Composite optimization for the resource allocation problem

Anastasiya Ivanova\textsuperscript{a,b}, Pavel Dvurechensky \textsuperscript{c,d}, Alexander Gasnikov \textsuperscript{a,b,d} and Dmitry Kamzolov \textsuperscript{a}

\textsuperscript{a}Moscow Institute of Physics and Technology, Moscow, Russia; \textsuperscript{b}National Research University Higher School of Economics, Moscow, Russia; \textsuperscript{c}Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany; \textsuperscript{d}Institute for Information Transmission Problems RAS, Moscow, Russia

ABSTRACT

In this paper, we consider resource allocation problem stated as a convex minimization problem with linear constraints. To solve this problem, we use gradient and accelerated gradient descent applied to the dual problem and prove the convergence rate both for the primal iterates and the dual iterates. We obtain faster convergence rates than the ones known in the literature. We also provide economic interpretation for these two methods. This means that iterations of the algorithms naturally correspond to the process of price and production adjustment in order to obtain the desired production volume in the economy. Overall, we show how these actions of the economic agents lead the whole system to the equilibrium.

ARTICLE HISTORY

Received 8 March 2019
Accepted 27 December 2019

KEYWORDS
Decentralized pricing; primal–dual method; accelerated gradient method; gradient method; composite optimization

2010 MATHEMATICS SUBJECT CLASSIFICATIONS
90C25; 68Q25; 91B32

1. Introduction

In this paper, we consider a resource allocation problem in an economy consisting of distributed set of producers which are managed by a centralized price adjustment mechanism. Our approach is based on the state-of-the-art convex optimization methods, i.e. we consider the resource allocation as a convex optimization problem, solve it by first-order methods, provide convergence analysis and give an economic interpretation of the steps of these methods.

The problem of optimal resource allocation is to maximize producers aggregated profits by sharing available resources. Popularized and advocated mainly in the monograph [2], the mechanisms of decentralized resource allocation gained a lot of attention in economics and operations research since then, see e.g. [7,15,21] and references therein. Each producer seeks to minimize its own costs and, in total, all the producers need to produce a certain amount of products. This problem can be cast as an optimization problem with the objective corresponding to the aggregated cost function of all producers and constraints corresponding to the condition for the necessary volume of production. We assume that constraints are linear and separable. In this optimization problem, primal variables are production bundles and dual variables represent prices of resources.

CONTACT
Dmitry Kamzolov kamzolov.dmitry@phystech.edu, dkamzolov@yandex.ru

© 2020 Informa UK Limited, trading as Taylor & Francis Group
Solution of this optimization problem depends on the way, how agents in this economy can exchange information with each other. One of the mechanisms is called Walrasian in the literature [3]. This mechanism is an iterative process. At each iteration, some agents formulate a supply for others. Further, the second agents, based on the supply, form the optimal demand. Then, the first agents again form the optimal supply based on demand and so on until the system comes to Walrasian equilibrium. In the economic literature, this iterative process is called the Walrasian tâtonnement. For simplicity, let us consider an economy with only one good. We assume that there are some small producers and one big consumer (Center). Each producer has its own cost function and this functions are unknown to the Center. The Center sets a price for the good and producers provide the produced quantity for this price. The goal of the Center is to iteratively find such a price that all the factories produce certain amount of good in total. This will be a Walrasian equilibrium in the system. One example of such situation can be a centrally governed economic like in USSR. The price set by the Center plays the role of control parameter of the system and the goal is to provide production determined by plan. Another example could be farmers as producers and some large retail company as the Center. Retail company knows the demand for the good and sets the price so that the produced amount is equal to the demand. In the literature, an iterative numerical algorithm, which corresponds to Walrasian mechanism is based either on the dichotomy method (for one-product economy) [18] or the ellipsoid method (for many-product economy) [15]. These algorithms are effective in the case when the amount of producers is small.

