} \right\rceil.$$

To remedy this drawback we propose the following scaling approach for problem (20) with scaling factor $s > 1$:
Define $b_A(s) = sb_A$, $b_B(s) = sb_B$, $X_i(s) = sX_i$, $x_i(s) = sx_i$ $(i = 1,\ldots,n)$, and solve

$$\min_{x_i(s) \in X_i(s)(i=1,\ldots,n)} \left\{ \sum_{i=1}^n \Phi_i\left(\frac{1}{s} x_i(s)\right) \; : \; \sum_{i=1}^n A_i x_i(s) = b_A(s), \; \sum_{i=1}^n B_i x_i(s) = b_B(s) \right\}. \tag{28}$$

This scaling approach results in optimal dual multipliers of the form:

$$\left\|(\mu(s),\lambda(s))^*\right\| = \frac{1}{s} \left\|(\mu,\lambda)^*\right\|,$$

and the number of iterations $k$ necessary to reach the desired accuracy $\epsilon$ in Theorem 3.1.3 is given by

$$k = \left\lceil \frac{s}{\epsilon} \sqrt{4 \sum_{i=1}^n D_{X_i} \sum_{i=1}^n \frac{\|A_i\|^2 + \|B_i\|^2}{\sigma_{X_i}}} \right\rceil - 1. \tag{29}$$

For the bounds on the duality gap and the constraint violation we obtain:

$$-\frac{1}{s}\left\|(\mu,\lambda)^*\right\|\left(\frac{1}{s}\left\|(\mu,\lambda)^*\right\|+\sqrt{\frac{1}{s^2}\|(\mu,\lambda)^*\|^2+2}\right)\epsilon \leq \sum_{i=1}^{n}\Phi_i(\hat{x}_i)-f^* \leq \epsilon, \tag{30}$$

and

$$s\left\|\begin{matrix}\sum_{i=1}^{n}A_i\hat{x}_i-b_A\\ \left[\sum_{i=1}^{n}B_i\hat{x}_i-b_B\right]^+\end{matrix}\right\| \leq \epsilon\left(\frac{1}{s}\left\|(\mu,\lambda)^*\right\|+\sqrt{\frac{1}{s^2}\|(\mu,\lambda)^*\|^2+2}\right). \tag{31}$$

It is clear, that increasing the scaling factor $s>1$ raises the Lipschitz constant $L_c$ and by implication the number of iterations $k$ in (29) in the same order of magnitude as decreasing $\epsilon$ does, but with respect to the lower bound on the duality gap in (30) and the bound on the constraint violation in (31), it is favourable to increase $s$ instead of decreasing $\epsilon$.

### 3.3 Distributed Proximal Center Algorithm with event-triggered communication

We will show now that the complexity $O(1/\epsilon)$ of the PCA can be maintained applying the DNA-EC developed in subsection 2.2 to update the dual multipliers in parallel which yields the DPCA-EC. To do so we again use Assumption 2.2.2 which requires a suitable structure of the matrices $A_i$ and $B_i$ in (20) to avoid a complete IDG. We omit details here and anticipate that a suitable structure of the constraints is given in network flow problems such as the DC-OPF problem considered in section 4.
To ensure a parallel implementation of step 1 in Algorithm 3.1.1 after introducing event-triggered communication with $(u,h)^{l,k}$ $(l=1,\ldots,s)$ defined according to (5), we further assume that the agents use the same outdated information of common neighbors in the IDG to compute their sub-blocks of $\nabla f_c$ and that $agent_l$ uses the outdated version of his sub-block $l$ of the dual multipliers in the computation of $\nabla_l f_c((u,h)^{l,k})$:

**Assumption 3.3.1**

$$\textit{If}\ i\in(N_{IDG}(l)\cup\{l\})\cap N_{IDG}(j)\ \textit{then}\ (u,h)_i^{l,k}=(u,h)_i^{j,k}\ \textit{for}\ k\geq 0\ \textit{and}\ i,j,l=1,\ldots,s.$$

**Remark 3.3.2**

*(i) Note that this assumption is not restrictive at all as the convergence result for the DNA-EC in Theorem 2.2.8 still holds substituting $\sqrt{\eta_{l^*}}$ by $\sqrt{\eta_{l^*}+1}$ in (17), where $\eta_{l^*}$ is the maximal degree of the IDG.*
*(ii) Assumption 3.3.1 allows the definition of a "global" vector $(\bar{u},\bar{h})^k$ that contains the possibly outdated information available to the agents:*

$$(\bar{u},\bar{h})_l^k=(u,h)_l^{l,k}\quad \textit{for}\quad k\geq 0\quad \textit{and}\quad l=1,\ldots,s. \tag{32}$$

*Note that with Assumption 2.2.2 it follows that $\nabla_l f_c((\bar{u},\bar{h})^k)=\nabla_l f_c((u,h)^{l,k})$.*

To state the DPCA-EC we denote in the following by $agent_{x_i}$ the agent that updates $x_i^k$ in iteration $k$ and has available all the possibly outdated sub-blocks of $(\bar{u},\bar{h})^k$ contained in the vector $(u,h)^{x_i,k}$ necessary to compute $x_i^{k+1}$ for $k\geq 0$ and $i=1,\ldots,n$. Moreover, denoting by $agent_l$ the agent that updates sub-block $l$ $(l=1,\ldots,s)$ of the dual multipliers, the DPCA-EC is:

**Algorithm 3.3.3 (DPCA-EC)** *For $k \geq 0$ do*
*For $i = 1,\dots,n$ given $(u,h)^{x_i,k}$, agent$_{x_i}$*

1. *computes*

$$x_i^{k+1} = \operatorname*{arg\,min}_{x_i \in X_i} \left\{ \Phi_i(x_i) + \left\langle u^{x_i,k}, A_i x_i \right\rangle + \left\langle h^{x_i,k}, B_i x_i \right\rangle + c d_{X_i}(x_i) \right\}.$$

*For $l = 1,\dots,s$ given the vectors $x_i^{k+1}$ necessary for the computation of $\nabla_l f_c((u,h)^{l,k})$, agent$_l$*

2. *computes* $\nabla_l f_c((u,h)^{l,k}) = \begin{pmatrix} \sum_{i=1}^n A_i x_i^{k+1} - b_A \\ \sum_{i=1}^n B_i x_i^{k+1} - b_B \end{pmatrix}_l$,
3. *finds*

$$(\tilde{\mu},\tilde{\lambda})_l^k = \operatorname*{arg\,max}_{(\mu,\lambda)_l \in (Q_A \times Q_B)_l} \left\{ \left\langle \nabla_l f_c((u,h)^{l,k}), (\mu,\lambda)_l - (u,h)_l^k \right\rangle - L_c \Delta_k (\eta_l + 1) \left\| (\mu,\lambda)_l - (u,h)_l^k \right\|_1 - \frac{L_c}{2} \left( (\mu,\lambda)_l - (u,h)_l^k \right)^2 \right\},$$

4. *finds* $(\tilde{v},\tilde{t})_l^k = \operatorname*{arg\,max}_{(v,t)_l \in (Q_A \times Q_B)_l} \left\{ -\frac{L_c}{\sigma} d_l((v,t)_l) + \sum_{j=0}^{k} \frac{j+1}{2} \left\langle \nabla f_c((u,h)^{l,j}), (v,t)_l - (u,t)_l^j \right\rangle \right\}$,
5. *sets* $(u,t)_l^{k+1} = \frac{2}{k+3} (\tilde{v},\tilde{t})_l^k + \frac{k+1}{k+3} (\tilde{\mu},\tilde{\lambda})_l^k$,
6. *and exchanges information:*

    ***if*** $\left\| (u,h)_l^{l,k} - (u,h)_l^{k+1} \right\|_1 > \Delta_{k+1}$ ***then***
    $(u,h)_l^{x_i,k+1} = (u,h)_l^{k+1}$,
    ***else***
    $(u,h)_l^{x_i,k+1} = (u,h)_l^{l,k}$.

