

Penalized Push-Sum Algorithm for Constrained Distributed Optimization with Application to Energy Management in Smart Grid

Tatiana Tatarenko, Jan Zimmermann, Volker Willert, Jürgen Adamy

Abstract— We study distributed convex constrained optimization on a time-varying multi-agent network. Each agent has access to its own local cost function, its local constraints, and its instant number of out-neighbors. The collective goal is to minimize the sum of the cost functions over the set of all constraints. We utilize the *push-sum protocol* to be able to solve this distributed optimization problem. We adapt the push-sum optimization algorithm, which has been studied in context of *unconstrained* optimization so far, to *convex constrained* optimization by introducing an appropriate choice of penalty functions and penalty parameters. Under some additional technical assumptions on the gradients we prove convergence of the distributed penalty-based push-sum algorithm to the optimal value of the global objective function. We apply the proposed penalty-based push-sum algorithm to the problem of *distributed energy management in smart grid* and discuss the advantages of this novel procedure in comparison with existing ones.

I. INTRODUCTION

Due to the emergence of large-scaled networked systems with limited information, distributed multi-agent optimization problems have gained a lot of attention recently. In such systems a number of agents, represented by nodes over some communication graph, aim to optimize a global objective by taking only local information into account. Beside the various applications of distributed optimization such as robust sensor network control [8], signal processing [14], network routing [5], and machine learning [15], [16], an important and promising area of applicability is energy management of future smart grid [3], [9], [17]. Smart grid is equipped with advanced communication technologies enabling efficient and distributed energy management between the grid's users [7], [17]. However, from a technical point of view, it is important to keep communication costs limited and choose a communication protocol that requires minimal coordination between agents and stays robust against changes in the network topology [11]. That is why, in this paper, we develop a communication-based optimization algorithm with the desired features mentioned above.

For this purpose we utilize the *push-sum* communication protocol. This protocol was initially introduced in [2] and used in [15] for distributed optimization. The push-sum protocol is applicable to time-dependent network topology and it can overcome the restrictive assumptions on the

The authors are with the Control Methods and Robotics Lab at TU Darmstadt, Germany.

The work was gratefully supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) within the SPP 1984 "Hybrid and multimodal energy systems: System theoretical methods for the transformation and operation of complex networks".

communication graph structure such as double stochastic communication matrices [2], [15]. The work [4] studied this algorithm over directed and time-varying communication in the case of well-behaved convex functions. The authors in [13] extended the results to a broader class of non-convex functions. However, all works on the push-sum algorithm presented in the literature so far dealt with unconstrained optimization. As in many applications, including energy management in smart grid, agents face a number of constraints. In this paper, we adapt the push-sum algorithm to the case of convex constrained optimization by introducing an appropriate choice of penalty functions and penalty parameters. Under some standard technical assumptions, we prove convergence of the resulting procedure to the optimal value of the system's objective function.

Another contribution of this paper consists in the application of the proposed procedure to the problem of energy management in smart grid. In contrast to the recent communication-based procedure proposed in [17], the penalty-based push-sum algorithm presented in this work is based on a *time-dependent* directed communication topology with *column-stochastic* matrices, where each agent merely needs to know the current number of its out-neighbors to define the elements of the communication matrix at each iteration. Moreover, it uses only one communication step per iteration, whereas the procedure in [17] requires two communication steps per iteration, with a row-stochastic and a column-stochastic communication matrix at the consequent communication iterations. Thus, the proposed penalty-based push-sum algorithm keeps communication costs lower and is able to adapt to changes in the communication topology.

The paper is organized as follows. In Section II we introduce the penalty-based push-sum algorithm and prove its convergence. Section III deals with formulation of the general non-convex energy management problem in smart grid, presents its convex reformulation for which the penalty-based push-sum procedure can be applied, and demonstrates some simulation results. Section IV concludes the paper.

Notations. We will use the following notations throughout this paper: We denote the set of integers by \mathbb{Z} and the set of non-negative integers by \mathbb{Z}^+ . For the metric ρ of a metric space $(X, \rho(\cdot))$ and two subsets $B_1 \subset X$ and $B_2 \subset X$, we let $\rho(B_1, B_2) = \max\{\sup_{x \in B_1} \inf_{y \in B_2} \rho(x, y), \sup_{y \in B_2} \inf_{x \in B_1} \rho(x, y)\}$. We denote the set $\{1, \dots, n\}$ by $[n]$. We use boldface to distinguish between the vectors in a multi-dimensional space and scalars. We denote the dot product of two vectors \mathbf{a} and \mathbf{b} by $\langle \mathbf{a}, \mathbf{b} \rangle$. $\|\cdot\|$ denotes the standard Euclidean

norm, whereas $\|\cdot\|_1$ is used to denote l^1 -norm in the vector space. Throughout this work, all time indices such as t belong to \mathbb{Z}^+ . For vectors $\mathbf{v}_i \in X^d$, $i \in [n]$, of elements in some vector space X (over \mathbb{R}), we let $\bar{\mathbf{v}} = \frac{1}{n} \sum_{i=1}^n \mathbf{v}_i$. We say the function $F : \mathbb{R}^d \rightarrow \mathbb{R}$ is inf-compact, if the set $\{\mathbf{x} \in \mathbb{R}^d : F(\mathbf{x}) \leq a\}$ is compact for all $a \in \mathbb{R}$. The function $1_{\{A\}}(x)$ denotes the indicator of the set A ($1_{\{A\}}(x) = 1$, if $x \in A$ and $1_{\{A\}}(x) = 0$, otherwise). The notation $o(x)$ as $x \rightarrow x_0$ is for some function $f(x)$ such that $\lim_{x \rightarrow x_0} \frac{f(x)}{x} = 0$.

II. PUSH-SUM ALGORITHM FOR DISTRIBUTED CONSTRAINED OPTIMIZATION

In this section we adapt the push-sum algorithm to the case of constrained convex optimization and prove convergence of the resulting procedure.

A. Problem Formulation and Adapted Push-Sum Algorithm

Let us consider the following general problem:

$$\begin{aligned} \min F(\mathbf{z}) &= \sum_{i=1}^n F_i(\mathbf{z}), \quad \mathbf{z} \in \mathbb{R}^d, \\ \text{s.t. } c_1(\mathbf{z}) &\leq 0, c_2(\mathbf{z}) \leq 0, \dots, c_n(\mathbf{z}) \leq 0, \end{aligned} \quad (1)$$

where $F_i : \mathbb{R}^d \rightarrow \mathbb{R}$, $c_i : \mathbb{R}^d \rightarrow \mathbb{R}$, $i = 1, \dots, n$, are some differentiable convex functions. Let \mathbf{f}_i denote the gradient of the function F_i , $i \in [n]$, $\mathbf{f} = \sum_{i=1}^n \mathbf{f}_i$. This problem is formulated in a multi-agent system consisting of n agents. Each agent i has access to its local cost function F_i and its local constraint described by the inequality $c_i(\mathbf{z}) \leq 0$ ¹. By the set S we denote the set of solutions for (1). By F^* we denote the optimal value of the objective function F in the problem (1).

At each time t , node i can only communicate to its out-neighbors in some directed graph $G(t)$, where the graph $G(t)$ has the vertex set $[n]$ and the edge set $E(t)$. We use $N_i^{in}(t)$ and $N_i^{out}(t)$ to denote the in- and out-neighborhoods of node i at time t . Each node i is always considered to be an in- and out-neighbor of itself. We use $d_i(t)$ to denote the out-degree of node i , and we assume that every node i knows its out-degree at every time t . The goal of the agents is to solve distributively the constrained minimization problem (1). We introduce the following standard definition for the sequence $G(t)$.

Definition 1. We say that a sequence of graphs $\{G(t)\}$ is B -strongly connected, if, for any time $t \geq 0$, the graph

$$G(t : t + B) = ([n], E(t) \cup E(t + 1) \cup \dots \cup E(t + B - 1)),$$

is strongly connected. In other words, the union of the graphs over every B time intervals is strongly connected.