In this paper, following [25], we consider a different price adjustment mechanism. In particular, we consider the resource allocation problem without centralized price control, each producer setting up its own price for selling products to the Center. The Center knows the amount which needs to be produced by all the producers in total, selects the most advantageous offers (i.e. selects the offers with the best price) and tries to purchase the product in the required volume. Producers adjust the volume of product and the prices, based on the volume bought from them by the Center and the demand from the Center for this particular factory. In their paper [25], the authors use dual subgradient method with averaging as a numerical algorithm for this problem. The main advantage of their algorithm is that they provide convergence rate for the whole primal–dual sequence unlike the optimization literature which gives convergence rate for the running average. Their algorithm has an optimal convergence rate for convex nonsmooth optimization problems $O(1/\sqrt{t})$, $t$ being the iteration counter. In contrast to [25], in this paper, under an additional assumption of strongly convexity of the primal objective, we consider the dual optimization problem as a composite minimization problem, meaning that the objective in the dual problem is a sum of two functions a smooth and a simple non-smooth. We use gradient descent to obtain convergence rate $O(1/t)$ and accelerated gradient descent to obtain convergence rate $O(1/t^2)$.

Since our methods use distributed calculations, we mention as a related work some papers on distributed optimization [6,9,10,13,20,22,23,28–30]. These papers consider decentralized distributed optimization as opposed to our situation of centralized setup with one center.

The paper is organized as follows. In Section 2, we consider the primal problem and describe its economical interpretation. In Section 3, we describe the method of subgradient projection for the resource allocation problem and give the interpretation for the
step. In Section 4, we use composite gradient method for resource allocation problem and obtain estimation for the convergence rate. In Section 5, we consider accelerated composite gradient descent. And in Section 6, we show some experiments that verify our theory.

2. Problem statement

In this section, we provide the statement of the resource allocation problem. For simplicity, we start with a one-product economy. The case of many products will be considered in Appendix. We assume that there is a Center and \( n \) producers which produce one product. Each producer has its own cost function \( f_k(x_k), k = 1, \ldots, n \), representing the total cost of production of a volume \( x_k \in \mathbb{R} \) – the volume of product produced by the producer \( k \) in 1 year. Since the producers are independent, the cost functions of the producers are unknown to the Center, and each producer knows only its own cost function. Each producer is also entitled to set its own price for product, but the price does not affect the quality of the product, i.e. all producers produce the same product, only at different prices. The Center buys product from the producers and chooses its strategy in such a way that the total production volume per year by all producers is not less than \( C \). To do so, the Center needs to find \( y_k \) – the volume of product which is purchased from the producer \( k \). Then, each producer produces at least the volume \( y_k \) of the product, and the goal of the system is to minimize the total cost of production. Thus we consider the following resource allocation problem

\[
(P) \quad \min_{\sum_{k=1}^{n} y_k \geq C, x_k \geq y_k, y_k \geq 0, x_k \geq 0, k = 1, \ldots, n} f(x) = \sum_{k=1}^{n} f_k(x_k),
\]

where cost functions \( f_k(x_k), k = 1, \ldots, n \), are increasing and \( \mu \)-strongly convex, i.e. \( f_k''(x_k) \geq \mu, \forall x \geq 0, k = 1, \ldots, n \).

**Remark 2.1:** We point that, from the optimization point of view the variables \( y_k, k = 1, \ldots, n \), are not necessary. We introduce them in order to the constructed dual problem and primal–dual method for solving the primal–dual pair of problems has an economic interpretation. The reformulation (P) is one of our contributions.

**Remark 2.2:** Assumption of strong convexity of functions \( f_k(x_k), k = 1, \ldots, n \), holds, for example, when these functions are twice continuously differentiable and have positive second derivative. Economically this means that the production cost grows faster than the volume of the production. In other words, the production cost of a new unit of volume grows as the volume of production grows. For example, this happens for Agriculture. If the producer grows wheat, then the more he wants to produce from 1 hectare, the more he should invest in fertilizers, chemicals from pests or even genetic technology. For a factory the producer has to invest more and more in more advanced facilities such as robots, production machines and so on.

Introducing dual variables \( p_k, k = 1, \ldots, n \), and using the duality theory, we obtain