**Remark 3.3.4**

*(i) Note that the communication takes place only between the primal and the dual agents, , i.e., the graph representing the communication topology is different from the IDG in subsection 2.2. We show in the next section how intuitive choices of agent$_{x_i} \in \{$agent$_1,\dots,$agent$_s\}$ ($i = 1,\dots,n$) for solving the DC-OPF problem ensures local communication in the classical sense, i.e., the communication network has the same topology as the power network. Moreover, this choice can be applied in general for solving network-flow problems with the DPCA-EC.*
*(ii) Obviously, $x_i^{k+1}$ has to be computed only if $(u,h)^{x_i,k} \neq (u,h)^{x_i,k-1}$, i.e., the event-triggered dual communication (the exchange of dual iterates) in step* 6 *of the DPCA-EC induces a not closer specified event-triggered primal communication in step* 1.

To prepare the convergence proof of the DPCA-EC under the necessary assumption that $Q_A \times Q_B$ is closed and bounded (cf. Theorem 2.2.8), we provide the following lemma using the quantity $P(\Delta)$ defined by

$$P(\Delta) = 2(\eta_{l^*} + 1)\beta L_c C g'(\delta), \tag{33}$$

where $C = \max_{(\mu,\lambda),(\chi,\xi) \in Q_A \times Q_B} \|(\mu,\lambda) - (\chi,\xi)\|_1$ is the diameter of $Q_A \times Q_B$, $\beta \in \mathbb{R}_+$ is the coefficient in the threshold $\Delta_k = \beta\delta^k$ with $\delta \in (0,1)$, and $g(\delta) = \sum_{j=0}^{\infty} \delta^j = 1/(1-\delta)$ ( cf. Theorem 2.2.8).

**Lemma 3.3.5** *The following inequality holds for $(\tilde{\mu},\tilde{\lambda})^k$ in the DPCA-EC and $(\bar{u},\bar{h})^k$ defined according to (32) for $k \geq 0$:*

$$\frac{(k+1)(k+2)}{4} f_c((\tilde{\mu},\tilde{\lambda})^k) \geq \max_{(\mu,\lambda)\in Q_A\times Q_B} \Bigg\{ -\frac{L_c}{\sigma} d(\mu,\lambda) + \sum_{j=0}^{k} \frac{j+1}{2} \Big( f_c((\bar{u},\bar{h})^j) + \Big\langle \nabla f_c((\bar{u},\bar{h})^j), (\mu,\lambda) - (\bar{u},\bar{h})^j \Big\rangle \Big) \Bigg\} - P(\varDelta). \tag{34}$$

*Proof* For the choice $\alpha_k = (k+1)/2$ for $k \geq 0$ in Lemma 2.2.7 we obtain

$$\frac{(k+1)(k+2)}{4} f_c((\tilde{\mu},\tilde{\lambda})^k) \geq \max_{(\mu,\lambda)\in Q_A\times Q_B} \Bigg\{ -\rho^k - \frac{L_c}{\sigma} d(\mu,\lambda) + \sum_{j=0}^{k} \frac{j+1}{2} \Big( f_c((u,h)^j) + \Big\langle \nabla f_c((\bar{u},\bar{h})^j), (\mu,\lambda) - (u,h)^j \Big\rangle \Big) \Bigg\} - \sum_{j=1}^{k} E_j,$$

where $\rho^k$ is defined by

$$\rho^k = (\eta_{l^*}+1)\, L_c C \sum_{j=1}^{k} \frac{j+1}{2} \beta \delta^j,$$

according to Remark 3.3.2. It can be easily shown (cf. with the proof of Theorem 2.2.8) that

$$\rho^k \leq \frac{P(\varDelta)}{4} \quad \text{and} \quad \sum_{j=1}^{k} E_j \leq \frac{P(\varDelta)}{4},$$

which yields

$$\frac{(k+1)(k+2)}{4} f_c((\tilde{\mu},\tilde{\lambda})^k) \geq \max_{(\mu,\lambda)\in Q_A\times Q_B} \Bigg\{ -\frac{L_c}{\sigma} d(\mu,\lambda) + \sum_{j=0}^{k} \frac{j+1}{2} \Big( f_c((u,h)^j) + \Big\langle \nabla f_c((\bar{u},\bar{h})^j), (\mu,\lambda) - (u,h)^j \Big\rangle \Big) \Bigg\} - \frac{P(\varDelta)}{2}.$$

We have

$$\begin{aligned}
&\Big\langle \nabla f_c((\bar{u},\bar{h})^j), (\mu,\lambda) - (u,h)^j \Big\rangle \\
&= \Big\langle \nabla f_c((u,h)^j), (\mu,\lambda) - (u,h)^j \Big\rangle + \sum_{l=1}^{s} \Big\langle \nabla_l f_c((u,h)^{l,j}) - \nabla_l f_c((u,h)^j), (\mu,\lambda)_l - (u,h)^j_l \Big\rangle \\
&\geq \Big\langle \nabla f_c((u,h)^j), (\mu,\lambda) - (u,h)^j \Big\rangle - L_c \sum_{l=1}^{s} \underbrace{\Big\| (u,h)^{l,j} - (u,h)^j \Big\|}_{\leq (\eta_l+1)\varDelta_j} \Big\| (\mu,\lambda)_l - (u,h)^j_l \Big\| \\
&\geq \Big\langle \nabla f_c((u,h)^j), (\mu,\lambda) - (u,h)^j \Big\rangle - (\eta_{l^*}+1)\, L_c \varDelta_j \underbrace{\Big\| (\mu,\lambda) - (u,h)^j \Big\|_1}_{\leq C},
\end{aligned}$$

and accordingly

$$\Big\langle \nabla f_c((u,h)^j), (\mu,\lambda) - (\bar{u},\bar{h})^j \Big\rangle \geq \Big\langle \nabla f_c((\bar{u},\bar{h})^j), (\mu,\lambda) - (\bar{u},\bar{h})^j \Big\rangle - (\eta_{l^*}+1)\, L_c C \varDelta_j.$$

Finally, with the concavity of $f_c$ we obtain for all $(\mu,\lambda) \in Q_A \times Q_B$ that

$$\begin{aligned}
&\sum_{j=0}^{k} \frac{j+1}{2} \Big( f_c((u,h)^j) + \Big\langle \nabla f_c((u,h)^j), (\mu,\lambda) - (u,h)^j \Big\rangle - (\eta_{l^*}+1) L_c C \Delta_j \Big) - \frac{P(\Delta)}{2} \\
\geq &\sum_{j=0}^{k} \frac{j+1}{2} \Big( f_c((u,h)^j) + \Big\langle \nabla f_c((u,h)^j), (\mu,\lambda) - (u,h)^j \Big\rangle \Big) - \frac{3P(\Delta)}{4} \\
= &\sum_{j=0}^{k} \frac{j+1}{2} \Big( f_c((u,h)^j) + \Big\langle \nabla f_c((u,h)^j), (\mu,\lambda) - (u,h)^j + (\bar{u},\bar{h})^j - (\bar{u},\bar{h})^j \Big\rangle \Big) - \frac{3P(\Delta)}{4} \\
\geq &\sum_{j=0}^{k} \frac{j+1}{2} \Big( f_c((\bar{u},\bar{h})^j) + \Big\langle \nabla f_c((u,h)^j), (\mu,\lambda) - (\bar{u},\bar{h})^j \Big\rangle \Big) - \frac{3P(\Delta)}{4} \\
\geq &\sum_{j=0}^{k} \frac{j+1}{2} \Big( f_c((\bar{u},\bar{h})^j) + \Big\langle \nabla f_c((\bar{u},\bar{h})^j), (\mu,\lambda) - (\bar{u},\bar{h})^j \Big\rangle - (\eta_{l^*}+1) L_c C \Delta_j \Big) - \frac{3P(\Delta)}{4} \\
\geq &\sum_{j=0}^{k} \frac{j+1}{2} \Big( f_c((\bar{u},\bar{h})^j) + \Big\langle \nabla f_c((\bar{u},\bar{h})^j), (\mu,\lambda) - (\bar{u},\bar{h})^j \Big\rangle \Big) - P(\Delta).
\end{aligned}$$

□

Applying Lemma 3.3.5 results in an estimate of the duality gap after $k$ iterations of the DPCA-EC.