In the following analysis we assume that the sequence of the communication graphs $\{G(t)\}$ under consideration is

¹For the sake of notation simplicity, we assume that the local constraint of each agent i is expressed by only one function c_i . The analysis below is applicable to problems, where agents have more than one constraint function.

B -strongly connected, which guarantees enough information “mixing” during communication between agents over time.

To deal with the problem described above, we aim to develop a distributed optimization procedure based on the push-sum protocol [2]. However, this protocol uses specific ratios of local agents’ values to cancel out the effect of information imbalances caused by limited agents’ knowledge of their neighborhoods in time-dependent and directed communication network [2], [4]. That is why optimization methods based on projection onto the set of constraints cannot be applied here, as they violate the balance properties guaranteed by taking the corresponding ratio. To overcome this limitation and to incorporate the constraints of the problem (1) into the optimization algorithm, we leverage the idea of penalty function methods [6]. We choose the convex penalty functions $\{\Psi_i(\mathbf{z})\}_i$ such that²

$$\begin{aligned} \Psi_i(\mathbf{z}) &= g(c_i(\mathbf{z})) \\ g(u) &= \begin{cases} \log\left(\frac{e^u + e^{-u}}{2}\right), & \text{if } u > 0 \\ 0, & \text{if } u \leq 0. \end{cases} \end{aligned} \quad (2)$$

Let $\Psi = \sum_{i=1}^n \Psi_i$, $\psi_i(\mathbf{z})$ denote the gradient of the function $\Psi_i(\mathbf{z})$ and $\psi = \sum_{i=1}^n \psi_i$. Note that for each $i \in [n]$ the vector-function $\psi_i(\mathbf{z})$ is uniformly bounded over \mathbb{R}^d , given that $\nabla c_i(\mathbf{z})$ is uniformly bounded over \mathbb{R}^d . By adding the penalty function value Ψ to the objective function value F in (1), we obtain the following unconstrained penalized optimization problem:

$$\min_{\mathbf{z} \in \mathbb{R}^d} F_t(\mathbf{z}) = F(\mathbf{z}) + r_t \Psi(\mathbf{z}) = \sum_{i=1}^n [F_i(\mathbf{z}) + r_t \Psi_i(\mathbf{z})], \quad (3)$$

where r_t is some positive penalty parameter. Note that as the functions F_i and Ψ_i are convex for all i and $r_t > 0$, the unconstrained problem above is convex. Let S_t denote the set of solutions for (3). The connection between the penalized unconstrained problem (3) and the initial constrained one in (1) is shown in the following proposition (see [6]):

Proposition 1. Let the function F be inf-compact and $r_t \rightarrow \infty$ as $t \rightarrow \infty$. Then S_t and S are not empty and S_t converges to S as t goes to infinity, namely $\lim_{t \rightarrow \infty} \rho(S, S_t) = 0$. Moreover, $\lim_{t \rightarrow \infty} F_t^* = F^*$, where $F_t^* = \min_{\mathbf{z} \in \mathbb{R}^d} F_t(\mathbf{z})$.

Next, we apply the push-sum algorithm from [4] to the penalized problem (3). We proceed with the formal algorithm formulation. At every moment of time $t \in \mathbb{Z}^+$ each node i maintains vector variables $\mathbf{z}_i(t)$, $\mathbf{x}_i(t)$, $\mathbf{w}_i(t) \in \mathbb{R}^d$, as well as a scalar variable $y_i(t)$ such that $y_i(0) = 1$ for all $i \in [n]$.

²Other candidates for penalty function can be found in [1], Chapter 5.

These quantities are updated as follows:

$$\mathbf{w}_i(t+1) = \sum_{j \in N_i^{\text{in}}(t)} \frac{\mathbf{x}_j(t)}{d_j(t)}, \quad (4a)$$

$$y_i(t+1) = \sum_{j \in N_i^{\text{in}}(t)} \frac{y_j(t)}{d_j(t)}, \quad (4b)$$

$$\mathbf{z}_i(t+1) = \frac{\mathbf{w}_i(t+1)}{y_i(t+1)}, \quad (4c)$$

$$\mathbf{x}_i(t+1) = \mathbf{w}_i(t) - a_t[\mathbf{f}_i(\mathbf{z}_i(t+1)) + r_t \psi_i(\mathbf{z}_i(t+1))], \quad (4d)$$

where $a_t \geq 0$ is a time-dependent step size for all t . The version of the push-sum algorithm above corresponds to the one proposed in [4], where the optimization step (4d) is augmented by the penalty term $r_t \psi_i(\mathbf{z}_i(t+1))$. Note that the algorithm above is based on a time-dependent communication topology, where each agent i merely needs to know its current out-degree $d_i(t)$ to follow the algorithm's steps.

B. Convergence of the Algorithm

In what follows, we analyze the convergence property of the algorithm (4) under the following assumptions regarding the gradient functions.

Assumption 1. The gradients \mathbf{f}_i and ∇c_i are uniformly bounded over \mathbb{R}^d for all $i \in [n]$.

Remark 1. Since \mathbf{f}_i is assumed to be bounded for any $i \in [n]$, there exists a positive constant L_i such that $\|\mathbf{f}_i(\mathbf{z})\| \leq L_i$ for any $\mathbf{z} \in \mathbb{R}^d$. Let $L = \max_{i \in [n]} L_i$. Moreover, due to the bounded gradients ∇c_i , the gradient function ψ_i is bounded for any $i \in [n]$ (see (2)). Thus, there exists a positive constant M_i such that $\|\psi_i(\mathbf{z})\| \leq M_i$ for any $\mathbf{z} \in \mathbb{R}^d$. Let $M = \max_{i \in [n]} M_i$.

Assumption 2. The gradients \mathbf{f}_i and ∇c_i are Lipschitz continuous over \mathbb{R}^d for all $i \in [n]$.

Remark 2. According to the choice of the penalty functions in (2), Lipschitz continuity of ∇c_i over \mathbb{R}^d implies Lipschitz continuity of ψ_i over \mathbb{R}^d for all $i \in [n]$. Thus, given Assumption 2, there exist positive constants l_i and m_i such that $\|\mathbf{f}_i(\mathbf{z}_1) - \mathbf{f}_i(\mathbf{z}_2)\| \leq l_i \|\mathbf{z}_1 - \mathbf{z}_2\|$ and $\|\psi_i(\mathbf{z}_1) - \psi_i(\mathbf{z}_2)\| \leq m_i \|\mathbf{z}_1 - \mathbf{z}_2\|$ for any $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^d$ and all $i \in [n]$ respectively. Let $l = \max_{i \in [n]} l_i$ and $m = \max_{i \in [n]} m_i$.

Moreover, we make the following assumption regarding the parameters a_t, r_t .

Assumption 3.

$$a_t \leq a_s \text{ for all } t \geq s, \quad \sum_{t=0}^{\infty} a_t = \infty, \quad (5a)$$

$$r_t \geq 1, \quad r_t \rightarrow \infty, \quad (5b)$$

$$\sum_{t=0}^{\infty} a_t^2 r_t^3 < \infty, \quad r_{t+1} - r_t = o(a_t) \text{ as } a_t \rightarrow 0. \quad (5c)$$

Remark 3. Note that the conditions (5a), (5c) imply that $a_t \rightarrow 0$ as $t \rightarrow \infty$. Appropriate sequences $\{a_t\}$ and $\{r_t\}$ that meet the assumption above can be, for example, $a_t = \frac{1}{t^{0.5+b}}$, $r_t = t^{0.25b}$, where $0 < b < 0.4$.

Let \mathcal{C} denote the constrained solution set of optimization problem (1). For the proof of convergence we will need the implication $\psi(\mathbf{z}) = \mathbf{0} \Rightarrow \Psi(\mathbf{z}) = 0$. Since for any $\mathbf{z} \notin \mathcal{C}$

$$\psi(\mathbf{z}) = \sum_{i=1}^n \frac{e^{2c_i(\mathbf{z})} - 1}{e^{2c_i(\mathbf{z})} + 1} \nabla c_i(\mathbf{z}),$$

we need the following assumption:

Assumption 4. For any $\mathbf{z} \notin \mathcal{C}$ the vectors $\{\nabla c_i(\mathbf{z})\}$ are not linear dependent with coefficients $\left\{ \frac{e^{2c_i(\mathbf{z})} - 1}{e^{2c_i(\mathbf{z})} + 1} \right\}$ correspondingly.

