



2017/13

DP

Yurii Nesterov and Vladimir Shikhman

Dual subgradient method
with averaging for optimal
resource allocation



CORE

Voie du Roman Pays 34, L1.03.01

Tel (32 10) 47 43 04

Fax (32 10) 47 43 01

Email: immaq-library@uclouvain.be

[https://uclouvain.be/en/research-institutes/
immaq/core/discussion-papers.html](https://uclouvain.be/en/research-institutes/immaq/core/discussion-papers.html)

CORE DISCUSSION PAPER

2017/13

Dual subgradient method with averaging for optimal resource allocation

Yu. Nesterov ^{*} V. Shikhman [†]

April 21, 2017

Abstract

A dual subgradient method is proposed for solving convex optimization problems with linear constraints. As novelty, the recovering of primal solutions can be avoided. Instead, the optimal convergence rate for the whole sequence of primal-dual iterates is obtained. This is due to the primal-dual averaging strategies which are incorporated into the iterative scheme. We apply our dual subgradient method with averaging to optimal resource allocation within a multi-agent environment. The proposed dual subgradient method naturally corresponds to a distributed process of production/price adjustments and effectively leads to a market equilibrium.

Keywords: convex programming, dual subgradient methods, rate of convergence, averaging, resource allocation

^{*}Center for Operations Research and Econometrics (CORE), Catholic University of Louvain (UCL), 34 voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium; e-mail: yurii.nesterov@uclouvain.be.

[†]Center for Operations Research and Econometrics (CORE), Catholic University of Louvain (UCL), 34 voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium; e-mail: vladimir.shikhman@uclouvain.be.

The research of the first author has been supported by a grant “Action de recherche concertée ARC 04/09-315” from the “Direction de la recherche scientifique - Communauté française de Belgique”. The research of the second author has been supported by a grant “Chargé de recherches 98444” from the “Fonds de la Recherche Scientifique - FNRS, Belgique”. The scientific responsibility rests with its authors.

1 Introduction

Dual subgradient methods are used for multi-agent convex optimization subject to shared constraints. For the latter primal problem a corresponding dual is stated by the Lagrangian relaxation of shared constraints. It is crucial that subgradients of the dual objective function represent primal feasibility. This fact motivates to apply subgradient-type methods from convex optimization in order to solve the dual problem. In the framework of dual subgradient methods it is typical to assume that given a current dual iterate, agents solve their individual optimization problems and form primal iterates. Then, dual variables are updated accordingly to the aggregated primal feasibility. By doing so, a sequence of primal-dual iterates emerges. Dual subgradient methods differ in how primal iterates are formed and in how dual variables are then updated.

The main goal of this paper is to present a dual subgradient method enjoying the optimal convergence rate for the whole primal-dual sequence. Note that in the literature convergence rates are provided for the running averages rather than for the whole primal-dual sequence. This issue is usually referred to as the recovering of primal solutions by averaging. The difficulty of primal recovering is due to the typical nonsmoothness of the dual objective function at an optimum. Thus, even knowing the optimal dual does not necessarily lead to an optimal primal solution. Averaging schemes for recovering of primal-dual solutions have been extensively studied, see e.g. [7, 10, 11, 12, 18, 19].

There are two main drawbacks for the recovering of primal-dual solutions. First, the convergence rates for averages of primal-dual iterates do not prevent uncontrollable jumps of function values at some particular iterations. The latter may cause oscillation of the multi-agent system during the adjustment process. In fact, the dual subgradient method can be viewed as an adjustment strategy for approaching a stable state of such a system. The dual variables play the role of current control parameters, such as e.g. prices, and the subgradient of the dual objective function represent observed reactions of the system, such as e.g. excess supply. It is important to implement the variants of the control, which asymptotically stabilize the system. Hence, it is not reasonable to accumulate knowledge about potentially good controls and system reactions, which will be never implemented in practice. This is the second obstacle for the recovering of primal-dual solutions. The convergence of historical averages of primal-dual iterates provides statistical rather than actual information about the dynamics of a system. To achieve the latter, we need to guarantee the convergence rate for the primal-dual sequence itself. We mention those papers where the convergence for the whole primal-dual sequence has been under consideration. In [17], a primal-dual sequence is generated by combining the usual subgradient method with an auxiliary penalty problem. In order to determine step sizes, the differences of functional values for both the dual and auxiliary penalty problems are used. This enforces the convergence of primal-dual iterates at the geometric rate. However, the applicability of the method from [17] in a multi-agent environment is hampered, since the functional values are due to the private information. In [2], a variant of the fast gradient method is applied to generate the primal-dual sequence. Assuming strong concavity of agents' objective functions, $O\left(\frac{1}{T}\right)$ rate of convergence for the primal-dual sequence is established. In [16], primal-dual iterates are formed by averaging strategies using an idea from heavy ball methods. Convergence of the whole primal-dual sequence is shown, but its rate is missing. In this paper, we turn our attention to the generation of a primal-dual sequence

by the so-called quasi-monotone subgradient methods from [14]. Our primal-dual iterates are also formed by averaging, but the construction differs from that in [16]. Based on a current dual iterate, agents solve their individual optimization problems and form primal iterates by averaging. Then, a dual forecast is set to be proportional to the feasibility of these average primal iterates. The factors of proportionality play the role of step sizes in the scheme. Finally, the next dual iterate averages dual forecasts over time. For the sequence of primal-dual iterates generated by our dual subgradient method with averaging we provide the convergence rate of the order $O\left(\frac{1}{\sqrt{t+1}}\right)$. Note that the latter convergence rate is optimal for convex nonsmooth optimization, and we do not assume strong concavity as in [2].

Additionally, we apply our dual subgradient method with averaging to optimal resource allocation. It is worth to mention that the range of applications for dual subgradient methods is rather wide. It includes network resource allocation problem [2], multicommodity network flows [9], traffic equilibrium and road pricing [10]. The problem of optimal resource allocation is to maximize producers' aggregated profits by sharing available resources. Popularized and advocated mainly in the monograph [1], the mechanisms of decentralized resource allocation gained a lot of attention in economics and operations research since then, see e.g. [3, 6, 8] and references therein. Resource allocation can be seen as a generalized Nash equilibrium problem with common constraints, cf. recent surveys [4, 5]. In fact, each producer seeks to maximize the own profit while the available resources are shared. We assume that these resource constraints are linear and separable w.r.t. the producers. In this economic context, primal variables are production bundles and dual variables represent prices of resources. Our dual subgradient method naturally corresponds to a distributed process of production/price adjustments. It effectively leads to the clearance of the resource market, i.e. the situation when supply meets demand at equilibrium prices. We emphasize that by our adjustment process the resource market clears in the limit as such, and not only on average, as it is typical for the application of standard subgradient schemes. In economic terms, our results explain how resource markets may dynamically stabilize by decentralized adjustment rather than oscillate.

The paper is organized as follows. Section 2 is devoted to the dual subgradient method with averaging and its convergence properties. In Section 3 optimal resource allocation problem is solved. Proofs of auxiliary statements are given in Appendix.