**Theorem 3.3.6** *After k iterations of the DPCA-EC we obtain an approximate solution*

$$\hat{x}_i = \sum_{j=0}^{k} \frac{2(j+1)}{(k+1)(k+2)} x_i^{j+1} \quad \text{for} \quad i = 1, \dots, n$$

*to problem (20) and* $(\hat{\mu},\hat{\lambda}) = (\tilde{\mu},\tilde{\lambda})^k$ *which satisfies the following upper bound on the duality gap:*

$$\sum_{i=1}^{n} \Phi_i(\hat{x}_i) - f_0(\hat{\mu},\hat{\lambda}) \leq c \sum_{i=1}^{n} D_{X_i} - \max_{(\mu,\lambda)\in Q_A\times Q_B} \Bigg\{ -\frac{4L_c}{(k+1)^2\sigma} d(\mu,\lambda) + \Bigg\langle \sum_{i=1}^{n} A_i\hat{x}_i - b_A, \mu \Bigg\rangle + \Bigg\langle \sum_{i=1}^{n} B_i\hat{x}_i - b_B, \lambda \Bigg\rangle \Bigg\} + \frac{4P(\Delta)}{(k+1)^2}$$

*Proof* Using inequality (34) the proof is almost identical to the proof of Theorem 3.4 in [10]. □

With Theorem 3.3.6 we can now prove the convergence of the DPCA-EC maintaining the efficiency estimate of $O(1/\epsilon)$:

**Theorem 3.3.7** *Assume that there exists* $R > 0$ *such that the set* $Q_A \times Q_B$ *has the form* $Q_A \times Q_B = \Big\{ (\mu,\lambda) \in \mathbb{R}^{m_A} \times \mathbb{R}_+^{m_B} : \|\mu\|_{\max} \leq R,\ \|\lambda\|_{max} \leq R \Big\}$ *and contains a* $(\mu,\lambda)^* \in M^* \times \Lambda^*$ *with* $\|(\mu,\lambda)^*\| < R$. *Denote by D a finite constant with* $D \geq \max_{(\mu,\lambda)\in Q_A\times Q_B} d(\mu,\lambda)$. *Moreover, let* $\Delta_k$, $g(\delta)$, *and C be defined as in Theorem 2.2.8. Taking* $c = \epsilon / \Big( 2 \sum_{i=0}^{n} D_{X_i} \Big)$ *with* $\epsilon > 0$ *in (21) and*

$$k+1 = \left\lceil 2 \sqrt{\frac{L_c E(\Delta)}{\epsilon}} \right\rceil,$$

*where* $L_c = \left(2\sum_{i=0}^n D_{X_i}/\epsilon\right)\left(\sum_{i=0}^n \left(\|A_i\|^2 + \|B_i\|^2\right)/\sigma_{X_i}\right)$*, and*
$E(\Delta) = (2D + \sigma 4(\eta_{l^*}+1)\beta C g'(\delta))/\sigma$*. Then after k iterations of the DPCA-EC an approximate solution to the problem (20) is given by*

$$\hat{x}_i = \sum_{j=0}^{k} \frac{2(j+1)}{(k+1)(k+2)} x_i^{j+1} \quad for \quad i = 1,\dots,n,$$

*and with* $(\hat{\mu},\hat{\lambda}) = (\tilde{\mu},\tilde{\lambda})^k$ *the following bounds on the duality gap are satisfied:*

$$-\frac{\|(\mu,\lambda)^*\|}{R-\|(\mu,\lambda)^*\|}\epsilon \le \sum_{i=1}^n \Phi_i(\hat{x}_i) - f_0(\hat{\mu},\hat{\lambda}) \le \epsilon, \tag{35}$$

*as well as the following bound on the constraint violation:*

$$\left\| \begin{matrix} \sum_{i=1}^n A_i\hat{x}_i - b_A \\ \left[\sum_{i=1}^n B_i\hat{x}_i - b_B\right]^+ \end{matrix} \right\| \le \frac{\epsilon}{R-\|(\mu,\lambda)^*\|}. \tag{36}$$

*Proof* The proof is similar to the proof of Theorem 3.6 in [10]. If we have a look at the result of Theorem 3.3.6

$$\sum_{i=1}^n \Phi_i(\hat{x}_i) - f_0(\hat{\mu},\hat{\lambda}) \le c\sum_{i=1}^n D_{X_i} - \max_{(\mu,\lambda)\in Q_A\times Q_B}\left\{-\frac{4L_c}{(k+1)^2\sigma}d(\mu,\lambda) + \left\langle \sum_{i=1}^n A_i\hat{x}_i - b_A, \mu\right\rangle + \left\langle \sum_{i=1}^n B_i\hat{x}_i - b_B, \lambda\right\rangle\right\} + \frac{4P(\Delta)}{(k+1)^2} \tag{37}$$

the task is to minimize the right-hand side of the inequality with respect to $c$.
For the maximization part we obtain with the definition of $D$ and
$Q_A\times Q_B = \left\{(\mu,\lambda)\in\mathbb{R}^{m_A}\times\mathbb{R}_+^{m_B} : \|\mu\|_{\max}\le R,\ \|\lambda\|_{\max}\le R\right\}$ that

$$\begin{aligned}
&\max_{(\mu,\lambda)\in Q_A\times Q_B}\left\{-\frac{4L_c}{(k+1)^2\sigma}d(\mu,\lambda) + \left\langle \sum_{i=1}^n A_i\hat{x}_i - b_A, \mu\right\rangle + \left\langle \sum_{i=1}^n B_i\hat{x}_i - b_B, \lambda\right\rangle\right\}\\
&\ge -\frac{4L_cD}{(k+1)^2\sigma} + \max_{\mu\in Q_A}\left\langle \sum_{i=1}^n A_i\hat{x}_i - b_A, \mu\right\rangle + \max_{\lambda\in \mathbb{Q}_B}\left\langle \sum_{i=1}^n B_i\hat{x}_i - b_B, \lambda\right\rangle\\
&= -\frac{4L_cD}{(k+1)^2\sigma} + R\left\|\sum_{i=1}^n A_i\hat{x}_i - b_A\right\|_1 + R\left\|\left[\sum_{i=1}^n B_i\hat{x}_i - b_B\right]^+\right\|_1\\
&\ge -\frac{4L_cD}{(k+1)^2\sigma} + R\left\| \begin{matrix} \sum_{i=1}^n A_i\hat{x}_i - b_A \\ \left[\sum_{i=1}^n B_i\hat{x}_i - b_B\right]^+ \end{matrix} \right\|,
\end{aligned}$$

and for inequality (37) we obtain

$$\sum_{i=1}^n \Phi_i(\hat{x}_i) - f_0(\hat{\mu},\hat{\lambda}) \le c\sum_{i=1}^n D_{X_i} + \frac{4L_cD}{(k+1)^2\sigma} - R\left\| \begin{matrix} \sum_{i=1}^n A_i\hat{x}_i - b_A \\ \left[\sum_{i=1}^n B_i\hat{x}_i - b_B\right]^+ \end{matrix} \right\| + \frac{4P(\Delta)}{(k+1)^2}. \tag{38}$$

$$\le c\sum_{i=1}^n D_{X_i} + \frac{4L_cD}{(k+1)^2\sigma} + \frac{4P(\Delta)}{(k+1)^2}. \tag{39}$$

With

$$L_c = \sum_{i=1}^{n} \frac{||A_i||^2 + ||B_i||^2}{c\sigma_{X_i}} \quad \text{and} \quad P(\Delta) = 2\,(\eta_{l^*}+1)\beta L_c C g'(\delta)$$

we can express the right-hand side of (39) as a function $h(c)$ with

$$\begin{aligned} h(c) &= c\sum_{i=1}^{n} D_{X_i} + L_c\left(\frac{4D}{(k+1)^2\sigma} + \frac{8\,(\eta_{l^*}+1)\beta C g'(\delta)}{(k+1)^2}\right) \\ &= c\sum_{i=1}^{n} D_{X_i} + \frac{1}{c}\left(\sum_{i=1}^{n} \frac{||A_i||^2 + ||B_i||^2}{\sigma_{X_i}}\right)\frac{4D+\sigma 8\,(\eta_{l^*}+1)\beta C g'(\delta)}{(k+1)^2\sigma} \end{aligned}$$