Remark 4. Assumption 4 can be relaxed as follows. There is no point $\mathbf{z} \notin \mathcal{C}$ such that $\nabla c_i(\mathbf{z}) = \mathbf{0}$ for all $i \in [n]$. Indeed, if for any $\mathbf{z} \notin \mathcal{C}$ there exists at least one j such that $\nabla c_j(\mathbf{z}) \neq \mathbf{0}$, the penalty functions Ψ_i , $i \in [n]$, can be replaced by $m_i \Psi_i$ for some constants m_i , $i \in [n]$, to guarantee that the adjusted condition in Assumption 4, namely $\psi(\mathbf{z}) = \sum_{i=1}^n m_i \frac{e^{2c_i(\mathbf{z})} - 1}{e^{2c_i(\mathbf{z})} + 1} \nabla c_i(\mathbf{z}) \neq \mathbf{0}$, holds.

According to the procedure (4), the running average of $\{\mathbf{x}_i(t)\}_{i \in [n]}$, namely $\bar{\mathbf{x}}(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i(t)$, fulfills the following iterations (see also [4]):

$$\begin{aligned} \bar{\mathbf{x}}(t+1) &= \bar{\mathbf{x}}(t) - \frac{a_t}{n} \sum_{i=1}^n [\mathbf{f}_i(\mathbf{z}_i(t+1)) + r_t \psi_i(\mathbf{z}_i(t+1))] \\ &= \bar{\mathbf{x}}(t) - a_t (\mathbf{f}(\bar{\mathbf{x}}(t)) + r_t \psi(\bar{\mathbf{x}}(t))) \\ &\quad - a_t \left[\frac{1}{n} \sum_{i=1}^n \mathbf{f}_i(\mathbf{z}_i(t+1)) - \mathbf{f}(\bar{\mathbf{x}}(t)) \right] \\ &\quad - a_t r_t \left[\frac{1}{n} \sum_{i=1}^n \psi_i(\mathbf{z}_i(t+1)) - \psi(\bar{\mathbf{x}}(t)) \right]. \end{aligned} \quad (6)$$

Some helpful results that will be used in the convergence analysis are presented in the Appendix (see Theorems 2 and 3). In particular, Theorem 2(a) implies that all $\mathbf{x}_i(t+1)$, $i \in [n]$, converge with time to their running average $\bar{\mathbf{x}}(t)$, given that Assumptions 1 and 3 hold.

Further we utilize the following notations:

$$\mathbf{q}(t, \bar{\mathbf{x}}(t)) = \frac{1}{n} \sum_{i=1}^n \mathbf{f}_i(\mathbf{z}_i(t+1)) - \mathbf{f}(\bar{\mathbf{x}}(t)),$$

$$\mathbf{p}(t, \bar{\mathbf{x}}(t)) = \frac{1}{n} \sum_{i=1}^n \psi_i(\mathbf{z}_i(t+1)) - \psi(\bar{\mathbf{x}}(t)).$$

We will use the following lemma which bounds the norms of the vectors $\mathbf{q}(t, \bar{\mathbf{x}}(t))$ and $\mathbf{p}(t, \bar{\mathbf{x}}(t))$ introduced above.

Lemma 1. Let Assumptions 1-3 hold. Then, there exist $q(t)$ and $p(t)$ such that the following holds for the process (4):

$$\|\mathbf{q}(t, \bar{\mathbf{x}}(t))\| \leq \frac{l}{n} \sum_{i=1}^n \|\mathbf{z}_i(t+1) - \bar{\mathbf{x}}(t)\| = q(t),$$

$$\|\mathbf{p}(t, \bar{\mathbf{x}}(t))\| \leq \frac{m}{n} \sum_{i=1}^n \|\mathbf{z}_i(t+1) - \bar{\mathbf{x}}(t)\| = p(t),$$

such that

$$\sum_{t=0}^{\infty} a_t r_t^3 q(t) < \infty, \quad \sum_{t=0}^{\infty} a_t r_t^3 p(t) < \infty.$$

Proof. Due to Assumption 2,

$$\begin{aligned} \|\mathbf{q}(t, \bar{\mathbf{x}}(t))\| &= \left\| \frac{1}{n} \left[\sum_{i=1}^n \mathbf{f}_i(\mathbf{z}_i(t+1)) - \sum_{i=1}^n \mathbf{f}_i(\bar{\mathbf{x}}(t)) \right] \right\| \\ &\leq \frac{l}{n} \sum_{i=1}^n \|\mathbf{z}_i(t+1) - \bar{\mathbf{x}}(t)\|. \end{aligned}$$

Let $q(t) = \frac{l}{n} \sum_{i=1}^n \|\mathbf{z}_i(t+1) - \bar{\mathbf{x}}(t)\|$. Next, let us consider the series

$$\sum_{t=1}^{\infty} b_t a_t \|\mathbf{f}_i(\mathbf{z}_i(t+1)) + r_t \psi_i(\mathbf{z}_i(t+1))\|_1.$$

If $b_t = a_t r_t^3$,

$$\sum_{t=1}^{\infty} b_t a_t \|\mathbf{f}_i(\mathbf{z}_i(t+1)) + r_t \psi_i(\mathbf{z}_i(t+1))\|_1 < \infty,$$

as $\mathbf{f}_i(\mathbf{z}_i(t+1))$ and $\psi_i(\mathbf{z}_i(t+1))$ are bounded, (5b) and (5c) hold for a_t and r_t . Thus, we can use Theorem 2 from Appendix to conclude that

$$\sum_{t=0}^{\infty} a_t r_t^3 q(t) = l \sum_{t=0}^{\infty} \frac{a_t r_t^3}{n} \sum_{i=1}^n \|\mathbf{z}_i(t+1) - \bar{\mathbf{x}}(t)\| < \infty.$$

Analogously, one can show that $\sum_{t=0}^{\infty} a_t r_t^3 p(t) < \infty$. \square

Remark 5. Note that due to the choice of the parameters in Assumption 3 and the fact that under Assumptions 1-3 both $p(t)$ and $q(t)$ tend to 0 as $t \rightarrow \infty$ (see Theorem 2 in Appendix), $\sum_{t=0}^{\infty} a_t^q r_t^{q_2} q^2(t) < \infty$, $\sum_{t=0}^{\infty} a_t^{p_1} r_t^{p_2} p^2(t) < \infty$, and $\sum_{t=0}^{\infty} a_t^2 r_t^2 p(t) q(t) < \infty$ for all integers $q_1, p_1 \geq 1$, $q_2, p_2 \in \{0, 1, 2, 3\}$.

Now we state the main result for the penalty-based push-sum algorithm (4).

Theorem 1. Let the function F in the problem (1) be inf-compact. Let Assumptions 1-3 hold. Then all local variables $\mathbf{z}_i(t+1)$, $i \in [n]$, in the procedure (4) reach a consensus as $t \rightarrow \infty$ and each limit point of this consensus corresponds to a solution to the problem (1), given that the sequence of the communication graphs $\{G(t)\}$ under consideration is B -strongly connected.