Our notation is quite standard. We denote by \mathbb{R}^n the space of n -dimensional column vectors $x = (x^{(1)}, \dots, x^{(n)})^T$, and by \mathbb{R}_+^n (\mathbb{R}_{++}^n) the set of all vectors with nonnegative (positive) components, respectively. For x and y from \mathbb{R}^n , we introduce the standard scalar product and the Hadamard product

$$\langle x, y \rangle = \sum_{j=1}^n x^{(j)} y^{(j)}, \quad x \circ y = \left(x^{(j)} y^{(j)} \right)_{j=1}^n \in \mathbb{R}^n.$$

The positive part of the real value $a \in \mathbb{R}$ is denoted by $(a)_+ = \max\{a, 0\}$. For $x = (x^{(1)}, \dots, x^{(n)})^T \in \mathbb{R}^n$ we denote

$$\begin{aligned} x_+ &= \left((x^{(1)})_+, \dots, (x^{(n)})_+ \right)^T, \quad x^2 = \left((x^{(1)})^2, \dots, (x^{(n)})^2 \right)^T, \\ \frac{1}{x} &= \left(\frac{1}{x^{(1)}}, \dots, \frac{1}{x^{(n)}} \right)^T \quad \text{if } x^{(j)} \neq 0, \quad j = 1, \dots, n. \end{aligned}$$

Given $d \in \mathbb{R}_{++}^m$, we use the following norms for $v = (v^{(1)}, \dots, v^{(m)})^T \in \mathbb{R}^m$:

$$\|v\|_{2,d} = \sqrt{\sum_{j=1}^n d^{(j)} (v^{(j)})^2}, \quad \|v\|_{\infty,d} = \max_{j=1,\dots,n} d^{(j)} |v^{(j)}|.$$

We denote by $\|x\|$ any, but fixed norm of $x = (x^{(1)}, \dots, x^{(n)})^T \in \mathbb{R}^n$. For an $m \times n$ matrix $A \in \mathbb{R}^{m \times n}$ we use the norm

$$\|A\|_{2,d} = \max_{\|x\| \leq 1} \|Ax\|_{2,d}.$$

For vectors $p_1, \dots, p_I \in \mathbb{R}^m$, denote by $\max_{i=1,\dots,I} p_i \in \mathbb{R}^m$ the vector with coordinates

$$\left(\max_{i=1,\dots,I} p_i \right)^{(k)} = \max_{i=1,\dots,I} p_i^{(k)}, \quad k = 1, \dots, m.$$

2 Dual subgradient method with averaging

We consider the following optimization problem:

$$f^* \stackrel{\text{def}}{=} \max_{x \in X} f(x) \quad \text{s.t.} \quad Ax \leq b, \quad (1)$$

where the feasible set $X \subset \mathbb{R}^n$ is convex and compact, the objective function f is concave on \mathbb{R}^n , $A \in \mathbb{R}^{m \times n}$ is an $m \times n$ real matrix, and $b \in \mathbb{R}^m$. Let X^* denote the solution set of the primal optimization problem (1). The dual objective function is

$$\Psi(p) \stackrel{\text{def}}{=} \max_{x \in X} [f(x) - \langle p, Ax \rangle] + \langle p, b \rangle, \quad (2)$$

where $p \in \mathbb{R}^m$ is the vector of dual (or Lagrange) multipliers for the linear inequality constraints in (1). In applications the maximization problem (2) is usually solvable within the distributed framework, see Section 3. Note that Ψ is convex as the maximum of linear functions w.r.t. p . Let $X^*(p)$ denote the solution set of the optimization problem (2) for given $p \in \mathbb{R}^m$. Note that $X^*(p)$ is convex and compact.

The dual optimization problem for (1) reads as

$$\Psi^* \stackrel{\text{def}}{=} \min_{p \in \mathbb{R}_+^m} \Psi(p). \quad (3)$$

Its solution set is denoted by P^* . In order to relate primal and dual optimization problems (1) and (3), we assume Slater constraint qualification throughout this Section 2.

Assumption 1 (Slater constraint qualification) *There exists $\bar{x} \in X$ with $A\bar{x} < b$.*

Under Slater constraint qualification, the dual optimization problem (3) is solvable, its solution set P^* is nonempty and compact, e.g., [15]. Moreover, the strong duality of convex optimization provides $f^* = \Psi^*$. Hence, any primal-dual solution $(x^*, p^*) \in X^* \times P^*$ fulfills $f(x^*) = \Psi(p^*)$. Together with the weak duality of convex optimization, the reverse

implication is also true. Any feasible pair $(x^*, p^*) \in X \times \mathbb{R}_+^n$ with $Ax^* \leq b$ fulfilling $f(x^*) = \Psi(p^*)$ solves (1) and (3), respectively, e.g. [15].

Our aim is to approximate primal-dual solutions of (1) and (3) by solving the dual optimization problem (3) via quasi-monotone subgradient method from [14]. For that, let us recall the formula for the convex subdifferential of Ψ , e.g. [15]:

$$\partial\Psi(p) = \{b - Ax(p) \mid x(p) \in X^*(p)\}, \quad (4)$$

where $X^*(p)$ is the solution set of the optimization problem (2). In the framework of dual subgradient methods it is typical to assume that the optimization problem (2) can be solved. Throughout this section, we use

Assumption 2 (Primal availability) *For all $p \in \mathbb{R}^n$ at least one solution of the optimization problem (2) is available:*

$$x(p) \in X^*(p) = \arg \max_{x \in X} [f(x) - \langle p, Ax \rangle].$$

For a sequence of parameters $\gamma[t] \in \mathbb{R}_+^m$, $t \geq 0$, we form primal-dual pairs $(x[t], p[t])$ by the following iteration:

Dual Subgradient Method with Averaging	
<p>1. Take $x(p[t]) \in X^*(p[t])$ and form primal average</p> $x[t] \stackrel{\text{def}}{=} \frac{1}{t+1} \sum_{r=0}^t x(p[r]) = \frac{t}{t+1} x[t-1] + \frac{1}{t+1} x(p[t]).$ <p>2. Forecast $p^+[t+1] \stackrel{\text{def}}{=} \frac{1}{\Gamma[t]} \circ [Ax[t] - b]_+$</p> <p style="padding-left: 20px;">by using average parameter $\Gamma[t] \stackrel{\text{def}}{=} \frac{1}{t+1} \sum_{r=0}^t \gamma[r] \in \mathbb{R}^m.$</p> <p>3. Form dual average</p> $p[t+1] \stackrel{\text{def}}{=} \frac{1}{t+2} \sum_{r=0}^{t+1} p^+[r] = \frac{t+1}{t+2} p[t] + \frac{1}{t+2} p^+[t+1],$ <p style="padding-left: 20px;">where $p[0] = p^+[0] = 0.$</p>	(DSMA)

In what follows we show convergence of $(x[t], p[t])$ from DSMA toward primal-dual solutions of (1) and (3). First, Lemmata 1 and 2 provide estimations of the following expression from above and below:

$$\Psi(p[t]) - f(x[t]) + \left\langle \frac{1}{\Gamma[t]}, \frac{1}{2} (Ax[t] - b)_+^2 \right\rangle. \quad (5)$$

In (5) the term $\Psi(p[t]) - f(x[t])$ stands for the primal-dual gap, and $\frac{1}{2} (Ax[t] - b)_+^2$ represents the quadratic penalty for the linear constraints in (1).

For given $d \in \mathbb{R}_{++}^n$, let us set the following control sequence of parameters:

$$\Delta[t] \stackrel{\text{def}}{=} \frac{1}{t+1} \sum_{r=0}^t \left\| \frac{1}{\sum_{s=0}^{r-1} \gamma[s]} \right\|_{\infty, \frac{1}{d}}, \text{ where } \sum_{s=0}^{-1} \gamma[s] \stackrel{\text{def}}{=} \gamma[0]. \quad (6)$$

Lemma 1 (Estimation from above) *Let $(x[t], p[t])$ be generated by DSMA with parameters $\gamma[t]$. Then, for all $t \geq 0$ it holds:*

$$\Psi(p[t]) - f(x[t]) + \left\langle \frac{1}{\Gamma[t]}, \frac{1}{2} (Ax[t] - b)_+^2 \right\rangle \leq C_1 \cdot \Delta[t],$$

where

$$C_1 \stackrel{\text{def}}{=} \frac{1}{2} \left(\|A\|_{2,d} \max_{x \in X} \|x\| + \|b\|_{2,d} \right)^2.$$

The proof of Lemma 1 is based on the estimate sequence technique from [13]. For readers' convenience it is postponed to Appendix, as well the proof of next Lemma 2.