To get the minimum of $h$ we have to solve

$$\begin{aligned} h'(c) &= \sum_{i=1}^{n} D_{X_i} - \frac{1}{c^2}\left(\sum_{i=1}^{n} \frac{||A_i||^2 + ||B_i||^2}{\sigma_{X_i}}\right)\frac{4D+\sigma 8\,(\eta_{l^*}+1)\beta C g'(\delta)}{(k+1)^2\sigma} = 0 \\ \iff c^*_{1,2} &= \pm\sqrt{\left(\sum_{i=1}^{n} \frac{||A_i||^2 + ||B_i||^2}{\sigma_{X_i}}\right)\frac{4D+\sigma 8\,(\eta_{l^*}+1)\beta C g'(\delta)}{(k+1)^2\sigma\sum_{i=1}^{n} D_{X_i}}}. \end{aligned}$$

and as $c$ in (21) has to be positive, we choose

$$c^* = \frac{1}{k+1}\sqrt{\left(\sum_{i=1}^{n} \frac{||A_i||^2 + ||B_i||^2}{\sigma_{X_i}}\right)\frac{4D+\sigma 8\,(\eta_{l^*}+1)\beta C g'(\delta)}{\sigma\sum_{i=1}^{n} D_{X_i}}}. \tag{40}$$

Finally, we get

$$h(c^*) = \frac{2}{k+1}\sqrt{\left(\sum_{i=1}^{n} \frac{||A_i||^2 + ||B_i||^2}{\sigma_{X_i}}\right)\frac{(4D+\sigma 8\,(\eta_{l^*}+1)\beta C g'(\delta))\sum_{i=1}^{n} D_{X_i}}{\sigma}},$$

and with

$$k+1 = \frac{2}{\epsilon}\sqrt{\left(\sum_{i=1}^{n} \frac{||A_i||^2 + ||B_i||^2}{\sigma_{X_i}}\right)\frac{(4D+\sigma 8\,(\eta_{l^*}+1)\beta C g'(\delta))\left(\sum_{i=1}^{n} D_{X_i}\right)}{\sigma}},$$

we obtain the right-hand side of inequality (35) and the value for $c = c^*$.
With inequality (24) and inequality (38) we get

$$\left(R - \left\|(\mu,\lambda)^*\right\|\right)\left\|\begin{matrix}\sum_{i=1}^{n} A_i\hat{x}_i - b_A \\ \left[\sum_{i=1}^{n} B_i\hat{x}_i - b_B\right]^+\end{matrix}\right\| \le c\sum_{i=1}^{n} D_{X_i} + \frac{4L_c D}{(k+1)^2\sigma} + \frac{4P(\Delta)}{(k+1)^2},$$

and the bound on the constraint violation (36) follows immediately by replacing $c$ with $c^*$. Again applying inequality (24) yields the lower bound on the duality gap. □

### 3.4 Optimal convexity parameters

In this subsection we show how to optimally determine the convexity parameter $\sigma_{X_i}$ $(i = 1,\dots,n)$ for the following choice of prox-functions used for smoothing the dual function of (20):

$$d_{X_i}(x_i) = \frac{\sigma_{X_i}}{2}\|x_i\|^2 \quad \text{with} \quad x_i \in X_i \quad \text{for} \quad i = 1,\dots,n. \tag{41}$$

This is important as arbitrarily chosen convexity parameters can lead to a significantly larger Lipschitz constant $L_c$ which increases the amount of iterations $k$ needed in the PCA and the DPCA-EC necessary to reach the desired accuracy $\epsilon$.
We set $v = (\|A_1\|^2 + \|B_1\|^2, \dots, \|A_n\|^2 + \|B_n\|^2)^T$, $d_X = (\max_{x_1\in X_1}\|x_1\|^2, \dots, \max_{x_n\in X_n}\|x_n\|^2)^T$, and $\sigma_X = (\sigma_{X_1}, \dots, \sigma_{X_n})^T$. Then the smoothing parameter $c$ and the Lipschitz constant $L_c$ in Theorem 3.3.7 can be written in the following way:

$$c = \frac{\epsilon}{d_X^T\sigma_X} \quad \text{and} \quad L_c(\sigma_X) = \frac{d_X^T\sigma_X}{\epsilon}\sum_{i=1}^n \frac{v_i}{\sigma_{X_i}}.$$

Note that the scaling $\sigma_X \to \zeta\sigma_X$ with $\zeta > 0$ does not change $L_c(\sigma_X)$. We thus can constrain the minimization of $L_c(\sigma_X)$ by the normalization condition $d_X^T\sigma_X = 1$ and consider the following optimization problem:

$$\underset{\sigma_{X_i}>0\ (i=1,\dots,n)}{\arg\min}\left\{L_c(\sigma_X)\ :\ d_X^T\sigma_X = 1\right\} = \underset{\sigma_{X_i}>0\ (i=1,\dots,n)}{\arg\min}\left\{\sum_{i=1}^n \frac{v_i}{\sigma_{X_i}}\ :\ d_X^T\sigma_X = 1\right\}.$$

The optimality condition is

$$\left.\begin{aligned} -\tfrac{v_i}{\sigma_{X_i}^2} + \mu d_{X_i} &= 0 \\ \sigma_{X_i} &> 0 \end{aligned}\right\} \Longleftrightarrow \sigma_{X_i} = \sqrt{\frac{v_i}{\mu d_{X_i}}} \quad \text{for} \quad i = 1,\dots,n.$$

Using the constraint

$$1 = d_X^T\sigma_X = \frac{1}{\sqrt{\mu}}\sum_{i=1}^n \sqrt{v_i d_{X_i}},$$

the dual multiplier $\mu$ is given by

$$\mu = \left(\sum_{i=1}^n \sqrt{v_i d_{X_i}}\right)^2,$$

and we obtain the optimal convexity parameter

$$\sigma_{X_i} = \frac{1}{\sum_{j=1}^n \sqrt{v_j d_{X_j}}}\sqrt{\frac{v_i}{d_{X_i}}} \quad \text{for} \quad i = 1,\dots,n.$$

Note that the same result is obtained if $c = 2\epsilon/d_X^T\sigma_X$ as in Theorem 3.1.3. Moreover, the optimal convexity parameters can be computed in closed form and in parallel, which is advantageous in a distributed setting.

### 3.5 Optimal smoothing parameters

To further reduce the number of iterations of the PCA, we introduce multiple positive smoothing parameters $c_{X_i}$ $(i = 1, \dots, n)$ instead of a single parameter $c$ in (21) and show how to optimally choose them. Similar to Theorem 3.1 in [10] it can be shown that the Lipschitz constant $L_c$ with $c = (c_{X_1}, \dots, c_{X_n})^T$ of the gradient of the resultant augmented dual function

$$f_c(\mu,\lambda) = \min_{x_i \in X_i (i=1,\dots,n)} \left\{ \sum_{i=1}^{n} \Phi_i(x_i) + \left\langle \sum_{i=1}^{n} A_i x_i - b_A, \mu \right\rangle + \left\langle \sum_{i=1}^{n} B_i x_i - b_B, \lambda \right\rangle + \sum_{i=1}^{n} c_{X_i} d_{X_i}(x_i) \right\} \quad (42)$$

is

$$L_c(\sigma_X) = \sum_{i=1}^{n} \frac{\|A_i\|^2 + \|B_i\|^2}{c_{X_i}\sigma_{X_i}}. \quad (43)$$

Taking the smoothing parameters $c_{X_i}$ $(i = 1, \dots, n)$ into account, we can rewrite Theorem 3.1.3 as follows:

**Theorem 3.5.1** *Let $Q_A \times Q_B = \mathbb{R}^{m_A} \times \mathbb{R}_+^{m_B}$, $d(\mu,\lambda) = (\sigma/2)\|(\mu,\lambda)\|^2$, and define $D_X = \left(D_{X_1}, \dots, D_{X_n}\right)^T$. Take $c = \left(c_{X_1}, \dots, c_{X_n}\right)^T$ in (42) such that $c^T D_X = \epsilon$ and $k+1 = \lceil 2\sqrt{L_c/\epsilon}\rceil$. Then after $k$ iterations of Algorithm 3.1.1 an approximate solution to the problem (20) is given by*

$$\hat{x}_i = \sum_{j=0}^{k} \frac{2(j+1)}{(k+1)(k+2)} x_i^{j+1} \quad \text{for} \quad i = 1, \dots, n$$