Proof. First, we will show that $\lim_{t \rightarrow \infty} \Psi(\bar{\mathbf{x}}(t)) = 0$. In particular, it will mean that all limit points of $\{\bar{\mathbf{x}}(t)\}$ belong to the feasible set $C = \{\mathbf{z} \in \mathbb{R}^d \mid c_i(\mathbf{z}) \leq 0, i \in [n]\}$. Taking the Mean-value Theorem and relation (6) into account, and using the notation

$$\tilde{\mathbf{f}}(t, \bar{\mathbf{x}}(t)) = \mathbf{f}(\bar{\mathbf{x}}(t)) + r_t \psi(\bar{\mathbf{x}}(t)) + \mathbf{q}(t, \bar{\mathbf{x}}(t)) + r_t \mathbf{p}(t, \bar{\mathbf{x}}(t)), \quad (7)$$

we get that for some $\theta \in [0, 1]$ and $\tilde{\mathbf{x}}(t) = \bar{\mathbf{x}}(t) - \theta a_t \tilde{\mathbf{f}}(t, \bar{\mathbf{x}}(t))$

$$\begin{aligned} \Psi(\bar{\mathbf{x}}(t+1)) &= \Psi(\bar{\mathbf{x}}(t)) - a_t \langle \psi(\bar{\mathbf{x}}(t)), \tilde{\mathbf{f}}(t, \bar{\mathbf{x}}(t)) \rangle \\ &\quad + a_t [\langle \psi(\bar{\mathbf{x}}(t)), \tilde{\mathbf{f}}(t, \bar{\mathbf{x}}(t)) \rangle - \langle \psi(\tilde{\mathbf{x}}(t)), \tilde{\mathbf{f}}(t, \bar{\mathbf{x}}(t)) \rangle]. \end{aligned} \quad (8)$$

According to Remark 2 and the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} \langle \psi(\bar{\mathbf{x}}(t)), \tilde{\mathbf{f}}(t, \bar{\mathbf{x}}(t)) \rangle - \langle \psi(\tilde{\mathbf{x}}(t)), \tilde{\mathbf{f}}(t, \bar{\mathbf{x}}(t)) \rangle &\leq \|\tilde{\mathbf{f}}(t, \bar{\mathbf{x}}(t))\| \|\psi(\bar{\mathbf{x}}(t)) - \psi(\tilde{\mathbf{x}}(t))\| \\ &\leq \|\tilde{\mathbf{f}}(t, \bar{\mathbf{x}}(t))\| mn \|\bar{\mathbf{x}}(t) - \tilde{\mathbf{x}}(t)\| \\ &\leq \theta m n a_t \|\tilde{\mathbf{f}}(t, \bar{\mathbf{x}}(t))\|^2, \end{aligned} \quad (9)$$

where m is the constant defined in Remark 2.

Next, using Remark 2, Lemma 1, and due to the Cauchy-Schwarz inequality, we get

$$\begin{aligned} \|\tilde{\mathbf{f}}(t, \bar{\mathbf{x}}(t))\|^2 &\leq \|\mathbf{f}(\bar{\mathbf{x}}(t))\|^2 \\ &\quad + r_t^2 \|\psi(\bar{\mathbf{x}}(t))\|^2 + \|\mathbf{q}(t, \bar{\mathbf{x}}(t))\|^2 + r_t^2 \|\mathbf{p}(t, \bar{\mathbf{x}}(t))\|^2 \\ &\quad + 2\|\mathbf{f}(\bar{\mathbf{x}}(t))\| r_t \|\psi(\bar{\mathbf{x}}(t))\| + 2\|\mathbf{f}(\bar{\mathbf{x}}(t))\| \|\mathbf{q}(t, \bar{\mathbf{x}}(t))\| \\ &\quad + 2r_t \|\mathbf{q}(t, \bar{\mathbf{x}}(t))\| \|\mathbf{p}(t, \bar{\mathbf{x}}(t))\| + 2r_t \|\psi(\bar{\mathbf{x}}(t))\| \|\mathbf{q}(t, \bar{\mathbf{x}}(t))\| \\ &\quad + 2r_t^2 \|\psi(\bar{\mathbf{x}}(t))\| \|\mathbf{p}(t, \bar{\mathbf{x}}(t))\| + 2\|\mathbf{f}(\bar{\mathbf{x}}(t))\| \|\mathbf{p}(t, \bar{\mathbf{x}}(t))\| \\ &\leq k_1 \left(\frac{1}{2} + q(t) + r_t p(t) \right) + k_2 \left(\frac{1}{2} r_t^2 + r_t q(t) + r_t^2 p(t) \right) \\ &\quad + k_3 r_t + q^2(t) + r_t^2 p^2(t) + 2r_t p(t) q(t) := h_0(t) \end{aligned} \quad (10)$$

for some positive constants k_1, k_2 , and k_3 . Finally, using the definition of $\tilde{\mathbf{f}}(t, \bar{\mathbf{x}}(t))$ in (7) and the Cauchy-Schwarz inequality again, we obtain for some positive constants k_4 and k_5 that

$$\begin{aligned} \langle \psi(\bar{\mathbf{x}}(t)), \tilde{\mathbf{f}}(t, \bar{\mathbf{x}}(t)) \rangle &\geq \langle \psi(\bar{\mathbf{x}}(t)), \mathbf{f}(\bar{\mathbf{x}}(t)) \rangle + r_t \|\psi(\bar{\mathbf{x}}(t))\|^2 \\ &\quad - k_4 \|\mathbf{q}(t, \bar{\mathbf{x}}(t))\| - k_5 r_t \|\mathbf{p}(t, \bar{\mathbf{x}}(t))\| \\ &\geq \|\psi(\bar{\mathbf{x}}(t))\| (-\|\mathbf{f}(\bar{\mathbf{x}}(t))\| + r_t \|\psi(\bar{\mathbf{x}}(t))\|) \\ &\quad - k_4 q(t) - k_5 r_t p(t) \\ &\geq \|\psi(\bar{\mathbf{x}}(t))\| \mathbb{1}_{\{\|\psi(\bar{\mathbf{x}}(t))\| \neq 0\}} - k_4 q(t) - k_5 r_t p(t), \end{aligned} \quad (11)$$

where the first inequality is due to the bounded $\|\psi\|$ and the last inequality is due to the fact that $0 < r_t \rightarrow \infty$ as $t \rightarrow \infty$ and $\|\mathbf{f}\|$ are bounded.

By substituting (9)-(11) to (8) we can write:

$$\begin{aligned} \Psi(\bar{\mathbf{x}}(t+1)) &\leq \Psi(\bar{\mathbf{x}}(t)) - a_t \|\psi(\bar{\mathbf{x}}(t))\| \mathbb{1}_{\{\|\psi(\bar{\mathbf{x}}(t))\| \neq 0\}} + h(t), \end{aligned}$$

where $h(t) = a_t^2 (h_0(t) + k_4 q(t) + k_5 r_t p(t))$. According to the choice of a_t and r_t , Lemma 1, and Remark 5, we obtain that $\sum_{t=0}^{\infty} h(t) < \infty$. Thus, using the well-known result on the sequences of non-negative variables presented in Theorem 3 (see Appendix), we conclude that $\lim_{t \rightarrow \infty} \Psi(\bar{\mathbf{x}}(t))$ exists, is finite, and $\sum_{t=0}^{\infty} a_t \|\psi(\bar{\mathbf{x}}(t))\|^2 < \infty$. Thus, $\liminf_{t \rightarrow \infty} \|\psi(\bar{\mathbf{x}}(t))\| = 0$, since $\sum_{t=0}^{\infty} a_t = \infty$ (see (5a)). It implies existence of a subsequence $\{t_k\} \subseteq \{t\}$ such that $\lim_{k \rightarrow \infty} \|\psi(\bar{\mathbf{x}}(t_k))\| = 0$ and, as, given Assumption