Lemma 2 (Estimation from below) *Let $(x[t], p[t])$ be generated by DSMA with parameters $\gamma[t]$. Then, for all $t \geq 0$ it holds:*

$$\Psi(p[t]) - f(x[t]) + \left\langle \frac{1}{\Gamma[t]}, \frac{1}{2} (Ax[t] - b)_+^2 \right\rangle \geq \Psi(p[t]) - \Psi^* - C_2 \cdot \|\Gamma[t]\|_{\infty, d},$$

where

$$C_2 \stackrel{\text{def}}{=} \frac{1}{2} \left(\min_{p^* \in P^*} \|p^*\|_{2, \frac{1}{d}} \right)^2.$$

Now, let us specify parameters $\gamma[t] \in \mathbb{R}_+^m$, $t \geq 0$, in Step 2. of DSMA.

Definition 1 (Admissible parameters) *The sequence of parameters $\gamma[t] \in \mathbb{R}_+^m$, $t \geq 0$, is called admissible if it satisfies:*

- (i) $\gamma[t] \rightarrow 0$,
- (ii) $\sum_{r=0}^t \gamma[r] \rightarrow \infty$.

Lemma 3 examines the behavior of the control sequences $\Gamma[t]$ and $\Delta[t]$.

Lemma 3 (Control sequences) *Let the sequence of parameters $\gamma[t] \in \mathbb{R}_+^m$, $t \geq 0$, be admissible. Then, $\Gamma[t] \rightarrow 0$ and $\Delta[t] \rightarrow 0$. Moreover, the best achievable order of convergence is $O\left(\frac{1}{\sqrt{t+1}}\right)$.*

Proof:

Definition 1(i) requires that $\gamma[t] \rightarrow 0$. Hence, by averaging $\Gamma[t] = \frac{1}{t+1} \sum_{r=0}^t \gamma[r] \rightarrow 0$.

From Definition 1(ii) we have $\sum_{s=0}^{r-1} \gamma[s] \rightarrow \infty$. Thus, $\left\| \frac{1}{\sum_{s=0}^{r-1} \gamma[s]} \right\|_{\infty, \frac{1}{d}} \rightarrow 0$, and again by

averaging $\Delta[t] = \frac{1}{t+1} \sum_{r=0}^t \left\| \frac{1}{\sum_{s=0}^{r-1} \gamma[s]} \right\|_{\infty, \frac{1}{d}} \rightarrow 0$. In order to achieve the best order of

convergence, we take admissible parameters $\gamma[t] = \frac{1}{\sqrt{t+1}}, t \geq 0$. Note that (i)-(ii) from Definition 1 are satisfied for this particular choice. Then,

$$\sum_{r=0}^t \gamma[r] = \sum_{r=0}^t \frac{1}{\sqrt{r+1}} \leq \int_{r=-\frac{1}{2}}^{t+\frac{1}{2}} \frac{1}{\sqrt{r+1}} dr = 2 \left(\sqrt{t+\frac{3}{2}} - \sqrt{\frac{1}{2}} \right) = O(\sqrt{t+1}).$$

Hence,

$$\Gamma[t] = \frac{1}{t+1} \sum_{r=0}^t \gamma[r] = \frac{1}{t+1} O(\sqrt{t+1}) = O\left(\frac{1}{\sqrt{t+1}}\right),$$

$$\Delta[t] = \frac{1}{t+1} \sum_{r=0}^t \left\| \frac{1}{\sum_{s=0}^{r-1} \gamma[s]} \right\|_{\infty, \frac{1}{d}} = \frac{1}{t+1} \sum_{r=0}^t \frac{1}{O(\sqrt{r+1})}$$

$$= \frac{1}{t+1} O\left(\int_{r=0}^t \frac{1}{\sqrt{r+1}} dr\right) = O\left(\frac{1}{\sqrt{t+1}}\right).$$

The order of convergence for both $\Gamma[t] \rightarrow 0$ and $\Delta[t] \rightarrow 0$ is $O\left(\frac{1}{\sqrt{t+1}}\right)$. □

Let us state the main convergence result for DSMA.

Theorem 1 (Convergence of DSMA) *Let admissible parameters $\gamma[t]$ be applied in DSMA. Then, the sequence $(x[t], p[t])$ generated by DSMA converges toward the set of primal-dual solutions of (1) and (3). Moreover, the rate of convergence $O\left(\frac{1}{\sqrt{t+1}}\right)$ follows from the valid inequality for all $t \geq 0$:*

$$\Psi(p[t]) - \Psi^* - C_2 \|\Gamma[t]\|_{\infty, d} \leq \Psi(p[t]) - f(x[t]) + \left\langle \frac{1}{\Gamma[t]}, \frac{1}{2} (Ax[t] - b)_+^2 \right\rangle \leq C_1 \Delta[t]. \quad (7)$$

Proof:

Due to the compactness of the feasible set X and the dual optimal solution set P^* , the following expressions

$$C_1 = \frac{1}{2} \left(\|A\|_{2,d} \max_{x \in X} \|x\| + \|b\|_{2,d} \right)^2, \quad C_2 = \frac{1}{2} \left(\min_{p^* \in P^*} \|p^*\|_{2, \frac{1}{d}} \right)^2$$

are finite constants. Lemmata 1 and 2 provide

$$\Psi(p[t]) - \Psi^* - C_2 \|\Gamma[t]\|_{\infty, d} \leq \Psi(p[t]) - f(x[t]) + \left\langle \frac{1}{\Gamma[t]}, \frac{1}{2} (Ax[t] - b)_+^2 \right\rangle \leq C_1 \Delta[t].$$

This inequality is composed by the primal and dual objective functions $f(x[t])$, $\Psi(p[t])$, computed at the current iterates, and the quadratic penalty $\frac{1}{2}(Ax[t] - b)_+^2$ for violation of the linear constraints in (1). Moreover, since the parameters $\gamma[t]$, $t \geq 0$, are chosen admissible, Lemma 3 gives

$$\Gamma[t] \rightarrow 0, \quad \Delta[t] \rightarrow 0.$$

Overall, $x[t]$ converges to the solution set of (1), and $p[t]$ converges to the solution set of (2) by order $O\left(\frac{1}{\sqrt{t+1}}\right)$. \square

3 Optimal resource allocation

We apply the dual subgradient method with averaging from Section 2 for optimal resource allocation. In particular, problems of optimal management and optimal distribution of resources are solved. The problem of optimal resource management is to maximize producers' aggregated profits by sharing scarce resources. Here, a manager updates prices of resources, which play the role of dual variables, in order to maintain the aggregated availability of scarce resources. The problem of optimal resource distribution is to maximize producers' aggregated profits minus distribution cost of scarce resources. Here, producers update their own prices of resources, which also play the role of dual variables, in order to maintain the individual availability of scarce resources. In both cases, we show that DSMA optimally manages/distributes resources by properly adjusting their prices.

3.1 Resource management

We consider the following resource management problem:

$$\max_{\substack{x_i \in X_i \\ i=1, \dots, I}} \sum_{i=1}^I f_i(x_i) \quad \text{s.t.} \quad \sum_{i=1}^I A_i x_i \leq b. \quad (8)$$

Here, $x_i \in \mathbb{R}_+^n$ represents a production bundle of goods $1, \dots, n$. The i -th producer's feasible set $X_i \subset \mathbb{R}_+^n$ is convex and compact, the profit function f_i is concave on \mathbb{R}^n . $A_i \in \mathbb{R}_+^{m \times n}$ is an $m \times n$ transformation matrix whose j -th column indicates the amounts of resources $1, \dots, m$ needed to produce one entity of good j by i -th producer. Available scarce resources $b \in \mathbb{R}_+^m$ are shared.