*and with $(\hat{\mu}, \hat{\lambda}) = (\mu,\lambda)^k$ the following bounds on the duality gap are satisfied:*

$$-\left\|(\mu,\lambda)^*\right\|\left(\left\|(\mu,\lambda)^*\right\| + \sqrt{\|(\mu,\lambda)^*\|^2 + 2}\right)\epsilon \le \sum_{i=1}^{n} \Phi_i(\hat{x}_i) - f_0(\hat{\mu},\hat{\lambda}) \le \epsilon,$$

*as well as the following bound on the constraint violation:*

$$\left\| \begin{matrix} \sum_{i=1}^{n} A_i \hat{x}_i - b_A \\ \left[\sum_{i=1}^{n} B_i \hat{x}_i - b_B\right]^+ \end{matrix} \right\| \le \epsilon \left( \left\|(\mu,\lambda)^*\right\| + \sqrt{\|(\mu,\lambda)^*\|^2 + 2} \right).$$

*Proof* Applying Lemma 3.1.2 the proof is almost identical to the proof of Theorem 3.7 in [10]. □

To answer the arising question of how to optimally choose the smoothing parameters $c_{X_i}$ $(i = 1, \dots, n)$ that minimize the Lipschitz constant and satisfy the condition $c^T D_X = \epsilon$, we consider an optimization problem similar to the one in the previous subsection:

$$\operatorname*{arg\,min}_{c_{X_i} > 0 \ (i=1,\dots,n)} \left\{ \sum_{i=1}^{n} \frac{v_i}{c_{X_i}\sigma_{X_i}} \ : \ c^T D_X = \epsilon \right\},$$

and obtain the solution

$$c_{X_i} = \frac{\epsilon}{\sum_{j=1}^{n} \sqrt{\frac{v_j D_{X_j}}{\sigma_{X_j}}}} \sqrt{\frac{v_i}{\sigma_{X_i} D_{X_i}}} \quad \text{for} \quad i = 1, \dots, n. \quad (44)$$

in the same way. Note that for the choice of prox-functions $d_{X_i}(x_i) = \left(\sigma_{X_i}/2\right)\|x_i\|^2$ and $c_{X_i}$ according to (44) the dual augmented function $f_c(\mu,\lambda)$ in (42) as well as the Lipschitz constant $L_c$ of the gradient of $f_c(\mu,\lambda)$ in (43) do not depend on the convexity parameters $\sigma_{X_i} > 0$ as we have $c_{X_i}\sigma_{X_i} = \left(\epsilon/\sum_j \sqrt{v_j d_{X_j}/2}\right)\sqrt{2v_i/d_{X_i}}$ for $i = 1,\ldots,n$, with $d_X = (d_{X_1},\ldots,d_{X_n})^T$ defined as in the previous subsection.

## 4 Application to the DC optimal power flow problem

In this section we discuss the results of the DPCA-EC applied to solve the DC-OPF problem. The objective of the DC-OPF problem is to determine the most economical distribution of power generation to serve given loads in a power system, restricted by operational limits of generation and transmission facilities as well as Kirchhoff's Current Law expressed by the power flow equations.
We show that the problem can be solved in parallel with event-triggered and local communication between the agents, i.e., the communication topology equals the topology of the power system. We model the power system according to [5,22] by a directed graph $G = \{V,E\}$ with $V = \{v_1,\ldots,v_n\}$ representing the set of buses $i = 1,\ldots,n$, where all buses contain a load and for simplicity of notation are directly connected to a generator. The generalization of only $p < n$ buses directly connected to a generator is straightforward. The edge $e_{ij} \in E \subseteq V\times V$ with $|E| = m$ represents the transmission line from bus $i$ to bus $j$. Let $I$ be the $m\times n$ incidence matrix of the graph $G$ and define a diagonal matrix $D \in \mathbb{R}^{m\times m}$ with $d_{ll} = 1/x_l$ where $x_l$ denotes the reactance of the $l$th transmission line. Moreover, we define the variables of the DC-OPF problem and their feasible sets:

- Let $\theta_i \in \Theta_i$ be the phase angle of the voltage at bus $i$, given the compact and convex set $\Theta_i = \left[\theta_i^{min},\theta_i^{max}\right]$ for $i = 1,\ldots,n$, and set $\theta = (\theta_1,\ldots,\theta_n)^T$.
- Let $P_i^g \in P_i$ be the generated power at bus $i$, given the compact and convex set $P_i = \left[P_i^{gmin},P_i^{gmax}\right]$ for $i = 1,\ldots,n$, and set $P_g = (P_1^g,\ldots,P_n^g)^T$.

In the DC model of a power system the active power flow between two neighboring buses $i$ and $j$ is approximated by

$$F_{ij} = \frac{1}{x_{ij}}(\theta_i-\theta_j) = -\frac{1}{x_{ji}}(\theta_j-\theta_i) = -F_{ji},$$

where $x_{ij} = x_{ji}$ is the reactance of the transmission line connecting bus $i$ and bus $j$. Accordingly, the power flow on every transmission line can be expressed with $A\theta$ after defining the weighted incidence matrix $A = DI \in \mathbb{R}^{m\times n}$. Given the loads $P^d = (P_1^d,\ldots,P_n^d)^T$ at each bus and the upper and lower line flow limits $F^{max} = (F_1^{max},\ldots,F_m^{max})^T$ and $F^{min} = (F_1^{min},\ldots,F_m^{min})^T$ on each transmission line, the convex DC-OPF problem that we consider has the form:

$$\min_{P_i^g\in P_i,\theta_i\in\Theta_i} \sum_{i=1}^{n} \Phi_i(P_i^g) \tag{45}$$

subject to

$$B\theta = P^g - P^d \tag{46}$$

$$F^{min} \leq A\theta \leq F^{max} \tag{47}$$

where $\Phi_i(x) = a_{i2}x^2 + a_{i1}x + a_{i0}$ is the quadratic cost of power production at bus $i$ with non-negative coefficients $a_{i2}, a_{i1}, a_{i0}$ for $i = 1,\dots,n$ and matrix $B \in \mathbb{R}^{n\times n}$ is defined as:

$$B = -I^T DI \Longrightarrow B_{ij} = \begin{cases} -\sum_{k\in N(i)} \frac{1}{x_{ik}}, & \text{if } i = j, \\ \frac{1}{x_{ij}}, & \text{if } j \in N(i), \\ 0, & \text{else,} \end{cases}$$

where $N(i) = \left\{ j \mid (v_i, v_j) \in E \vee (v_j, v_i) \in E \right\}$ denotes the set of indices of $v_i$'s neighbors. Clearly, constraint (47) limits the power flow on each transmission line, whereas constraint (46) expresses the power flow equations.

To smoothen the Lagrangian of the DC-OPF problem we choose the optimal smoothing parameters $c = (c_1,\dots,c_{2n})^T$ according to (44) in subsection 3.5 and the prox-functions $d_i(x_i) = (\sigma_i/2)\,x_i^2$ with $\sigma_i > 0$ for $i = 1,\dots,2n$, corresponding to the number of primal variables $P_i^g$ and $\theta_i$. Note that the convexity parameter $\sigma_i > 0$ can be chosen arbitrarily as mentioned in subsection 3.5. We obtain the following augmented dual objective function:

$$\begin{aligned} f_c(\mu,\lambda) = \min_{P_i^g\in P_i,\ \theta_i\in\Theta_i} &\left\{ \sum_{i=1}^{n} \Phi_i(P_i^g) + \sum_{l=1}^{m} \lambda_l \Big( (A\theta)_l - F_l^{max} \Big) + \sum_{l=1}^{m} \lambda_{l+m} \Big( (-A\theta)_l + F_l^{min} \Big) + \right. \\ &\left. \sum_{i=1}^{n} \mu_i \Big( P_i^g - (B\theta)_i - P_i^d \Big) + \sum_{i=1}^{n} \frac{c_i\sigma_i}{2} P_i^{g2} + \sum_{i=1}^{n} \frac{c_{i+n}\sigma_{i+n}}{2} \theta_i^2 \right\} \\ = &\sum_{i=1}^{n} \min_{P_i^g\in P_i} \left\{ \Phi_i(P_i^g) + \mu_i P_i^g + \frac{c_i\sigma_i}{2} P_i^{g2} \right\} + \\ &\sum_{i=1}^{n} \min_{\theta_i\in\Theta_i} \left\{ \left( \sum_{l\in L(i)} (\lambda_l - \lambda_{l+m}) A_{li} - \sum_{j\in N(i)\cup\{i\}} \mu_j B_{ji} \right) \theta_i + \frac{c_{i+n}\sigma_{i+n}}{2} \theta_i^2 \right\} + \\ &\sum_{l=1}^{m} \Big( \lambda_{l+m} F_l^{min} - \lambda_l F_l^{max} \Big) - \sum_{i=1}^{n} \mu_i P_i^d, \end{aligned} \tag{48}$$

where $L(i) = \{l \mid v_i \in e_l \wedge e_l \in E\}$ denotes the set of indices of lines connected to bus $i$. As mentioned above, the DC-OPF problem can be solved in parallel by the DPCA-EC with local communication for the following intuitive distribution of agents:

At each bus $i$ an $agent_i$ is installed that controls the dual variable $\mu_i$ corresponding to the power flow balance constraint at bus $i$, as well as the primal variables $P_i^g$ and $\theta_i$ for $i = 1,\dots,n$. Moreover, at each transmission line two agents, $agent_{l+n}$ and $agent_{l+n+m}$, are installed that control the dual multipliers $\lambda_l$ and $\lambda_{l+m}$ corresponding to the upper and lower line flow limits at transmission line $l$ for $l = 1,\dots,m$. This setting containing $2m + n$ agents ensures that the communication takes place only locally, i.e., between topologically neighbored agents with respect to the power network as can be seen in the following adjusted notation of the DPCA-EC to solve the DC-OPF problem. We we choose $d = (\sigma/2)\,\|(\mu,\lambda)\|^2$ according to the Assumptions 2.2.1 made in subsection 2.2. Note that for this choice of prox-function the convexity parameter $\sigma > 0$ can be chosen arbitrarily (cf. step 4 of Algorithm 3.3.3):

**Algorithm 4.0.2** *For $k \geq 0$ do in parallel:*

1. *For $i = 1,\ldots,n$, given $\bar{u}_j^k$ if $j \in N(i) \cup \{i\}$ and $\bar{h}_l^k, \bar{h}_{l+m}^k$ if bus i is connected to transmission line l, i.e., if $l \in L(i)$, where $\bar{u}^k$ and $\bar{h}^k$ are defined according to (32), agent$_i$ solves*

$$P_i^{g,k+1} = \underset{P_i^g \in P_i}{\arg\min} \left\{ \Phi_i(P_i^g) + \bar{u}_i^k P_i^g + \frac{c_i \sigma_i}{2} {P_i^g}^2 \right\},$$

$$\theta_i^{k+1} = \underset{\theta_i \in \Theta_i}{\arg\min} \left\{ \left( \sum_{l \in L(i)} (\bar{h}_l^k - \bar{h}_{l+m}^k) A_{li} - \sum_{j \in N(i) \cup \{i\}} \bar{u}_j^k B_{ji} \right) \theta_i + \frac{c_{i+n}\sigma_{i+n}}{2} {\theta_i}^2 \right\},$$

   *if not already done in a previous iteration. After that he sends $\theta_i^{k+1}$ to agent$_j$ with $j \in N(i)$, agent$_{l+n}$, and agent$_{l+n+m}$ with $l \in L(i)$.*
2. *Given*

$$L_c = \sum_{i=1}^{n} \left( 2 \sum_{l \in L(i)} A_{li}^2 + \sum_{j \in N(i) \cup \{i\}} B_{ji}^2 \right) \Big/ (c_{i+n}\sigma_{i+n}) + \sum_{i=1}^{n} 1/(c_i \sigma_i),$$

   *For $i = 1,\ldots,n$, agent$_i$ computes*

$$\nabla_i^k := \frac{\partial f_c(\bar{u}^k, \bar{h}^k)}{\partial \mu_i} = P_i^{g,k+1} - \sum_{j \in N(i) \cup \{i\}} B_{ij} \theta_j^{k+1} - P_i^d$$

   *and solves*

$$\tilde{\mu}_i^k = \underset{\mu \in \mathbb{R}}{\arg\max} \left\{ \mu \nabla_i^k - (\eta_i + 1) L_c \Delta_k \left| \mu - \mu_i^k \right| - \frac{L_c}{2} (\mu - \mu_i^k)^2 \right\},$$

$$\tilde{v}_i^k = \underset{v \in \mathbb{R}}{\arg\max} \left\{ -\frac{L_c}{2} v^2 + v \sum_{j=0}^{k} \frac{j+1}{2} \nabla_i^j \right\},$$

   *where $G(l) = \left\{ i, j \mid e_l = (v_i, v_j) \in E \right\}$ is the set of indices of buses connected by transmission line l. For $l = 1,\ldots,m$, agent$_{l+n}$ computes the partial derivative*

$$\nabla_{l+n}^k := \frac{\partial f_c(\bar{u}^k, \bar{h}^k)}{\partial \lambda_l} = \sum_{i \in G(l)} A_{li} \theta_i^{k+1} - F_l^{max}$$

   *and solves*

$$\tilde{\lambda}_l^k = \underset{\lambda \in \mathbb{R}_+}{\arg\max} \left\{ \lambda \nabla_{l+n}^k - (\eta_{l+n} + 1) L_c \Delta_k \left| \lambda - \lambda_l^k \right| - \frac{L_c}{2} (\lambda - \lambda_l^k)^2 \right\},$$

$$\tilde{t}_l^k = \underset{t \in \mathbb{R}_+}{\arg\max} \left\{ -\frac{L_c}{2} t^2 + t \sum_{j=0}^{k} \frac{j+1}{2} \nabla_{l+n}^j \right\}.$$

   *Agent$_{l+n+m}$ computes*

$$\nabla_{l+n+m}^k := \frac{\partial f_c(\bar{u}^k, \bar{h}^k)}{\partial \lambda_{l+m}} = \sum_{i \in G(l)} -A_{li} \theta_i^{k+1} + F_l^{min}$$

*and solves*

$$\tilde{\lambda}^k_{l+m} = \arg\max_{\lambda\in\mathbb{R}_+} \left\{ \lambda \nabla^k_{l+n+m} - (\eta_{l+n+m}+1)\, L_c \Delta_k \left|\lambda - \lambda^k_{l+m}\right| - \frac{L_c}{2}(\lambda - \lambda^k_{l+m})^2 \right\},$$

$$\tilde{t}^k_{l+m} = \arg\max_{t\in\mathbb{R}_+} \left\{ -\frac{L_c}{2} t^2 + t \sum_{j=0}^{k} \frac{j+1}{2} \nabla^k_{l+n+m} \right\}.$$

3. *For $i = 1,\ldots,n$ and $l = 1,\ldots,m$ the agent$_i$, agent$_{l+n}$, and agent$_{l+n+m}$ update $u^{k+1}_i$, $h^{k+1}_l$, and $h^{k+1}_{l+m}$ according to:*

$$u^{k+1}_i = \frac{k+1}{k+3}\tilde{\mu}^k_i + \frac{2}{k+3}\tilde{\nu}^k_i,$$
$$h^{k+1}_l = \frac{k+1}{k+3}\tilde{\lambda}^k_l + \frac{2}{k+3}\tilde{t}^k_l,$$
$$h^{k+1}_{l+m} = \frac{k+1}{k+3}\tilde{\lambda}^k_{l+m} + \frac{2}{k+3}\tilde{t}^k_{l+m}.$$

4. *For $i = 1,\ldots,n$ and $l = 1,\ldots,m$ agent$_i$, agent$_{l+n}$, and agent$_{l+n+m}$ exchange information:*