4, $\|\psi(\mathbf{z})\| = 0$ if and only if $\Psi(\mathbf{z}) = 0$, we conclude that $\lim_{k \rightarrow \infty} \Psi(\bar{\mathbf{x}}(t_k)) = 0$. Thus,

$$\lim_{t \rightarrow \infty} \Psi(\bar{\mathbf{x}}(t)) = 0. \quad (12)$$

Next, let us notice that $F(\bar{\mathbf{x}}(t+1)) = F_t(\bar{\mathbf{x}}(t+1)) - r_t \Psi(\bar{\mathbf{x}}(t+1)) = F_{t+1}(\bar{\mathbf{x}}(t+1)) - r_{t+1} \Psi(\bar{\mathbf{x}}(t+1))$. Hence, taking into account that $\Psi(\bar{\mathbf{x}}(t)) = o(1)$ as $t \rightarrow \infty$ (see (12)), $r_{t+1} - r_t = o(a_t)$, $a_t \rightarrow 0$ as $t \rightarrow \infty$, and using Mean-value Theorem, we get

$$\begin{aligned} F_{t+1}(\bar{\mathbf{x}}(t+1)) &= F_t(\bar{\mathbf{x}}(t+1)) + (r_{t+1} - r_t) \Psi(\bar{\mathbf{x}}(t+1)) \\ &= F_t(\bar{\mathbf{x}}(t)) - a_t \tilde{\mathbf{f}}(t, \bar{\mathbf{x}}(t)) + o(a_t) \\ &= F_t(\bar{\mathbf{x}}(t)) - a_t \langle \nabla F_t(\bar{\mathbf{x}}(t)), \tilde{\mathbf{f}}(t, \bar{\mathbf{x}}(t)) \rangle + o(a_t) \\ &\quad + a_t \langle \nabla F_t(\bar{\mathbf{x}}(t)) - \nabla F_t(\mathbf{x}'(t)), \tilde{\mathbf{f}}(t, \bar{\mathbf{x}}(t)) \rangle, \end{aligned} \quad (13)$$

where $\mathbf{x}'(t) = \bar{\mathbf{x}}(t) - \beta a_t \tilde{\mathbf{f}}(t, \bar{\mathbf{x}}(t))$ for some $\beta \in [0, 1]$.

According to Assumption 2, there exists some $l_1 > 0$ such that

$$\begin{aligned} &\langle \nabla F_t(\bar{\mathbf{x}}(t)) - \nabla F_t(\mathbf{x}'(t)), \tilde{\mathbf{f}}(t, \bar{\mathbf{x}}(t)) \rangle \\ &\leq l_1 a_t r_t \|\tilde{\mathbf{f}}(t, \bar{\mathbf{x}}(t))\|^2 \leq l_1 a_t r_t h_0(t), \end{aligned}$$

where for the first inequality we used the Cauchy-Schwarz inequality and the last inequality is due to (10). Hence, due to the Cauchy-Schwarz inequality and the definition of $\tilde{\mathbf{f}}(t, \bar{\mathbf{x}}(t))$ in (7), we obtain from (13) that

$$\begin{aligned} F_{t+1}(\bar{\mathbf{x}}(t+1)) &\leq F_t(\bar{\mathbf{x}}(t)) \\ &\quad - a_t \|\nabla F_t(\bar{\mathbf{x}}(t))\|^2 - a_t \langle \mathbf{f}(\bar{\mathbf{x}}(t)), \mathbf{q}(t, \bar{\mathbf{x}}(t)) \rangle \\ &\quad - a_t r_t \langle \psi(\bar{\mathbf{x}}(t)), \mathbf{q}(t, \bar{\mathbf{x}}(t)) \rangle - a_t r_t \langle \mathbf{f}(\bar{\mathbf{x}}(t)), \mathbf{p}(t, \bar{\mathbf{x}}(t)) \rangle \\ &\quad - a_t r_t^2 \langle \psi(\bar{\mathbf{x}}(t)), \mathbf{p}(t, \bar{\mathbf{x}}(t)) \rangle + l_1 a_t^2 r_t h_0(t) + o(a_t) \\ &\leq F_t(\bar{\mathbf{x}}(t)) - a_t (\|\nabla F_t(\bar{\mathbf{x}}(t))\|^2 + o(1)) \\ &\quad + a_t q(t) \|\mathbf{f}(\bar{\mathbf{x}}(t))\| + a_t r_t q(t) \|\psi(\bar{\mathbf{x}}(t))\| \\ &\quad + a_t r_t p(t) \|\mathbf{f}(\bar{\mathbf{x}}(t))\| + a_t r_t^2 p(t) \|\psi(\bar{\mathbf{x}}(t))\| + l_1 a_t^2 r_t h_0(t) \\ &= F_t(\bar{\mathbf{x}}(t)) - a_t (\|\nabla F_t(\bar{\mathbf{x}}(t))\|^2 + o(1)) + h_1(t), \end{aligned}$$

where

$$\begin{aligned} h_1(t) &= a_t q(t) \|\mathbf{f}(\bar{\mathbf{x}}(t))\| + a_t r_t q(t) \|\psi(\bar{\mathbf{x}}(t))\| \\ &\quad + a_t r_t p(t) \|\mathbf{f}(\bar{\mathbf{x}}(t))\| + a_t r_t^2 p(t) \|\psi(\bar{\mathbf{x}}(t))\| + l_1 a_t^2 r_t h_0(t). \end{aligned}$$

Thus, $\sum_{t=0}^{\infty} h_1(t) < \infty$, due to Assumption 1, the choice of a_t and r_t , and Lemma 1 (see Remark 5). Hence, according to Theorem 3 from Appendix, we can conclude that

$$F_t(\bar{\mathbf{x}}(t)) \text{ has a limit as } t \rightarrow \infty. \quad (14)$$

Moreover, $\sum_{t=1}^{\infty} a_t \|\nabla F_t(\bar{\mathbf{x}}(t))\|^2 < \infty$, which, due to (5a), implies $\liminf_{t \rightarrow \infty} \|\nabla F_t(\bar{\mathbf{x}}(t))\| = 0$. Let us choose a subsequence $\{t_k\} \subseteq \{t\}$ such that $\lim_{k \rightarrow \infty} \|\nabla F_{t_k}(\bar{\mathbf{x}}(t_k))\| = 0$. Due to convexity of F_t over \mathbb{R}^d for all t , the last limit implies that $\lim_{k \rightarrow \infty} [F_{t_k}(\bar{\mathbf{x}}(t_k)) - F_{t_k}^*] = 0$, where $F_{t_k}^* = \min_{\mathbf{x} \in \mathbb{R}^d} F_{t_k}(\mathbf{x})$. Next, due to Proposition 1 and as $r_t \rightarrow \infty$, we conclude that $\lim_{k \rightarrow \infty} F_{t_k}(\bar{\mathbf{x}}(t_k)) - F^* = 0$, which together with (14) implies that $\lim_{t \rightarrow \infty} F_t(\bar{\mathbf{x}}(t)) = F^*$ and, hence, every limit point of $\bar{\mathbf{x}}(t)$ is a solution to the problem

(1). Finally, by invoking Theorem 2(a), we conclude the result. \square

Remark 6. Note that if, additionally to the conditions in Theorem 1, the function F is assumed to be strictly convex, then there exists a unique solution \mathbf{z}^* to the problem (1). In this case, Theorem 1 implies convergence of all $\mathbf{z}_i(t)$ evolving according to the algorithm (4) to this optimum \mathbf{z}^* .

III. APPLICATIONS IN ENERGY MANAGEMENT

We consider a problem of energy management formulated and analyzed in [17]. Let \mathcal{N}_g and \mathcal{N}_d be the sets of distributed generators and responsive demands in a power grid with N_g and N_d elements respectively. Let $N = N_g + N_d$. A directed connected time-dependent graph $G(t) = ([N], \mathcal{E}(t))$ is used to represent the communication topology of the network in the grid, where $[N] = \mathcal{N}_g \cup \mathcal{N}_d$ is the set of the nodes containing generators and demands and $\mathcal{E}(t) \subseteq [N] \times [N]$ is the edge set. Note that $(j, i) \in \mathcal{E}(t)$ if and only if the node $i \in [N]$ can receive information from node $j \in [N]$ at time t . In contrast to the previous works [10], [17], in this paper we focus on a broader class of the communication topology containing time-dependent graphs and requiring each user to know only its current out-degree to construct an appropriate communication matrix. For this purpose, we will apply the penalty-based push-sum algorithm introduced and analyzed in Section II to the distributed optimization formulated below.