The dual objective function represents the total revenue

$$\Psi(p) \stackrel{\text{def}}{=} \sum_{i=1}^I \max_{x_i \in X_i} [f_i(x_i) - \langle p, A_i x_i \rangle] + \langle p, b \rangle, \quad (9)$$

where $p \in \mathbb{R}^m$ is the vector of prices for resources. In fact, the term $\langle p, A_i x_i \rangle$ is the cost of producing x_i to resource prices p . Hence, $f_i(x_i) - \langle p, A_i x_i \rangle$ can be viewed as i -th producer's revenue, and $\langle p, b \rangle$ is the revenue of resource holders. The dual optimization problem for (8) minimizes the total revenue (9):

$$\min_{p \in \mathbb{R}_+^m} \Psi(p). \quad (10)$$

The Slater constraint qualification for (8) plays the role of the productivity assumption. Namely, we assume that there exist feasible production bundles $\bar{x}_i \in X_i$, $i = 1, \dots, I$, such that the supply of resources strictly exceeds their aggregated demand, i.e. $\sum_{i=1}^I A_i \bar{x}_i < b$.

Let us interpret primal-dual solutions of (8) and (10) in economic terms. Due to the duality of convex optimization, primal-dual solutions of (8) and (10) coincide with the following market equilibria $\left((x_i)_{i=1}^I, p \right)$:

$$\left. \begin{array}{l} (i) \text{ producers maximize revenues, i.e.} \\ x_i \in \arg \max_{x_i \in X_i} [f_i(x_i) - \langle p, A_i x_i \rangle], i = 1, \dots, I, \\ (ii) \text{ market of resources is cleared, i.e.} \\ p \geq 0, \quad \sum_{i=1}^I A_i x_i \leq b, \quad \left\langle p, b - \sum_{i=1}^I A_i x_i \right\rangle = 0. \end{array} \right\} \quad (11)$$

Note that the availability of a maximum revenue bundle in (i) of (11) is posed in Assumption 2. It says that given prices of resources p , every producer is able to solve the own revenue maximization problem, i.e. to find

$$x_i(p) \in X_i^*(p) \stackrel{\text{def}}{=} \arg \max_{x_i \in X_i} [f_i(x_i) - \langle p, A_i x_i \rangle].$$

Further, (ii) in (11) is the necessary optimality condition for the dual optimization problem (10). It means that at equilibrium prices the aggregated use of resources never exceeds its available amount, and the resources with positive prices are completely exhausted.

Given any price p , the excess supply of resources comes from the convex subdifferential

$$\partial\Psi(p) = \left\{ b - \sum_{i=1}^I A_i x_i(p) \mid x_i(p) \in X_i^*(p), i = 1, \dots, I \right\}. \quad (12)$$

The latter suggests to use dual subgradient methods for price adjustment. Specifically, we apply DSMA for solving the resource management problem (8) and its dual problem (10). By doing so, we assume that a manager has access to the aggregated excess demand of resources $\sum_{i=1}^I A_i x_i - b$, where x_i is the realized production of i -th producer. Then, the manager updates prices accordingly to the excess demand of resources. Namely, let the manager have a sequence of parameters $\gamma[t] \in \mathbb{R}_+^m$, $t \geq 0$. Primal-dual pairs $\left((x_i[t])_{i=1}^I, p[t] \right)$ of production bundles and prices are formed by the following iteration from DSMA:

Resource Management with Averaging

1. i -th producer maximizes revenue by $x_i(p[t]) \in X_i^*(p[t])$
and updates production statistics $x_i[t] \stackrel{\text{def}}{=} \frac{1}{t+1} \sum_{r=0}^t x_i(p[r])$.
2. Manager observes current excess demand $\sum_{i=1}^I A_i x_i[t] - b$
and forms price forecast $p^+[t+1] \stackrel{\text{def}}{=} \frac{1}{\Gamma[t]} \circ \left[\sum_{i=1}^I A_i x_i[t] - b \right]_+$
by using average parameter $\Gamma[t] \stackrel{\text{def}}{=} \frac{1}{t+1} \sum_{r=0}^t \gamma[r] \in \mathbb{R}^m$.
3. Manager forms new prices by averaging forecasts
 $p[t+1] \stackrel{\text{def}}{=} \frac{1}{t+2} \sum_{r=0}^{t+1} p^+[r]$, where $p[0] = p^+[0] = 0$.

(MAN)

Since MAN is a variant of DSMA, its convergence follows directly from Theorem 1.

Theorem 2 (Convergence of Resource Management) *Let admissible parameters $\gamma[t]$ be applied in MAN. Then, the sequence $\left((x_i[t])_i^I, p[t] \right)$ of production bundles and prices generated by MAN converges toward the set of primal-dual solutions of (8) and (10), or equivalently, toward the set of market equilibria (11). Moreover, the achievable rate of convergence is of the order $O\left(\frac{1}{\sqrt{t+1}}\right)$.*

3.2 Resource distribution

Using notations from Section 3.2, we consider the following resource distribution problem:

$$\max_{\substack{(x_i)_i^I \in (X_i)_i^I \\ y \in \mathcal{Y}}} \sum_{i=1}^I f_i(x_i) - c(y) \quad \text{s.t.} \quad A_i x_i \leq y_i, \quad i = 1, \dots, I, \quad (13)$$

where

$$\mathcal{Y} \stackrel{\text{def}}{=} \left\{ y = (y_i)_{i=1}^I \in Y \mid \sum_{i=1}^I y_i \leq b \right\}$$

denotes the set of feasible distributions. Here, $y = (y_i)_{i=1}^I \in \mathbb{R}_+^{mn}$ represents a resource distribution between producers $1, \dots, I$. The set $Y \subset \mathbb{R}_+^{mn}$ is convex and closed. The condition $\sum_{i=1}^I y_i \leq b$ guarantees that the distributed resources do not exceed their available amounts. The latter also implies that the feasible set \mathcal{Y} is compact. The distribution

cost function $c(y)$ is assumed to be convex. Overall, the resource distribution problem (13) maximizes producers' aggregated profits minus the distribution cost of some available scarce resources. The linear constraint $A_i x_i \leq y_i$ in (13) means that the resources $A_i x_i$ needed for producing x_i do not exceed those resource shares y_i assigned to the i -th producer.

The dual objective function represents the total revenue

$$\Psi(p_1, \dots, p_I) \stackrel{\text{def}}{=} \sum_{i=1}^I \max_{x_i \in X_i} [f_i(x_i) - \langle p_i, A_i x_i \rangle] + \max_{y \in Y} \left[\sum_{i=1}^I \langle p_i, y_i \rangle - c(y) \right], \quad (14)$$

where $\mathbf{p} = (p_i)_{i=1}^I \in \mathbb{R}^{nm}$ is the vector of producers' prices for resources. In fact, the term $\langle p_i, A_i x_i \rangle$ is the cost of producing x_i to i -th producer's resource prices p_i . Hence, $f_i(x_i) - \langle p_i, A_i x_i \rangle$ can be viewed as i -th producer's revenue. The term $\sum_{i=1}^I \langle p_i, y_i \rangle$ represents the profit of resource holders by disposing resources y_i priced by p_i to the i -th producer. Hence, $\sum_{i=1}^I \langle p_i, y_i \rangle - c(y)$ is the revenue of resource holders caused by resource distribution $y = (y_i)_{i=1}^I$. The dual optimization problem for (13) minimizes the total revenue (14):

$$\min_{\substack{p_i \in \mathbb{R}_+^n \\ i=1, \dots, I}} \Psi(p_1, \dots, p_I). \quad (15)$$

The Slater constraint qualification for (13) plays the role of the productivity assumption. Namely, we assume that there exist feasible production bundles $\bar{x}_i \in X_i$, $i = 1, \dots, I$, and a feasible distribution $\bar{y} = (\bar{y}_i)_{i=1}^I \in Y$, such that the supply of resources to i -th producer strictly exceeds the corresponding demand, i.e. $A_i \bar{x}_i < \bar{y}_i$, $i = 1, \dots, I$.