\[
\min_{\sum_{k=1}^{n} y_k \geq C, x_k \geq y_k, y_k \geq 0, x_k \geq 0, k = 1, \ldots, n} f(x)
\]
\[
\begin{align*}
\min_{\sum_{k=1}^{n} y_k \geq C, y_k \geq 0; x_k \geq 0, k = 1, \ldots, n} & \left\{ f(x) + \sum_{k=1}^{n} \max_{p_k \geq 0} p_k (y_k - x_k) \right\} \\
= & -\min_{p_1, \ldots, p_n \geq 0} \left\{ \sum_{k=1}^{n} \max_{x_k \geq 0} (p_k x_k - f_k(x_k)) - \min_{\sum_{k=1}^{n} y_k \geq C; y_k \geq 0} \sum_{k=1}^{n} p_k y_k \right\} \\
& \times [y_k^* = C, \text{ where } k^* = \arg\min_k p_k \text{ and } y_k^* = 0 \text{ for } k \neq k^*] \\
= & -\min_{p_1, \ldots, p_n \geq 0} \left\{ \sum_{k=1}^{n} \max_{x_k \geq 0} (p_k x_k - f_k(x_k)) - C \min_{k=1, \ldots, n} p_k \right\} \\
= & -\min_{p_1, \ldots, p_n \geq 0} \left\{ \sum_{k=1}^{n} \left(p_k x_k(p_k) - f_k(x_k(p_k))\right) - C \min_{k=1, \ldots, n} p_k \right\},
\end{align*}
\]

where

\[x_k(p_k) = \arg\max_{x_k \geq 0} \left\{ p_k x_k - f_k(x_k) \right\}, \quad k = 1, 2, \ldots, n. \quad (1)\]

Then the dual problem (up to a sign) has the following form:

\[(D) \quad \varphi(p_1, \ldots, p_n) = \sum_{k=1}^{n} \left(p_k x_k(p_k) - f_k(x_k(p_k))\right) - C \min_{k=1, \ldots, n} p_k \rightarrow \min_{p_1, \ldots, p_n \geq 0}. \]

Note that the Slater’s constraint qualification condition holds for the primal problem \((P)\). Thus the strong duality holds and both the primal problem \((P)\) and the dual problem \((D)\) have solutions. Throughout the paper, we solve the dual problem by different first-order methods, interpret the steps of these methods and show, how the primal variables \(x_k, y_k, k = 1, \ldots, n\), can be reconstructed. We point that primal–dual gradient methods \([1,4,8,11,12,17]\) are not applicable here since the dual problem \((D)\) is a composite optimization problem. Moreover these are not distributed methods.

### 3. Subgradient descent

For the sake of completeness, in this section, we consider dual problem \((D)\) as a non-smooth optimization problem and apply subgradient method to solve it with the rate \(O(1/\sqrt{t})\). We also provide an economic interpretation of the numerical procedure based on subgradient method. The material of this section is not new and mostly follows \([18]\), but we include it to be able to compare, in the next sections, its convergence rate and interpretation with faster approaches based on composite gradient descent. We also mention an alternative switching subgradient scheme \([5,26]\), which can be used for non-smooth problems.

To solve the problem \((D)\), we use the projected subgradient method with the step given by

\[p_{t+1}^+ = \left[p_t^+ - hg(p_t^+)\right]_+, \quad (2)\]
where \( h \) is the stepsize, which we determine later. The subgradient of the objective function in the dual problem (D) can be written in the following form:

\[
g(p_1, \ldots, p_n) = x(p) - C\lambda(p),
\]

where \( x(p) = (x_1(p_1), \ldots, x_n(p_n))^\top \), \( \lambda(p) = (\lambda_1(p_1), \ldots, \lambda_n(p_n))^\top \) and \( \sum_{k=1}^n \lambda_k(p_k) = 1 \), \( \lambda_k(p_k) \geq 0 \) if \( k \in \text{Arg min}_{j=1,\ldots,n} p_j \) and \( \lambda_k(p_k) = 0 \), if \( k \notin \text{Arg min}_{j=1,\ldots,n} p_j \). Note that here and below \( g(p_1, \ldots, p_n) \in \partial \psi(p_1, \ldots, p_n) \) is an arbitrary subgradient, i.e. an arbitrary element of the convex compact set – subdifferential.

| General projected subgradient method |
|-------------------------------------|

**Input:** \( h \) – stepsize, \( p^0 \) – starting point.

1. Set \( p^0_k, k = 1, \ldots, n \) and calculate

\[
x_k(p_k^t) = \arg\max_{x_k \geq 0} \left\{ p_k^t x_k - f_k(x_k) \right\}, \quad k = 1, 2, \ldots, n.
\]

2. Form a vector \( \lambda(p^t) \) as \( \lambda^t = (\lambda_1^t, \ldots, \lambda_n^t)^\top \), where \( \sum_{k=1}^n \lambda_k^t = 1 \) and if \( k \in \text{Arg min}_{j=1,\ldots,n} p_j^t \), then \( \lambda_k^t \geq 0 \), otherwise, if \( k \notin \text{Arg min}_{j=1,\ldots,n} p_j^t \), then \( \lambda_k^t = 0 \)