We consider the following generation and demand capacities in the system:

$$p_i \in [p_i^m, P_i^M], \quad i \in \mathcal{N}_g, \quad p_j \in [p_j^m, P_j^M], \quad j \in \mathcal{N}_d.$$

The cost function $C_i : \mathbb{R} \rightarrow \mathbb{R}$ of each generator $i \in \mathcal{N}_g$ is:

$$C_i(p_i) = \begin{cases} a_i p_i^2 + b_i p_i + c_i, & \text{if } p_i \in [p_i^m, P_i^M], \\ (2a_i p_i^m + b_i) p_i, & \text{if } p_i \leq p_i^m, \\ (2a_i P_i^M + b_i) p_i, & \text{if } p_i \geq P_i^M. \end{cases} \quad (15)$$

where a_i, b_i, c_i are positive fitting parameters. Thus, the cost functions are strongly convex functions.

The utility function $U_j : \mathbb{R} \rightarrow \mathbb{R}$ of each demand $j \in \mathcal{N}_d$ has the following properties:

$$\begin{aligned} U_j(0) &= 0, \quad \frac{dU_j}{dp_j} > 0 \text{ (non-decreasing),} \\ \exists K_1, K_2 : \frac{dU_j}{dp_j} &< K_1, \\ K_2 &\leq \frac{d^2 U_j}{(dp_j)^2} \leq 0 \text{ (get saturated).} \end{aligned} \quad (16)$$

Thus, the utility functions are concave functions.

Let $\mathbf{p} \in \mathbb{R}^N$ be the vector with coordinates $p_i, i \in [N]$. The goal in the power grid is to solve distributively the following energy management problem³:

$$\min_{\mathbf{p}} \sum_{i \in \mathcal{N}_g} C_i(p_i) - \sum_{j \in \mathcal{N}_d} U_j(p_j) \quad (17)$$

³For more details on the problem formulation see [17].

$$s.t. \sum_{i \in \mathcal{N}_g} (p_i - l_i p_i^2) = \sum_{j \in \mathcal{N}_d} p_j \quad (17a)$$

$$p_i^m \leq p_i \leq P_i^M, \quad i \in \mathcal{N}_g \quad (17b)$$

$$p_j^m \leq p_j \leq P_j^M, \quad j \in \mathcal{N}_d. \quad (17c)$$

In the problem above the constraint (17a) corresponds to the balance between the generated and the demanded power in the network, where each parameter l_i , $i \in \mathcal{N}_g$, corresponds to the coefficient of the transmission losses induced by the generator i and satisfies $0 \leq l_i < a_i$. Note that the problem (17) is non-convex due to the non-convex constraint defined by (17a).

A. Problem reformulation with constraints based on local information

To implement the distributed penalized push-sum algorithm to the energy management problem (17), we need to find its appropriate convex reformulation such that any solution to this reformulation provides a solution for (17). Moreover, as the constraint (17a) contains the information on the ‘‘loss’’ parameter l_i of each $i \in \mathcal{N}_g$, we aim to find a reformulation, where no constraint requires knowledge about the local properties of other nodes in the network. First of all, let us notice that the problem (17) is equivalent to the following one:

$$\min_{\mathbf{p}, \mathbf{v}} \sum_{i \in \mathcal{N}_g} C_i(p_i) - \sum_{j \in \mathcal{N}_d} U_j(p_j) \quad (18)$$

$$s.t. \sum_{i \in \mathcal{N}_g} (p_i - v_i) = \sum_{j \in \mathcal{N}_d} p_j \quad (18a)$$

$$p_i^m \leq p_i \leq P_i^M, \quad i \in \mathcal{N}_g \quad (18b)$$

$$p_j^m \leq p_j \leq P_j^M, \quad j \in \mathcal{N}_d \quad (18c)$$

$$v_i = l_i p_i^2, \quad i \in \mathcal{N}_g, \quad (18d)$$

where $\mathbf{v} \in \mathbb{R}^{\mathcal{N}_g}$ is the vector with coordinates v_i , $i \in \mathcal{N}_g$. Thus, the strategy p_i of each generator is augmented by the auxiliary parameter v_i . However, the problem (18) is still non-convex due to the constraints in (18d). Following the idea in [17], we present a new reformulation, where each non-convex equality constraint is replaced by the corresponding convex inequality one⁴. Thus, we obtain the convex optimization problem

$$\min_{\mathbf{p}, \mathbf{v}} \sum_{i \in \mathcal{N}_g} C_i(p_i) - \sum_{j \in \mathcal{N}_d} U_j(p_j) \quad (19)$$

$$s.t. \sum_{i \in \mathcal{N}_g} (p_i - v_i) = \sum_{j \in \mathcal{N}_d} p_j \quad (19a)$$

$$p_i^m \leq p_i \leq P_i^M, \quad i \in \mathcal{N}_g \quad (19b)$$

$$p_j^m \leq p_j \leq P_j^M, \quad j \in \mathcal{N}_d \quad (19c)$$

$$v_i \geq l_i p_i^2, \quad i \in \mathcal{N}_g. \quad (19d)$$

⁴According to the definition of convex sets in [1], the set defined by $v_i = l_i p_i^2$, $v_i \neq 0$ is non-convex as it only contains the separate points $\{-\sqrt{v_i/l_i}, +\sqrt{v_i/l_i}\}$. However, the set defined by $v_i \geq l_i p_i^2$ is convex as it consists of the interval $[-\sqrt{v_i/l_i}, +\sqrt{v_i/l_i}]$ and by that admits an interpolation between any two points in the set that is in return part of the set.

Next, we establish the relation between the convex problem (19) and the initial one (17). This will be done under the following two technical assumptions.

Assumption 5. The upper and lower bounds for the feasible power generation and demand (see (17b) and (17c)) satisfy

$$\sum_{j \in \mathcal{N}_d} P_j^M \geq \sum_{i \in \mathcal{N}_g} (p_i^m - l_i (p_i^m)^2).$$

The assumption above repeats the sufficient condition for an appropriate convex reformulation of the problem (17) presented in [17] (see also Remark 1 in [17]). The next assumption is the Slater constraint qualification for the reformulated problem (19). It will enable the relation analysis based on the Karush-Kuhn-Tucker conditions for the optimal primal dual pair of the problem (19).

Assumption 6. There exists a feasible point $(\hat{\mathbf{p}}, \hat{\mathbf{v}})$ for the problem (19) such that $\sum_{i \in \mathcal{N}_g} (\hat{p}_i - \hat{v}_i) = \sum_{j \in \mathcal{N}_d} \hat{p}_j$, $p_i^m < \hat{p}_i < P_i^M$, $i \in \mathcal{N}_g$, $p_j^m < \hat{p}_j < P_j^M$, $j \in \mathcal{N}_d$, $\hat{v}_i > l_i \hat{p}_i^2$, $i \in \mathcal{N}_g$.

The next proposition states the desired relation between the initial problem (17) and the convex one (19) above.

Proposition 2. Let Assumptions 5 and 6 hold. Then any solution to the problem (19) is a solution to the problem (17).

Proof. See Appendix. \square

The optimization problem (19) can be considered a particular case of the general distributed optimization problem (1). Indeed, let $\mathbf{z} = (\mathbf{p}, \mathbf{v})$ be the vector of joint strategies of the generators and responsive demands in the network, $F_i(\mathbf{z}) = C_i(p_i)$, if $i \in \mathcal{N}_g$, $F_j(\mathbf{z}) = -U_j(p_j)$, if $j \in \mathcal{N}_d$, and the constraints (19a)-(19d) be distributed over these agents as follows:

$$LC_i = \{c_i^1(\mathbf{z}) = p_i - P_i^M \leq 0, c_i^2(\mathbf{z}) = p_i^m - p_i \leq 0,$$

$$c_i^3(\mathbf{z}) = \sum_{i \in \mathcal{N}_g} (p_i - v_i) - \sum_{j \in \mathcal{N}_d} p_j \leq 0,$$

$$c_i^4(\mathbf{z}) = - \sum_{i \in \mathcal{N}_g} (p_i - v_i) + \sum_{j \in \mathcal{N}_d} p_j \leq 0,$$

$$c_i^5(\mathbf{z}) = (l_i p_i^2 - v_i) 1_{\{p_i \in [p_i^m, P_i^M]\}} + (2l_i p_i^m - v_i) 1_{\{p_i < p_i^m\}} + (2l_i P_i^M - v_i) 1_{\{p_i > P_i^M\}} \leq 0\},$$

if $i \in \mathcal{N}_g$,

$$LC_j = \{c_j^1(\mathbf{z}) = p_j - P_j^M \leq 0, c_j^2(\mathbf{z}) = p_j^m - p_j \leq 0\},$$

if $j \in \mathcal{N}_d$.