Let us interpret primal-dual solutions of (13) and (15) in economic terms. Due to the duality of convex optimization, primal-dual solutions of (13) and (15) coincide with the following market equilibria $\left((x_i)_{i=1}^I, (y_i)_{i=1}^I, (p_i)_{i=1}^I \right)$:

$$\left. \begin{array}{l} (i) \text{ producers maximize revenues, i.e.} \\ \quad x_i \in \arg \max_{x_i \in X_i} [f_i(x_i) - \langle p_i, A_i x_i \rangle], i = 1, \dots, I, \\ (ii) \text{ distributor maximizes revenue, i.e.} \\ \quad y = (y_i)_{i=1}^I \in \arg \max_{y \in Y} \left[\sum_{i=1}^I \langle p_i, y_i \rangle - c(y) \right], \\ (iii) \text{ } i\text{-th market of resources is cleared, i.e.} \\ \quad p_i \geq 0, \quad A_i x_i \leq y_i, \quad \langle p_i, y_i - A_i x_i \rangle = 0, i = 1, \dots, I. \end{array} \right\} \quad (16)$$

Note that the solvability of revenue maximization in (i) and (ii) of (16) is posed in Assumption 2. It says that given individual prices of resources $\mathbf{p} = (p_i)_{i=1}^I$, every producer is able to find

$$x_i(p_i) \in X_i^*(p_i) \stackrel{\text{def}}{=} \arg \max_{x_i \in X_i} [f_i(x_i) - \langle p_i, A_i x_i \rangle],$$

and the distributor is able to find

$$y(\mathbf{p}) = (y_i(\mathbf{p}))_{i=1}^I \in Y^*(\mathbf{p}) \stackrel{\text{def}}{=} \arg \max_{y \in Y} \left[\sum_{i=1}^I \langle p_i, y_i \rangle - c(y) \right].$$

Further, (iii) in (16) is the necessary optimality condition for the dual optimization problem (15). It means that at equilibrium prices the individual use of resources by i -th producer never exceeds their distributed amounts assigned to i -th producer. Moreover, for the resources with positive equilibrium prices the latter balance equalizes.

Given any prices $\mathbf{p} = (p_i)_{i=1}^I$, the producers' individual excess supplies of resources come from the convex subdifferential

$$\partial\Psi(p_1, \dots, p_I) = \left\{ (y_i(\mathbf{p}) - A_i x_i(p_i))_{i=1}^I \mid x_i(p_i) \in X_i^*(p_i), (y_i(\mathbf{p}))_{i=1}^I \in Y^*(\mathbf{p}) \right\}. \quad (17)$$

The latter suggests to use dual subgradient methods for individual price adjustment. Specifically, we apply DSMA for solving the resource distribution problem (13) and its dual problem (15). By doing so, we assume that the i -th producer has access to the individual excess demand of resources $A_i x_i - y_i$, where x_i is the realized production of i -th producer, and y_i is the share of resources assigned to i -th producer by the distributor. In order to perform optimal distribution, the distributor is assumed to anticipate producer's individual prices. Then, producers update their prices accordingly to the individual excess demands of resources. Namely, let producers have a sequences of parameters $\gamma_i[t] \in \mathbb{R}_+^m$, $t \geq 0$, $i = 1, \dots, I$. Primal-dual pairs $\left((x_i[t])_{i=1}^I, (y_i[t])_{i=1}^I, (p_i[t])_{i=1}^I \right)$ of production bundles, distributions and prices are formed by the following iteration from DSMA:

Resource Distribution with Averaging	
<p>1. Distributor anticipates producers' prices $\mathbf{p}[t] \stackrel{\text{def}}{=} (p_i[t])_{i=1}^I$, maximizes revenue by $(y_i(\mathbf{p}[t]))_{i=1}^I \in Y^*(\mathbf{p}[t])$ and performs average distribution $y_i[t] \stackrel{\text{def}}{=} \frac{1}{t+1} \sum_{r=0}^t y_i(\mathbf{p}[r])$.</p> <p>2. i-th producer maximizes revenue by $x_i(p_i[t]) \in X_i^*(p_i[t])$ and updates production statistics $x_i[t] \stackrel{\text{def}}{=} \frac{1}{t+1} \sum_{r=0}^t x_i(p_i[r])$.</p> <p>3. i-th producer observes individual excess demand $A_i x_i[t] - y_i[t]$ and forms price forecast $p_i^+[t+1] \stackrel{\text{def}}{=} \frac{1}{\Gamma_i[t]} \circ [A_i x_i[t] - y_i[t]]_+$ by using average parameter $\Gamma_i[t] \stackrel{\text{def}}{=} \frac{1}{t+1} \sum_{r=0}^t \gamma_i[r] \in \mathbb{R}^m$.</p> <p>4. i-th producer forms new prices by averaging forecasts $p_i[t+1] \stackrel{\text{def}}{=} \frac{1}{t+2} \sum_{r=0}^{t+1} p_i^+[r]$, where $p_i[0] = p_i^+[0] = 0$.</p>	(DIS)

Since DIS is a variant of DSMA, its convergence follows directly from Theorem 1.

Theorem 3 (Convergence of Resource Distribution) *Let admissible parameters $\gamma_i[t]$ be applied by producers in DIS. Then, the sequence $\left((x_i[t])_{i=1}^I, (y_i[t])_{i=1}^I, (p_i[t])_{i=1}^I\right)$ of production bundles, distributions and prices generated by DIS converges toward the set of primal-dual solutions of (13) and (15), or equivalently, toward the set of market equilibria (16). Moreover, the achievable rate of convergence is of the order $O\left(\frac{1}{\sqrt{t+1}}\right)$.*

3.3 Resource Auction

In Section 3.3 we relate the problems of optimal resource management and distribution. For that, let us consider the case of zero distribution cost $c(y) \equiv 0$, and no distributional constraints $Y \equiv \mathbb{R}_+^{mn}$ in (13):

$$\max_{\substack{(x_i)_i^I \in (X_i)_i^I \\ \sum_{i=1}^I y_i \leq b, y_i \geq 0}} \sum_{i=1}^I f_i(x_i) \quad \text{s.t.} \quad A_i x_i \leq y_i, \quad i = 1, \dots, I. \quad (18)$$

The dual objective function for (18) reads

$$\begin{aligned} \psi(p_1, \dots, p_I) &\stackrel{\text{def}}{=} \sum_{i=1}^I \max_{x_i \in X_i} [f_i(x_i) - \langle p_i, A_i x_i \rangle] + \max_{\substack{\sum_{i=1}^I y_i \leq b \\ y_i \geq 0}} \sum_{i=1}^I \langle p_i, y_i \rangle \\ &= \sum_{i=1}^I \max_{x_i \in X_i} [f_i(x_i) - \langle p_i, A_i x_i \rangle] + \langle p_{max}, b \rangle, \end{aligned} \quad (19)$$

where $p_{max} \stackrel{\text{def}}{=} \max_{i=1, \dots, I} p_i$ is the vector of highest offer prices for resources. The dual optimization problem for (18) is

$$\min_{\substack{p_i \in \mathbb{R}_+^n \\ i = 1, \dots, I}} \psi(p_1, \dots, p_I). \quad (20)$$