3. Do the step

\[
p^{t+1} = \left( p^t - h(x(p^t) - C\lambda(p^t)) \right)_+.
\]

Next, we discuss the economic interpretation of the steps of the subgradient method. First of all, \( C\lambda_k(p^t) \) can be seen as the volume the Center intends to purchase from the \( k \)th producer on \( t \) iteration. Indeed, the dual variable \( \nu \) corresponds to the vector of prices set by the producers and non-zero components of the vector \( \lambda \) can correspond only to those producers, who have minimal price. Since \( \sum_{k=1}^n \lambda_k^t = 1 \) and all \( \lambda_k^t \) are non-negative, \( \lambda_k^t \) for \( k \in \text{Arg min}_{j=1,\ldots,n} p_j^t \) can be seen as the relative proportions of purchase from the producers having minimal price. Thus \( C\lambda_k(p^t) \) can be seen as the volume the Center intends to purchase from the \( k \)th producer on \( t \) iteration.

Further, each \( k \)th component of the subgradient (3) can be interpreted as the difference between the production \( x_k(p_k) \) of the \( k \)th factory and the volume \( C\lambda_k \) of the Center’s demand for this factory. For some factories, for which \( C\lambda_k^t \) is positive, it can happen that \( x_k(p_k) - C\lambda_k^t < 0 \). This is a signal for the \( k \)th producer that the demand exceeds the supply and the \( k \)th price can be increased together with the increase of the produced volume.

Finally, the subgradient step is the production adjustment steps for each producer, i.e. each producer counts how much its production differs from the desired volume of the Center’s purchase from this producer this year. If the Center does not want to buy anything from the producer or buys less than it produced, then the producer lowers the price. If the
Center is ready to buy more than the producer produced, the producer raises the price. In the case of equality, the producer does not change anything.

As a result the policy for the Center and producers is as follows.

---

**Subgradient method for the resource allocation**

**Input:** \( \varepsilon > 0 \) – accuracy, \( p^0 \) – starting point.

1. Set the stepsize \( h = \frac{\varepsilon}{nC^2} \).
2. Given the price vector \( p^t \) for the current year, producers calculate the optimal production plan for these prices as
   \[
   x_k(p^t_k) = \arg\max_{x_k \geq 0} \left\{ p^t_k x_k - f_k(x_k) \right\}, \quad k = 1, 2, \ldots, n,
   \]
   and communicates this information to the Center.
3. The Center determines the shares of purchases for each producer, i.e. forms a vector \( \lambda(p^t) \) as
   \[
   \lambda^t = (\lambda^t_1, \ldots, \lambda^t_n)^	op, \quad \sum^n_{k=1} \lambda^t_k = 1, \quad \lambda^t_k \geq 0 \text{ if } k \in \text{Arg min}_{j=1,\ldots,n} p_j^t \text{ and } \lambda^t_k = 0, \text{ if } k \notin \text{Arg min}_{j=1,\ldots,n} p_j^t
   \]
   and sends this vector to all factories.
4. Each factory adjusts the price for the next year as follows:
   \[
   p^{t+1} = \left(p^t - h(x(p^t) - C\lambda(p^t)) \right)_+.
   \]

---

To state the convergence rate result, we need introduce an upper bound for the optimal value of the prices.

**Lemma 3.1:** Let the \( p^* \) be a solution to the dual problem \( (D) \). Then

\[
\|p^*\|_2 \leq \sqrt{np_{\max}},
\]

where

\[
p_{\max} := \frac{n}{C} \left( \sum^n_{k=1} f_k \left( \frac{2C}{n} \right) - \sum^n_{k=1} f_k(0) \right).
\]

The proof of this lemma is deferred to the Appendix.

Then, we can formulate the following theorem about convergence rate...
Theorem 3.1 ([18]): Let Algorithm (4) be run with starting point $p^0$ satisfying $0 \leq p^0_k \leq p_{\text{max}}$, $k = 1, \ldots, n$ for

$$N = \left\lceil \frac{164 (C n p_{\text{max}})^2}{\varepsilon^2} \right\rceil$$

steps. Then

$$f(x^N) - f(x^*) \leq \varepsilon, \quad C - \sum_{k=1}^{n} x_k^N \leq \frac{\varepsilon}{3 p_{\text{max}}},$$

(7)

where $x^N = 1/N \sum_{t=0}^{N-1} x(p^t)$.