Note that the condition $c_i^5(\mathbf{z}) \leq 0$ for $i \in \mathcal{N}_g$ above corresponds to the constraint (19d). We modified this constraint without changing the problem (due to existence of the hard constraints (19b)) to be able to use the result from Theorem 1 requiring bounded gradients of the constraint functions. Given the properties of the cost and utility functions (see (15) and (16)), the objective function $F(\mathbf{z}) = \sum_{i \in \mathcal{N}_g} F_i(\mathbf{z}) + \sum_{j \in \mathcal{N}_d} F_j(\mathbf{z})$ is strongly convex and, hence, inf-compact.

Moreover, Assumptions 1 and 2 hold for the gradients of the functions F_i , F_j and local constraint functions $c_i^1, c_i^2, c_i^3, c_i^4, c_i^5, c_j^1, c_j^2, i \in \mathcal{N}_g, j \in \mathcal{N}_d$. Thus, the problem (19) is equivalent to

$$\begin{aligned} \min_{\mathbf{z}} F(\mathbf{z}) &= \sum_{i \in \mathcal{N}_g} F_i(\mathbf{z}) + \sum_{j \in \mathcal{N}_d} F_j(\mathbf{z}), \\ \text{s.t. } \mathbf{z} &\in \left(\bigcap_{i \in \mathcal{N}_g} LC_i \right) \cap \left(\bigcap_{j \in \mathcal{N}_d} LC_j \right). \end{aligned} \quad (20)$$

and the following result can be formulated

Corollary 1. *Let Assumptions 5 and 6 hold for the problem (17). Then under an appropriate choice of the parameters a_t, r_t (see Assumption 6), the penalty-based push-sum algorithm (4) applied to the reformulated problem (20) converges to the optimal solution to (17) as time tends to infinity.*

B. Simulation Results

In this section we will substantiate our theoretical result stated in Corollary 1 for the energy management problem (17) with simulations. For this purpose, we verify Theorem 1 by comparing the optimum \mathbf{p}^* of the problem (17) with the iterations of the penalty-based push-sum algorithm presented in Section II. Note that Assumption 4 is met for the proposed constraint set.

For our simulation we use a small setup of two generator and two consumer nodes. The cost functions of the generators are as in (15), whereas the utility functions of the demand nodes are

$$U_j(p_j) = \begin{cases} \omega_j p_j - \alpha_j p_j^2, & p_j \leq \frac{\omega_j}{2K_j \alpha_j} \\ (\omega_j - \frac{\omega_j}{K_j}) p_j - \frac{\omega_j^2}{4K_j^2 \alpha_j}, & p_j > \frac{\omega_j}{2K_j \alpha_j} \end{cases}, \quad (21)$$

where $K_j < 1$ is a positive constant for each $j \in \mathcal{N}_d$. Thus, the properties (16) are met. For the parameters for the cost functions, the lower and upper bounds on p_i and p_j , as well as for the transmission loss coefficients $l_i, i \in \mathcal{N}_g$, we rely on settings in [17]. We model our time-varying communication architecture with a changing signal s_t that chooses the current graph $G(s(t))$ sequentially from the set $\mathcal{G} = \{G_1, G_2\}$, where G_1 and G_2 are not strongly connected but their union is. The communication architecture is depicted in Figure 1. The results of our simulations

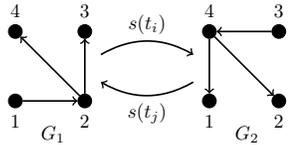


Fig. 1. Time-dependent directed communication architecture.

are shown in Table I and Figure 2. The table states the optimal values \mathbf{p}^* of the generators and consumers together with the result of the push-sum algorithm after 3×10^4 iterations. The figure shows the convergence of the relative error $\text{error}_k = \frac{|p_k - p_k^*|}{p_k^*}, k \in [4]$, to zero as time runs. We can notice that the relative error at the demand nodes

($k = 3, 4$) approaches 0 already after 500 iterations, whereas the generator's errors need significantly more time to get close to 0. This effect is due to a more complex structure of generators' constraints, for which an optimal choice of penalty parameters needs to be studied in the future work.

	p_1^G	p_2^G	p_1^D	p_2^D
\mathbf{p}^*	81.98	124.80	100.34	100.00
\mathbf{p}	81.69	122.75	99.61	99.64

TABLE I

OPTIMAL \mathbf{p}^* VS ALGORITHM OUTPUT AFTER 3×10^4 ITERATIONS.

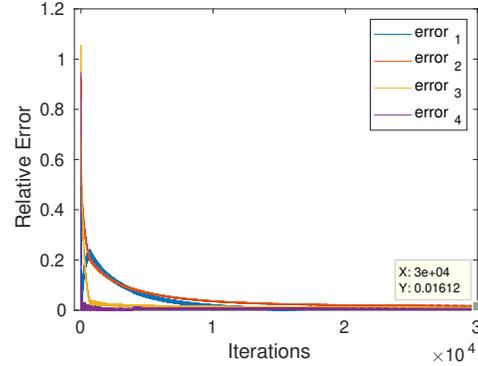


Fig. 2. Convergence of the relative error $\text{error}_k, k \in [4]$, between optimum p_k^* and algorithm output p_k . Highest relative error at 0.016 after $3 \cdot 10^4$ iterations.

IV. CONCLUSION

In this paper we extended the distributed push-sum algorithm to the case of constrained convex optimization. The penalty-based push-sum algorithm was presented and its convergence to a system's optimum was proven. We demonstrated applicability of the proposed procedure to distributed energy management in smart grid. Future work will focus on such questions as the convergence rate of the penalty-based push-sum algorithm and its dependence on the communication topology as well as an optimal choice of penalty functions and penalty parameters.

REFERENCES

- [1] D.P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1999.
- [2] D. Kempe, A. Dobra, and J. Gehrke. Gossip-based computation of aggregate information. In *Proceedings - Annual IEEE Symposium on Foundations of Computer Science, FOCS*, volume 2003-January, pages 482–491. IEEE Computer Society, 2003.
- [3] J. Mohammadi, G. Hug, and S. Kar. Asynchronous distributed approach for DC optimal power flow. In *2015 IEEE Eindhoven PowerTech*, pages 1–6, June 2015.
- [4] A. Nedić and A. Olshevsky. Distributed optimization over time-varying directed graphs. *IEEE Trans. Automat. Contr.*, 60(3):601–615, 2015.
- [5] G. Neglia, G. Reina, and S. Alouf. Distributed gradient optimization for epidemic routing: a preliminary evaluation. In *Proc. of 2nd IFIP Wireless Days 2009*, December 2009.
- [6] G. Pflug. On the convergence of a penalty-type stochastic optimization procedure. *Journal of Information and Optimization Sciences*, 2(3):249–258, 1981.
- [7] M. Pipattanasomporn, H. Feroze, and S. Rahman. Multi-agent systems in a distributed smart grid: Design and implementation. In *2009 IEEE/PES Power Systems Conference and Exposition*, pages 1–8, March 2009.