Lemma 4 *Let $\left((x_i)_{i=1}^I, (y_i)_{i=1}^I, (p_i)_{i=1}^I\right)$ solve primal-dual resource distribution problems (18) and (20). Then, $\left((x_i)_{i=1}^I, p_{max}\right)$ solves primal-dual resource management problems (8) and (10).*

Proof:

From

$$\max_{(x_i)_i^I \in (X_i)_i^I} \left[\sum_{i=1}^I f_i(x_i) \mid \sum_{i=1}^I A_i x_i \leq b \right] = \max_{\substack{(x_i)_i^I \in (X_i)_i^I \\ \sum_{i=1}^I y_i \leq b, y_i \geq 0}} \left[\sum_{i=1}^I f_i(x_i) \mid \begin{array}{l} A_i x_i \leq y_i \\ i = 1, \dots, I \end{array} \right]$$

we immediately see that the x -parts of solutions for (8) and (18) coincide. Moreover, this implies for the corresponding dual problems (10) and (20):

$$\min_{p \in \mathbb{R}_+^n} \Psi(p) = \min_{\substack{p_i \in \mathbb{R}_+^n \\ i = 1, \dots, I}} \psi(p_1, \dots, p_I).$$

Assume that $(p_i)_{i=1}^I$ solves (20). Then,

$$\min_{\substack{p_i \in \mathbb{R}_+^n \\ i = 1, \dots, I}} \psi(p_1, \dots, p_I) = \psi(p_1, \dots, p_I) \stackrel{(19)}{\geq} \Psi(p_{max}) \geq \min_{p \in \mathbb{R}_+^n} \Psi(p).$$

Hence, p_{max} solves (10). □

In what follows we intend to apply DIS to the resource distribution problem (18). For that, we discuss Assumption 2 which states the solvability of (19). Let us compute with $\mathbf{p} = (p_i)_{i=1}^I$:

$$Y^*(\mathbf{p}) \stackrel{\text{def}}{=} \arg \max_{\sum_{i=1}^I y_i \leq b, y_i \geq 0} \sum_{i=1}^I \langle p_i, y_i \rangle = \left\{ (\lambda_i \circ b)_{i=1}^I \mid (\lambda_i)_{i=1}^I \in \Lambda(\mathbf{p}) \right\},$$

where

$$\Lambda(\mathbf{p}) \stackrel{\text{def}}{=} \left\{ (\lambda_i)_{i=1}^I \in [0, 1]^m \mid \begin{array}{l} \sum_{i=1}^I \lambda_i^{(k)} = 1, \\ \lambda_i^{(k)} = 0 \text{ if } p_i^{(k)} \neq p_{max}^{(k)}, \\ i = 1, \dots, I, k = 1, \dots, m \end{array} \right\}$$

is the set of feasible shares of resources. Thus, Assumption 2 says that an auctioneer is able to find feasible shares of resources $(\lambda_i(\mathbf{p}))_{i=1}^I \in \Lambda(\mathbf{p})$. Namely, the auctioneer anticipates producers' individual prices $\mathbf{p} = (p_i)_{i=1}^I$ and distributes scarce resources b between those producers with highest offer prices $p_i = p_{max}$ accordingly to

$$y_i(\mathbf{p}) \stackrel{\text{def}}{=} \lambda_i(\mathbf{p}) \circ b, \quad i = 1, \dots, I.$$

This behavior corresponds to an auction. Namely, let producers have a sequences of parameters $\gamma_i[t] \in \mathbb{R}_+^m, t \geq 0, i = 1, \dots, I$. Primal-dual pairs $\left((x_i[t])_{i=1}^I, (y_i[t])_{i=1}^I, (p_i[t])_{i=1}^I \right)$ of production bundles, distributions and prices are formed by the following iteration from DIS:

Resource Auction with Averaging

1. Auctioneer anticipates producers' prices $\mathbf{p}[t] \stackrel{\text{def}}{=} (p_i[t])_{i=1}^I$,
 finds feasible shares $(\lambda_i(\mathbf{p}[t]))_{i=1}^I \in \Lambda(\mathbf{p}[t])$,
 performs average distribution $y_i[t] \stackrel{\text{def}}{=} \frac{1}{t+1} \sum_{r=0}^t \lambda_i(\mathbf{p}[r]) \circ b$.
2. i -th producer maximizes revenue by $x_i(p_i[t]) \in X_i^*(p_i[t])$
 and updates production statistics $x_i[t] \stackrel{\text{def}}{=} \frac{1}{t+1} \sum_{r=0}^t x_i(p_i[r])$.
3. i -th producer observes individual excess demand $A_i x_i[t] - y_i[t]$
 and forms price forecast $p_i^+[t+1] \stackrel{\text{def}}{=} \frac{1}{\Gamma_i[t]} \circ [A_i x_i[t] - y_i[t]]_+$
 by using average parameter $\Gamma_i[t] \stackrel{\text{def}}{=} \frac{1}{t+1} \sum_{r=0}^t \gamma_i[r] \in \mathbb{R}^m$.
4. i -th producer forms new prices by averaging forecasts
 $p_i[t+1] \stackrel{\text{def}}{=} \frac{1}{t+2} \sum_{r=0}^{t+1} p_i^+[r]$, where $p_i[0] = p_i^+[0] = 0$.

(AUC)

Since AUC is a variant of DIS, its convergence follows directly from Theorem 3.

Theorem 4 (Convergence of Resource Auction) *Let admissible parameters $\gamma_i[t]$ be applied by producers in DIS. Then, the sequence $\left((x_i[t])_{i=1}^I, (y_i[t])_{i=1}^I, (p_i[t])_{i=1}^I \right)$ of production bundles, distributions and prices generated by DIS converges toward the set of primal-dual solutions of (18) and (20). Moreover, the achievable rate of convergence is of the order $O\left(\frac{1}{\sqrt{t+1}}\right)$.*

Together with Lemma 4 we get the following

Corollary 1 (Resource Management via Auction) *Let the sequence of production bundles, distributions and prices $\left((x_i[t])_{i=1}^I, (y_i[t])_{i=1}^I, (p_i[t])_{i=1}^I \right)$ be generated by AUC with producers' admissible parameters $\gamma_i[t]$. Then, the sequence of production bundles and highest offer prices $\left((x_i[t])_{i=1}^I, p_{\max}[t] \stackrel{\text{def}}{=} \max_{i=1, \dots, I} p_i[t] \right)$ converges toward the set of primal-dual solutions of (8) and (10). Moreover, the achievable rate of convergence is of the order $O\left(\frac{1}{\sqrt{t+1}}\right)$.*

Corollary 1 states that the resource management problem can be alternatively solved by an auction. Here, producers update their individual prices rather than a manager does, cf. Section 3.1. The advantage of an auction as opposed to the management of resources is as follows. At an equilibrium price individual producers' demands meet the supplies of resources, which an auctioneer assigns to each of them. Hence, not only the aggregated demand and supply of resources equalize as it is in the case of resource management. Via auction the market of resources balances individually, and every producer gets as much of resources as necessary for optimal production.