   ***if*** $\left|\bar{u}^k_i - u^{k+1}_i\right| > \Delta_{k+1}$ ***then***
   $\bar{u}^{k+1}_i = u^{k+1}_i$ ***and*** *sends $u^{k+1}_i$ to agent$_j$ with $j \in N(i)$,*
   ***else***
   *agent$_i$ sends nothing and sets $\bar{u}^{k+1}_i = \bar{u}^k_i$ as well as agent$_j$ with $j \in N(i)$.*
   ***if*** $\left|\bar{h}^k_l - h^{k+1}_l\right| > \Delta_{k+1}$ ***then***
   *agent$_{l+n}$ $\bar{h}^{k+1}_l = h^{k+1}_l$ and sends $h^{k+1}_l$ to agent$_j$ with $l \in L(j)$,*
   ***else***
   *agent$_{l+n}$ sends nothing and sets $\bar{h}^{k+1}_l = \bar{h}^k_l$ as well as agent$_j$ with $l \in L(j)$.*

   *Agent$_{l+m+n}$ proceeds accordingly with $h^{k+1}_{l+m}$ and $\bar{h}^{k+1}_{l+m}$.*

**Remark 4.0.3** *A look at the derivatives $\nabla^k_j$ for $j = 1,\ldots,2m+n$ yields the number of neighbors $\eta_j$ of agent$_j$ in the IDG that describes the dependence of the jth component $\nabla_j f_c(\mu,\lambda)$ from the components controlled by other agents (cf. Assumption 2.2.2):*

$$\eta_j = \begin{cases} 3\sum_{i\in N(j)} |N(i)|, & \text{for} \quad j = 1,\ldots,n, \\ 3\sum_{i\in L(j)} |N(i)| - 2, & \text{for} \quad j = n+1,\ldots,n+m, \\ 3\sum_{i\in L(j-m)} |N(i)| - 2, & \text{for} \quad j = n+m+1,\ldots,n+2m. \end{cases}$$

Clearly, all argmin- and argmax-problems in the above algorithm can be solved analytically, similar to as it was shown in Example 2.2.4, which is very advantageous with respect to the computational complexity of each iteration. Another advantageous feature of the above algorithm is that the power variables $P^g_i$, which can be interpreted as private information [6] that should be kept secret if generators belong to different power suppliers, doesn't have to be exchanged between the agents.
Finally, for the sake of completeness let us note that in the case of $a_{i2} \neq 0$ for $i = 1,\ldots,n$ the cost functions $\Phi_i(x)$ are strongly convex with convexity parameters $2a_{i2}$. In this case we only need to smoothen the Lagrangian of problem (45) with respect to the phase angle variables $\theta_i$ resulting in the following augmented

dual function:

$$f_c(\mu,\lambda) = \sum_{i=1}^{n} \min_{P_i^g \in P_i} \left\{ \Phi_i(P_i^g) + \mu_i P_i^g \right\} + \sum_{i=1}^{n} \min_{\theta_i \in \Theta_i} \left\{ \left( \sum_{l \in L(i)} (\lambda_l - \lambda_{l+m}) A_{li} - \sum_{j \in N(i) \cup \{i\}} \mu_j B_{ji} \right) \theta_i + \frac{c_{i+n}\sigma_{i+n}}{2} \theta_i^2 \right\} + \sum_{l=1}^{m} \left( \lambda_{l+m} F_l^{min} - \lambda_l F_l^{max} \right) - \sum_{i=1}^{n} \mu_i P_i^d. \quad (49)$$

It is straightforward to show that the Lipschitz constant of the gradient $\nabla f_c(\mu,\lambda)$ of (49) is:

$$L_c = \sum_{i=1}^{n} \left( 2 \sum_{l \in L(i)} A_{li}^2 + \sum_{j \in N(i) \cup \{i\}} B_{ji}^2 \right) / (c_{i+n}\sigma_{i+n}) + \sum_{i=1}^{n} 1/(2a_{i2}).$$

We implemented the above algorithm as well as the PCA in Matlab $R2012a$ and compared them with respect to the communication exchange. For the tests we used data from the power systems test case archive of the University of Washington that can be found in [4]. The first problem that we discuss is the IEEE 57 bus test case with 80 transmission lines and where 7 of the 57 buses are each connected to a generator (for illustration we refer to [4]). We chose quadratic cost functions with $a_{i2} \neq 0$ for $i = 1, \ldots, 7$ and scaled the problem with $s = 30$ according to subsection 3.2. We applied the solver IPOPT [19] to determine the optimal scaled dual multipliers $\|\mu^*(s), \lambda^*(s)\| = 0.4998 < 1$ and the optimal value $f^*$ of (45). For comparison we state here the primal gap of the approximate solutions (51) at the starting point $(\mu^0, \lambda^0)$, i.e., at the center of the set $Q = \mathbb{R}^n \times \mathbb{R}_+^{2m}$ according to (3):

$$\left( \mu^0, \lambda^0 \right) = \arg\min_{(\mu,\lambda) \in Q} d_Q(\mu,\lambda), \quad (50)$$

which is $-23.0516$ $K\$/h$. Moreover, the constraint violation of the approximate solutions (51) at the starting point is 4.4571.

For the choice $\epsilon = 0.7$ after $k = 33826$ iterations of the PCA the approximate solutions

$$(\hat{P}^g, \hat{\theta}) = \sum_{j=0}^{k} \frac{2(j+1)}{(k+1)(k+2)} (P^{g,j+1}, \theta^{j+1}), \quad (51)$$

where $\hat{P}_i^g = 0$ if bus $i$ is not connected to a generator, should satisfy the following bounds on the primal gap and the constraint violation according to inequality (24) and Theorem 3.1.3 (combined with inequalities (30) and (31) in subsection 3.2):

$$-0.6996 \leq \sum_{i=1}^{n} \Phi_i(\hat{P}_i^g) - f^* \leq 0.7 \quad (52)$$

and

$$\left\| \begin{matrix} \hat{P}^g - B\hat{\theta} - P^d \\ \begin{bmatrix} A\hat{\theta} - F^{max} \\ -A\hat{\theta} + F_{min} \end{bmatrix}^+ \end{matrix} \right\| \leq 0.0467. \quad (53)$$

This is verified by the first row of the following Table 1 which shows the actual primal gap and constraint violation of the solution obtained without event-triggered communication. The rows 2 - 5 of Table 1 show the primal gap and the constraint violation of solutions obtained with the DPCA-EC for different choices of $\beta$ using the threshold $\Delta_k = \beta\delta^k$ with $\delta = 0.9998$. Here $\delta$ is chosen such that there holds $\delta^k \approx 0.025$ after half of the required number of iterations ($k \approx 33000/2$), which ensures that $\delta^k$ does not get too close to zero too early.

In column 3 of Table 1 we state the amount of the overall communication events, i.e., the exchange of primal and dual iterates between the agents described in Algorithm 4.0.2. In column 5 we separately state the amount of the dual communication events, i.e., the exchange of dual iterates, as the framework for event-triggered communication was introduced into the DNA-EC developed in subsection 2.2 that is used to update dual multipliers. It can be seen that only for $\beta \in \{10^{-2}, 10^{-3}\}$ the bounds on the constraint

| | $\beta \cdot \delta^k$ | primal gap | constraint violation | overall communication | dual communication |
|---|---|---|---|---|---|
| 1 | - | -0.1865 | 0.0127 | 32.5e6 (=100%) | 16.2e6 (=100%) |
| 2 | $10^{-2} \cdot 0.9998^k$ | -0.2776 | 0.2354 | 11.9e6 (≈ 36 %) | 2.0e6 (≈ 12 %) |
| 3 | $10^{-3} \cdot 0.9998^k$ | 0.2410 | 0.0635 | 16.3e6 (≈ 50 %) | 3.1e6 (≈ 19 %) |
| 4 | $10^{-4} \cdot 0.9998^k$ | -0.0946 | 0.0129 | 19.4e6 (≈ 59 %) | 4.0e6 (≈ 24 % ) |
| 5 | $10^{-5} \cdot 0.9998^k$ | -0.1473 | 0.0105 | 21.0e6 (≈ 64 %) | 4.9e6 (≈ 30 % ) |

Table 1: Results for the IEEE 57 bus test case

violation (53) are slightly exceeded whereas for $\beta \in \{10^{-4}, 10^{-5}\}$ the primal gap and the constraint violation stay within the given accuracy estimates (52) and (53) of the PCA with the difference that up to 50% of the overall communication and up to 76% of the dual communication could be saved compared to the solution obtained without event-triggered communication. Moreover, if instantaneous communication is assumed we measured an overall computation time of less than 2 seconds on a Intel Xeon Pentium X5570 Pro processor with 2.6 GHz which is due to the low computational complexity of each iteration.