- [8] M. Rabbat and R. Nowak. Distributed optimization in sensor networks. In *Proceedings of the 3rd International Symposium on Information Processing in Sensor Networks*, IPSN '04, pages 20–27, New York, NY, USA, 2004. ACM.
- [9] N. Rahbari-Asr, U. Ojha, Z. Zhang, and M. Chow. Incremental welfare consensus algorithm for cooperative distributed generation/demand response in smart grid. *IEEE Transactions on Smart Grid*, 5(6):2836–2845, Nov 2014.
- [10] N. Rahbari-Asr, U. Ojha, Z. Zhang, and M. Chow. Incremental welfare consensus algorithm for cooperative distributed generation/demand response in smart grid. *IEEE Transactions on Smart Grid*, 5(6):2836–2845, Nov 2014.
- [11] W. Ren and R. W. Beard. Consensus seeking in multiagent systems under dynamically changing interaction topologies. *IEEE Transactions on Automatic Control*, 50(5):655–661, May 2005.
- [12] H. Robbins and D. Siegmund. A convergence theorem for non negative almost supermartingales and some applications. In *Herbert Robbins Selected Papers*, pages 111–135. Springer, 1985.
- [13] T. Tatarenko and B. Touri. Non-convex distributed optimization. *IEEE Transactions on Automatic Control*, 62(8):3744–3757, Aug 2017.
- [14] B. Touri and A. Nedić. When infinite flow is sufficient for ergodicity. In *49th IEEE Conference on Decision and Control (CDC)*, pages 7479–7486. IEEE, 2010.
- [15] K.I. Tsianos, S. Lawlor, and M.G. Rabbat. Consensus-based distributed optimization: Practical issues and applications in large-scale machine learning. In *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1543–1550. IEEE, October 2012.
- [16] V. Willert and M. Schnaubelt. Data control: A feedback control design for clustering of networked data (in german). *Automatisierungstechnik*, 64(8):618–632, 2016.
- [17] C. Zhao, J. He, P. Cheng, and J. Chen. Consensus-based energy management in smart grid with transmission losses and directed communication. *IEEE Transactions on Smart Grid*, 8(5):2049–2061, Sep. 2017.

APPENDIX

Proof of Proposition 2

Proof. Consider the Lagrangian of problem (19), namely

$$\begin{aligned} L(\mathbf{p}, \mathbf{v}, \lambda, \boldsymbol{\mu}, \boldsymbol{\gamma}, \boldsymbol{\theta}) &= \sum_{i \in \mathcal{N}_g} C_i(p_i) - \sum_{j \in \mathcal{N}_d} U_j(p_j) \\ &+ \lambda \left(\sum_{i \in \mathcal{N}_g} (p_i - v_i) - \sum_{j \in \mathcal{N}_d} p_j \right) + \sum_{i \in \mathcal{N}_g} \mu_i (l_i p_i^2 - v_i) \\ &+ \sum_{k \in [N]} \gamma_k (p_k^m - p_k) + \sum_{k \in [N]} \theta_k (p_k - P_k^M), \end{aligned}$$

where $\boldsymbol{\mu}$, $\boldsymbol{\gamma}$, $\boldsymbol{\theta}$ are the vectors of the Lagrangian multipliers with corresponding dimensions. As the problem (19) is convex and Assumption 6 holds, we can use necessary and sufficient Karush-Kuhn-Tucker conditions for $[(\mathbf{p}^*, \mathbf{v}^*), (\lambda^*, \boldsymbol{\mu}^*, \boldsymbol{\gamma}^*, \boldsymbol{\theta}^*)]$ being an optimal primal dual pair⁵. Thus,

$$\frac{\partial L}{\partial p_i} = \frac{dC_i(p_i^*)}{dp_i} + \lambda^* + 2\mu_i^* l_i p_i^* - \gamma_i^* + \theta_i^* = 0, \quad i \in \mathcal{N}_g, \quad (22a)$$

$$\frac{\partial L}{\partial p_j} = -\frac{dU_j(p_j^*)}{dp_j} - \lambda^* - \gamma_j^* + \theta_j^* = 0, \quad j \in \mathcal{N}_d, \quad (22b)$$

$$\frac{\partial L}{\partial v_i} = -\lambda^* + \mu_i^* = 0, \quad i \in \mathcal{N}_g, \quad (22c)$$

⁵Note that existence of an optimal primal dual is guaranteed in this case as well (see, for example, Proposition 5.3.1 in [1]).

with

$$\gamma_k^* (p_k^m - p_k^*) = 0, \quad \gamma_k^* \geq 0, \quad k \in [N], \quad (23a)$$

$$\theta_k^* (p_k^* - P_k^M) = 0, \quad \theta_k^* \geq 0, \quad k \in [N], \quad (23b)$$

$$\mu_i^* (l_i (p_i^*)^2 - v_i^*) = 0, \quad \mu_i^* \geq 0, \quad i \in \mathcal{N}_g. \quad (23c)$$

Suppose that

$$v_{i'}^* > l_{i'} (p_{i'}^*)^2 \quad \text{for some } i' \in \mathcal{N}_g. \quad (24)$$

Then, due to (23c), $\mu_{i'}^* = 0$. Hence, according to (22c), $\lambda^* = 0$ and $\mu_i^* = 0$ for all $i \in \mathcal{N}_g$. Next, since Assumption 6 guarantees that $p_k^m \neq P_k^M$ for all $k \in [N]$, $\gamma_k^* = 0$ or $\theta_k^* = 0$ for all $k \in [N]$ (see (23a) and (23b)). Let us consider any $j \in \mathcal{N}_d$. Due to the fact that $\frac{dU_j(p_j^*)}{dp_j} > 0$ (see (16)), $\gamma_j^* \geq 0$ (see (23a)), and condition (22b), we conclude that $\gamma_j^* = 0$ and $\theta_j^* > 0$. Thus, $p_j^* = P_j^M$ (see (23b)) for all $j \in \mathcal{N}_d$. Analogously, for any $i \in \mathcal{N}_g$, due to the property of the cost functions, namely $\frac{dC_i(p_i^*)}{dp_i} > 0$ (see (15)), we obtain that $\gamma_i^* > 0$ and $\theta_i^* = 0$ (see (22a)). Thus, $p_i^* = p_i^m$ (see (23a)) for all $i \in \mathcal{N}_g$.

Next, taking into account (24) and the feasibility conditions (19a) and (19d), we get

$$\begin{aligned} \sum_{j \in \mathcal{N}_d} P_j^M &= \sum_{j \in \mathcal{N}_d} p_j^* = \sum_{i \in \mathcal{N}_g} (p_i^* - v_i^*) \\ &< \sum_{i \in \mathcal{N}_g} (p_i^m - l_i (p_i^m)^2), \end{aligned} \quad (25)$$

which contradicts Assumption 5. Thus, (24) cannot hold, which implies that $v_i^* = l_i (p_i^*)^2$ for all $i \in \mathcal{N}_g$. Hence, the optimal solution $(\mathbf{p}^*, \mathbf{v}^*)$ to the problem (19) necessarily satisfies the feasibility conditions of the problem (18), which is equivalent to the initial problem (17). By noticing that the objective function in the optimization problems is the same, we conclude the proof. \square

Supporting Theorems

Theorem 2. [4] Consider the sequences $\{\mathbf{z}_i(t)\}_t$, $i \in [n]$, generated by the algorithm (4). Assume that the graph sequence $\{G(t)\}$ is B -strongly connected and Assumptions 1 and 3 hold.

(a) Then $\lim_{t \rightarrow \infty} \|\mathbf{z}_i(t+1) - \bar{\mathbf{x}}(t)\| = 0$ for all $i \in [n]$.

Moreover,

(b) If $\{b_t\}$ is a non-increasing positive scalar sequence with $\sum_{t=1}^{\infty} b_t a_t \|\mathbf{f}_i(\mathbf{z}_i(t+1)) + r_t \boldsymbol{\psi}_i(\mathbf{z}_i(t+1))\|_1 < \infty$ for all $i \in [n]$, then $\sum_{t=0}^{\infty} b_t \|\mathbf{z}_i(t+1) - \bar{\mathbf{x}}(t)\| < \infty$ for all i , where $\|\cdot\|_1$ is the l^1 -norm in \mathbb{R}^d .

The next theorem is the well-known result on non-negative variables [12].

Theorem 3. Let z_n, β_n, ξ_n , and ζ_n be non-negative variables such that

$$z_{n+1} \leq z_n(1 + \beta_n) - \zeta_n + \xi_n.$$

Then $\lim_{n \rightarrow \infty} z_n$ exists and is finite and $\sum_{n=1}^{\infty} \zeta_n < \infty$ on $\{\sum_{n=1}^{\infty} \beta_n < \infty, \sum_{n=1}^{\infty} \xi_n < \infty\}$.