Appendix

Proof of Lemma 1:

It is sufficient to show that for all $t \geq 0$ it holds:

$$\Psi(p[t]) - \varphi_t \leq \rho_t, \quad (21)$$

where

$$\begin{aligned} \varphi_t &\stackrel{\text{def}}{=} \min_{p \in \mathbb{R}_+^m} \left[\ell_t(p) + \frac{1}{2} \langle \Gamma[t], p^2 \rangle \right], \quad \ell_t(p) \stackrel{\text{def}}{=} \frac{1}{t+1} \sum_{r=0}^t f(x(p[r])) + \langle b - Ax(p[r]), p \rangle, \\ \rho_t &\stackrel{\text{def}}{=} \frac{1}{t+1} \sum_{r=0}^t \left\langle \frac{1}{\sum_{s=0}^{r-1} \gamma[s]}, \frac{1}{2} (b - Ax(p[r]))^2 \right\rangle. \end{aligned}$$

Indeed, using $x[t] \stackrel{\text{def}}{=} \frac{1}{t+1} \sum_{r=0}^t x(p[r])$ and concavity of f , we have

$$\ell_t(p) \leq f(x[t]) + \langle b - Ax[t], p \rangle,$$

thus,

$$\varphi_t \leq f(x[t]) + \min_{p \in \mathbb{R}_+^m} \left[\langle b - Ax[t], p \rangle + \frac{1}{2} \langle \Gamma[t], p^2 \rangle \right] = f(x[t]) - \left\langle \frac{1}{\Gamma[t]}, \frac{1}{2} (Ax[t] - b)_+^2 \right\rangle. \quad (22)$$

Also, we obtain

$$\begin{aligned} \rho_t &\leq \frac{1}{t+1} \sum_{r=0}^t \left\| \frac{1}{\sum_{s=0}^{r-1} \gamma[s]} \right\|_{\infty, \frac{1}{d}} \left\langle d, \frac{1}{2} (b - Ax(p[r]))^2 \right\rangle \\ &= \frac{1}{t+1} \sum_{r=0}^t \left\| \frac{1}{\sum_{s=0}^{r-1} \gamma[s]} \right\|_{\infty, \frac{1}{d}} \frac{1}{2} \|b - Ax(p[r])\|_{2,d}^2 \\ &\leq \frac{1}{t+1} \sum_{r=0}^t \left\| \frac{1}{\sum_{s=0}^{r-1} \gamma[s]} \right\|_{\infty, \frac{1}{d}} \frac{1}{2} \left(\|A\|_{2,d} \max_{x \in X} \|x\| + \|b\|_{2,d} \right)^2 = \Delta[t] \cdot C_1. \end{aligned} \quad (23)$$

Overall, estimations (21), (22), and (23) provide the assertion of Lemma 1.

In order to complete the proof, we show the validity of inequalities (21) by induction. First, we prove

$$\Psi(p[0]) - \varphi_0 \leq \rho_0.$$

Indeed, in view of $p[0] = 0$ and $\Gamma[0] = \gamma[0] = \sum_{s=0}^{-1} \gamma[s]$, we have

$$\begin{aligned} \Psi(p[0]) - \varphi_0 &= f(x[0]) - \min_{p \in \mathbb{R}_+^m} \left[\ell_0(p) + \frac{1}{2} \langle \Gamma[0], p^2 \rangle \right] \\ &= - \min_{p \in \mathbb{R}_+^m} \left[\langle b - Ax[0], p \rangle + \frac{1}{2} \langle \Gamma[0], p^2 \rangle \right] \\ &\leq - \min_{p \in \mathbb{R}^m} \left[\langle b - Ax[0], p \rangle + \frac{1}{2} \langle \Gamma[0], p^2 \rangle \right] \\ &= \left\langle \frac{1}{\Gamma[0]}, \frac{1}{2} (b - Ax[0])^2 \right\rangle = \rho_0. \end{aligned}$$

Now, assume that the inequality (21) is valid for some $t \geq 0$. In order to show the inequality (21) for $t + 1$, we need the following auxiliary estimations:

$$\ell_{t+1}(p) = \frac{t+1}{t+2} \ell_t(p) + \frac{1}{t+2} [f(x(p[t+1])) + \langle b - Ax(p[t+1]), p \rangle], \quad (24)$$

$$\rho[t+1] = \frac{t+1}{t+2} \rho[t] + \frac{1}{t+2} \left\langle \frac{1}{\sum_{s=0}^t \gamma[t]}, (b - Ax(p[t+1]))^2 \right\rangle, \quad (25)$$

$$\Gamma[t+1] = \frac{t+1}{t+2} \Gamma[t] + \frac{1}{t+2} \gamma[t+1], \quad (26)$$

$$p[t+1] = \frac{t+1}{t+2} p[t] + \frac{1}{t+2} p^+[t+1]. \quad (27)$$

Further, we note that

$$p^+[t+1] = \frac{1}{\Gamma[t]} \circ [Ax[t] - b]_+$$

solves the optimization problem

$$\min_{p \in \mathbb{R}_+^m} \left[\langle b - Ax[t], p \rangle + \frac{1}{2} \langle \Gamma[t], p^2 \rangle \right].$$

Hence, $p^+[t+1]$ also solves

$$\frac{1}{t+1} \sum_{r=0}^t f(x(p[r])) + \min_{p \in \mathbb{R}_+^m} \left[\langle b - Ax[t], p \rangle + \frac{1}{2} \langle \Gamma[t], p^2 \rangle \right] = \min_{p \in \mathbb{R}_+^m} \left[\ell_t(p) + \frac{1}{2} \langle \Gamma[t], p^2 \rangle \right].$$

As consequences, we have

$$\varphi_t = \min_{p \in \mathbb{R}_+^m} \left[\ell_t(p) + \frac{1}{2} \langle \Gamma[t], p^2 \rangle \right] = \ell_t(p^+[t+1]) + \frac{1}{2} \langle \Gamma[t], (p^+[t+1])^2 \rangle, \quad (28)$$

and from the optimality conditions for φ_t :

$$\ell_t(p - p^+[t+1]) + \langle \Gamma[t] \circ p^+[t+1], p - p^+[t+1] \rangle \geq 0 \quad \text{for all } p \in \mathbb{R}_+^m. \quad (29)$$

Using the linearity of ℓ_t w.r.t. p , we estimate in view of (28) and (29):

$$\begin{aligned}
\ell_t(p) + \frac{1}{2} \langle \Gamma[t], p^2 \rangle &= \ell_t(p^+[t+1]) + \frac{1}{2} \langle \Gamma[t], (p^+[t+1])^2 \rangle \\
&\quad + \ell_t(p - p^+[t+1]) + \langle \Gamma[t] \circ p^+[t+1], p - p^+[t+1] \rangle \\
&\quad + \frac{1}{2} \langle \Gamma[t], (p - p^+[t+1])^2 \rangle \\
&\geq \varphi_t + \frac{1}{2} \langle \Gamma[t], (p - p^+[t+1])^2 \rangle.
\end{aligned} \tag{30}$$

Finally, let us proceed with the induction step for (21):