We obtained similar results for another DC-OPF problem that we built with data from the IEEE 118 bus test case [4] with 186 transmission lines and where 34 of the 118 buses are each connected to a generator. Again, we added quadratic cost functions with $a_{i2} \neq 0$ for $i = 1, \ldots, 34$ and scaled the problem with $s = 20$ to obtain optimal scaled dual multipliers with $\|\mu^*(s), \lambda^*(s)\| = 0.6369 < 1$. For comparison we state here again the primal gap of the approximate solutions (51) at the starting point $(\mu^0, \lambda^0)$ which is $-39.8456\ K\$/h$ . Moreover, the constraint violation of the approximate solutions (51) at the starting point is 5.0922. For $\epsilon = 1$ after $k = 122959$ iterations of the PCA the approximate solutions (51) should satisfy the following bounds:

$$-1.3935 \leq \sum_{i=1}^{n} \Phi_i(\hat{P}_i^g) - f^* \leq 1$$

and

$$\left\| \begin{matrix} \hat{P}^g - B\hat{\theta} - P^d \\ \begin{bmatrix} A\hat{\theta} - F^{max} \\ -A\hat{\theta} + F_{min} \end{bmatrix}^+ \end{matrix} \right\| \leq 0.1094.$$

Using a layout identical to Table 1, Table 2 shows the results for $\delta = 0.9999$ and $\beta$ in $\Delta_k = \beta\delta^k$ . Here, the measured overall computation time of less than 8 seconds.

## 4.1 Conclusions

We presented the DNA-EC and the DPCA-EC for distributed optimization with event-triggered communication. The DNA-EC extends an optimal first order method by Nesterov and is applicable to problems

| | $\beta \cdot \delta^k$ | primal gap | constraint violation | overall communication | dual communication |
|---|---|---|---|---|---|
| 1 | - | -0.3888 | 0.0324 | 27.4e7 (=100 %) | 13.7e7 (=100 %) |
| 2 | $10^{-2} \cdot 0.9999^k$ | 0.3101 | 0.4904 | 12.1e7 (≈ 44 %) | 2.5e7 (≈ 18 %) |
| 3 | $10^{-3} \cdot 0.9999^k$ | -0.0537 | 0.0616 | 14.5e7 (≈ 52 %) | 3.1e7 (≈ 22 %) |
| 4 | $10^{-4} \cdot 0.9999^k$ | -0.3316 | 0.0332 | 16.4e7 (≈ 59 %) | 3.6e7 (≈ 26 %) |
| 5 | $10^{-5} \cdot 0.9999^k$ | -0.3899 | 0.0326 | 17.7e7 (≈ 64 %) | 4.2e7 (≈ 30 %) |

Table 2: Results for the IEEE 118 bus test case

with convex and continuously differentiable objective function constrained by a block-separable set. The DPCA-EC extends the proximal center algorithm by Necoara and Suykens and is applicable to problems with separable and convex but not necessarily continuously differentiable objective function constrained by coupled linear constraints.
We showed convergence for both algorithms and gave accuracy estimates. Moreover, numerical results showed that the application of the DPCA-EC, to solve the DC-OPF problem for the IEEE 57 bus and IEEE 118 bus test cases in parallel and with local and event-triggered communication, can significantly reduce information exchange between the agents without loosing accuracy.

## 4.2 Acknowledgment

The support of this work by the German Research Foundation (DFG) within the Priority Program SPP 1305 “Control Theory of Digitally Networked Dynamical Systems” is gratefully acknowledged.

## References

1. Bakirtzis, A., Biskas, P.: A decentralized solution to the dc-opf of interconnected power systems. IEEE Transactions on Power Systems **18**(3), 1007–1013 (2003)
2. Bertsekas, D., Tsitsiklis, J.: Parallel and Distributed Computation: Numerical Methods. Prentice-Hall (1989)
3. Bhattacharya, S., Kumar, V., Likhachev, M.: Distributed optimization with pairwise constraints and its application to multi-robot path planning. In: Proceedings of the Conference on Robotics: Science and Systems (2010)
4. of Washington - Department of Electrical Engineering, U.: Power systems test case archive. http://www.ee.washington.edu/research/pstca/ (1999)
5. J, S., Tesfatsion, L.: Dc optimal power flow formulation and solution using quadprogj. Economics Department, Iowa State University **6014**, 1–36 (2006)
6. Jacobo, J., Roure, D.D.: A decentralised dc optimal power flow model. In: Proceedings of the International Conference on Deregulation and Restructuring and Power Technologies (2008)
7. Kim, B., Baldick, R.: A comparison of distributed optimal power flow algorithms. IEEE Transactions on Power Systems **15**(2), 599–604 (2000)
8. Lemmon, M.: Event-triggered feedback in control, estimation, and optimization. In: A.B.M. Heemels,, M. Johansson (eds.) Networked Control Systems, vol. 406, pp. 293–358. Springer (2010)
9. Lin, C., Lin, S.: Distributed optimal power flow with discrete control variables of large distributed power systems. IEEE Transactions on Power Systems **23**(3), 1383–1392 (2008)
10. Necoara, I., Suykens, J.: Application of a smoothing technique to decomposition in convex optimization. IEEE Transactions on Automatic Control **53**(11), 2674–2679 (2008)
11. Nedic, A., Ozdaglar, A.: Cooperative distributed multi-agent optimization. In: Y. Eldar, D. Palomar (eds.) Convex Optimization in Signal Processing and Communications. Cambridge University Press (2008)
12. Nesterov, Y.: Smooth minimization of non-smooth functions. Mathematical Programmming **103**(1), 127–152 (2005)
13. Palomar, D., Chiang, M.: A tutorial on decomposition methods for network utility maximization. IEEE Journal on Selected Areas in Communication **24**(8), 1439–1451 (2006)
14. Palomar, D., Chiang, M.: Alternative distributed algorithms for network utility maximization: Framework and applications. IEEE Transactions on Automatic Control **52**(12), 2254–2269 (2007)
15. Parker, L.: Decision making as optimization in multi-robot teams. In: Proceedings of the International Conference on Distributed Computing and Internet Technology (2012)

16. Purchala, K., Meeus, L., Van Dommelen, D., Belmans, R.: Usefulness of dc power flow for active power flow analysis. In: Proceedings of the IEEE Power Engineering Society General Meeting, vol. 1, pp. 454–459 (2005)
17. Samar, S., Boyd, S., Gorinevsky, D.: Distributed estimation via dual decomposition. In: Proceedings of the European Control Conference, pp. 1511–1516 (2007)
18. Speranzon, A., Fischione, C., Johansson, K.: Distributed and collaborative estimation over wireless sensor networks. In: Proceedings of the IEEE Conference on Decision and Control, pp. 1025–1030 (2006)
19. Wächter, A., Biegler, L.: On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Mathematical Programming **106**, 25–57 (2006)
20. Wan, P., Lemmon, M.: Distributed network utility maximization using event-triggered augmented lagrangian methods. In: Proceedings of the American Control Conference, vol. 1, pp. 3298–3303 (2009)
21. Wan, P., Lemmon, M.: Distributed network utility maximization using event-triggered barrier methods. In: Proceedings of the American Control Conference (2009)
22. Wan, P., Lemmon, M.: Optimal power flow in microgrids using event-triggered optimization. In: American Control Conference, pp. 2521–2526 (2010)
23. Wei, E., Ozdaglar, A., Jadbabaie, A.: A distributed newton method for network utility maximization. In: Proceedings of the IEEE Conference on Decision and Control, pp. 1816–1821 (2010)
24. Yang, B., Johansson, M.: Distributed optimization and games: A tutorial overview. In: A.B.M. Heemels,, M. Johansson (eds.) Networked Control Systems, vol. 406, pp. 109–148. Springer (2010)
25. Zhong, M., Cassandras, C.: Asynchronous distributed optimization with event-driven communication. IEEE Transactions on Automatic Control **55**(12), 2735–2750 (2010)