$$\begin{aligned}
\varphi_{t+1} &= \min_{p \in \mathbb{R}_+^m} \left[\ell_{t+1}(p) + \frac{1}{2} \langle \Gamma[t+1], p^2 \rangle \right] \\
&\stackrel{(26),(28)}{\geq} \min_{p \in \mathbb{R}_+^m} \left[\begin{aligned} &\frac{t+1}{t+2} \ell_t(p) + \frac{1}{t+2} [f(x(p[t+1])) + \langle b - Ax(p[t+1]), p \rangle] \\ &+ \frac{1}{2} \cdot \frac{t+1}{t+2} \langle \Gamma[t], p^2 \rangle \end{aligned} \right] \\
&\stackrel{(30)}{\geq} \min_{p \in \mathbb{R}_+^m} \left[\begin{aligned} &\frac{t+1}{t+2} \varphi_t(p) + \frac{1}{t+2} [f(x(p[t+1])) + \langle b - Ax(p[t+1]), p \rangle] \\ &+ \frac{1}{2} \cdot \frac{t+1}{t+2} \langle \Gamma[t], (p - p^+[t+1])^2 \rangle \end{aligned} \right] \\
&\stackrel{(21)}{\geq} \min_{p \in \mathbb{R}_+^m} \left[\begin{aligned} &\frac{t+1}{t+2} \Psi(p[t]) + \frac{1}{t+2} [f(x(p[t+1])) + \langle b - Ax(p[t+1]), p \rangle] \\ &+ \frac{1}{2} \cdot \frac{t+1}{t+2} \langle \Gamma[t], (p - p^+[t+1])^2 \rangle - \frac{t+1}{t+2} \rho_t \end{aligned} \right] \\
&\stackrel{(4)}{\geq} \min_{p \in \mathbb{R}_+^m} \left[\begin{aligned} &\frac{t+1}{t+2} [\Psi(p[t+1]) + \langle b - Ax(p[t+1]), p[t] - p[t+1] \rangle] \\ &+ \frac{1}{t+2} [f(x(p[t+1])) + \langle b - Ax(p[t+1]), p \rangle] \\ &+ \frac{1}{2} \cdot \frac{t+1}{t+2} \langle \Gamma[t], (p - p^+[t+1])^2 \rangle - \frac{t+1}{t+2} \rho_t \end{aligned} \right] \\
&\stackrel{(2),(27)}{=} \min_{p \in \mathbb{R}_+^m} \left[\begin{aligned} &\Psi(p[t+1]) + \frac{1}{t+2} \langle b - Ax(p[t+1]), p - p^+[t+1] \rangle \\ &+ \frac{1}{2} \cdot \frac{t+1}{t+2} \langle \Gamma[t], (p - p^+[t+1])^2 \rangle - \frac{t+1}{t+2} \rho_t \end{aligned} \right] \\
&\geq \min_{y \in \mathbb{R}^m} \left[\begin{aligned} &\Psi(p[t+1]) - \frac{t+1}{t+2} \rho_t \\ &+ \frac{1}{t+2} \langle b - Ax(p[t+1]), y \rangle + \frac{1}{2} \cdot \frac{t+1}{t+2} \langle \Gamma[t], y^2 \rangle \end{aligned} \right] \\
&= \Psi(p[t+1]) - \frac{t+1}{t+2} \rho_t - \frac{1}{t+2} \left\langle \frac{1}{\sum_{s=0}^t \gamma[t]}, (b - Ax(p[t+1]))^2 \right\rangle \\
&\stackrel{(25)}{=} \Psi(p[t+1]) - \rho_{t+1}.
\end{aligned}$$

□

Proof of Lemma 2:

Let $x^* \in X^*$ solve the primal optimization problem (1). Due to Assumption 1, there exists a vector $p^* \in P^*$ of dual (or Lagrange) multipliers such that

$$\langle \nabla f(x^*) - A^T p^*, x^* - x \rangle \geq 0 \quad \text{for all } x \in X, \quad \text{and } \langle p^*, Ax^* - b \rangle = 0.$$

Thus, we obtain

$$f(x) \leq f(x^*) + \langle \nabla f(x^*), x - x^* \rangle \leq f^* + \langle A^T p^*, x - x^* \rangle = f^* + \langle p^*, Ax - b \rangle.$$

Using the latter, we have

$$\begin{aligned} f(x[t]) - \left\langle \frac{1}{\Gamma[t]}, \frac{1}{2} (Ax[t] - b)_+^2 \right\rangle &\leq \max_{x \in X} \left[f(x) - \left\langle \frac{1}{\Gamma[t]}, \frac{1}{2} (Ax - b)_+^2 \right\rangle \right] \\ &\leq f^* + \max_{x \in X} \left[\langle p^*, Ax - b \rangle - \left\langle \frac{1}{\Gamma[t]}, \frac{1}{2} (Ax - b)_+^2 \right\rangle \right] \\ &\leq f^* + \max_{y \in \mathbb{R}^m} \left[\langle p^*, y \rangle - \left\langle \frac{1}{\Gamma[t]}, \frac{1}{2} (y)_+^2 \right\rangle \right] \\ &= f^* + \frac{1}{2} \langle p^{*2}, \Gamma[t] \rangle \leq f^* + \frac{1}{2} \|p^*\|_{2, \frac{1}{d}}^2 \cdot \|\Gamma[t]\|_{\infty, d}. \end{aligned}$$

Strong duality of convex optimization provides $f^* = \Psi^*$, which completes the proof. \square

References

- [1] ARROW, K., HURWICZ, L., (EDS.) *Decentralization and computation in resource allocation*, Cambridge University Press, 1977.
- [2] BECK, A., NEDIĆ, A., OZDAGLAR, A., TEBoulLE, M., *An $O(1/k)$ gradient method for network resource allocation problems*, IEEE Transactions on Control of Network Systems, Vol. 1, pp. 64-73, 2014.
- [3] CAMPBELL, D. E., *Resource allocation mechanisms*, Cambridge University Press, 1987.
- [4] FACCHINEI F., KANZOW C., *Generalized Nash equilibrium problems*, A Quarterly Journal of Operations Research, Vol. 5, pp. 173-210, 2007.
- [5] FISCHER, A., HERRICH, M., SCHÖNEFELD, K., *Generalized Nash equilibrium problems - recent advances and challenges*, Pesquisa Operacional, Vol. 34, pp. 521-558, 2014.
- [6] FRIEDMAN, E. J., OREN, S. S., *The complexity of resource allocation and price mechanisms under bounded rationality*, Economic Theory, Vol. 6, pp. 225-250, 1995.

- [7] GUSTAVSSON, E., PATRIKSSON, M., STRÖMBERG, A.-B., *Primal convergence from dual subgradient methods for convex optimization*, Mathematical Programming, Vol. 150, pp. 365-390, 2015.
- [8] KAKHBOD, A., *Resource allocation in decentralized systems with strategic agents: an implementation theory approach*, Springer, New York, 2013.
- [9] LARSSON, T., LIU, Z., *A Lagrangian relaxation scheme for structured linear programs with application to multicommodity network flows*, Optimization, Vol. 40, pp. 247284, 1997.
- [10] LARSSON, T., PATRIKSSON, M., STRÖMBERG, A.-B., *Ergodic primal convergence in dual subgradient schemes for convex programming*, Mathematical Programming, Vol. 86, pp. 283-312, 1999.
- [11] NEDIĆ, A., OZDAGLAR, A., *Approximate primal solutions and rate analysis for dual subgradient methods*, SIAM Journal Optimization, Vol. 19, pp. 1757-1780, 2009.
- [12] NEMIROVSKII, A. S., YUDIN, D. B., *Cezare convergence of gradient method approximation of saddle points for convex-concave functions*, Dokl. Akad. Nauk SSSR, Vol. 239, pp. 1056-1059, 1978.
- [13] NESTEROV, YU., *Introductory Lectures on Convex Optimization*, Kluwer, Boston, 2004.
- [14] NESTEROV, YU., SHIKHMAN, V., *Quasi-monotone subgradient methods for nonsmooth convex minimization*, Journal of Optimization Theory and Applications, Vol. 165, pp. 917-940, 2015.
- [15] ROCKAFELLAR, R. T., *Convex Analysis*, Princeton University Press, Princeton, 1970.
- [16] RUSZCZYŃSKI, A., *A merit function approach to the subgradient method with averaging*, Optimization Methods and Software, Vol. 23, pp. 161-172, 2008.
- [17] SHERALI, H. D., SEN, S., *A class of convergent primal-dual subgradient algorithms for decomposable convex programs*, Mathematical Programming, Vol. 35, pp. 279-297, 1986.
- [18] SHERALI, H. D., CHOI, G., *Recovery of primal solutions when using subgradient optimization methods to solve Lagrangian duals of linear programs*, Operations Research Letters, Vol. 19, pp. 105-113, 2096.
- [19] SHOR, N., *Minimization methods for nondifferentiable functions*, translated from the Russian by K. C. Kiwiel and A. Ruszczyński, Springer-Verlag, Berlin, 1985.