Note that the number of iterations $N$ to achieve accuracy $\varepsilon$ is very large. To improve the number of iterations, in the following sections we consider the methods based on the composite optimization approach.

4. Composite gradient method for the resource allocation problem

In this section, we consider a non-accelerated composite gradient method to solve the dual problem $(D)$, including its interpretation and convergence rate estimate. The problem $(D)$ can be rewritten as

$$\varphi(p_1, \ldots, p_n) = \psi(p_1, \ldots, p_n) + \xi(p_1, \ldots, p_n),$$

where

$$\psi(p_1, \ldots, p_n) = \sum_{k=1}^{n} \left\{ p_k x_k(p_k) - f_k(x_k(p_k)) \right\} = \langle p, x(p) \rangle - f(x(p))$$

(8)

is convex function. Gradient of this function

$$\nabla \psi(p) = x(p)$$

(9)

satisfies Lipschitz condition (see e.g. [24])

$$\| \nabla \psi(p^1) - \nabla \psi(p^2) \|_2 \leq L_{\psi} \| p^1 - p^2 \|_2, \quad \forall p^1, p^2 \geq 0,$$

(10)

where $L_{\psi} = n/\mu$ (see e.g.[18]) and

$$\xi(p_1, \ldots, p_n) = -C \min_{k=1, \ldots, n} p_k$$

is convex nonsmooth function. The idea is to use first-order information about $\psi$ and use the function $\xi$ as a whole, see the main projected gradient step (11). This allows the method to work in accordance with the smoothness properties of $\psi$, i.e. be faster since $\psi$ is smooth. The price is that the step (11) has to be simple enough. The next lemma shows that this is indeed the case and the solution can be found explicitly.
General composite projected gradient method

**Input:** $N > 0$ – number of steps, $L_\psi$ – Lipschitz constant of gradient $\psi$, $p^0$ – starting point.

1. Find
   
   $$x_k(p_k^t) = \arg\max_{x_k \geq 0} \left\{ p_k^tx_k - f_k(x_k) \right\}, \quad k = 1, 2, \ldots, n.$$

2. Do the step
   
   $$p^{t+1} = \arg\min_{p \geq 0} \left\{ \langle \nabla \psi(p^t), p - p^t \rangle - C \min_{k=1,\ldots,n} p_k + \frac{L_\psi}{2} \| p - p^t \|_2^2 \right\}. \quad (11)$$

**Lemma 4.1:** Let $\tilde{p}^{t+1} = p^t - (1/L_\psi)x(p^t)$. Then $p^{t+1}$ in (11) is defined as follows:

$$p_k^{t+1} = \max\left(p_{\text{center}}^{t+1}, \tilde{p}_k^{t+1}\right), \quad k = 1, \ldots, n,$$

where if $\sum_{k=1}^n (-p_k^{t+1})_+ \geq C/L_\psi$, then $p_{\text{center}}^{t+1} = 0$, else $p_{\text{center}}^{t+1} > 0$ is a solution of equation

$$\sum_{k=1}^n \left(p_{\text{center}}^{t+1} - \tilde{p}_k^{t+1}\right)_+ = \frac{C}{L_\psi}.$$

The proof of Lemma 4.1 can be found in Appendix. Note that the step (11) can be rewritten as

$$p^{t+1} = \left[ p^t - \frac{1}{L_\psi} \left( x(p^t) - C\lambda(p^{t+1}) \right) \right]_+, \quad (12)$$

where $\lambda(p^{t+1})$ is such that $\sum_{k=1}^n \lambda_k(p_k^{t+1}) = 1$, $\lambda_k(p_k^{t+1}) \geq 0$ if $k \in \arg\min_{j=1,\ldots,n} p_j^{t+1}$ and $\lambda_k(p_k^{t+1}) = 0$, if $k \notin \arg\min_{j=1,\ldots,n} p_j^{t+1}$. This equality looks very similar to equality (2).

**4.1. Composite gradient method for the resource allocation problem**

In this section, we apply general projected gradient method to the resource allocation problem and give its interpretation, i.e. describe the strategy of the Center and producers, and state the convergence rate theorem for this method. First of all, from (10), since $\nabla \psi(p) = x(p)$, the constant $L_\psi$ determines the relation between prices $p$ and production $x(p)$, i.e. says, what is the maximum change in the production if the price changes. From Lemma 4.1, each year $t$ the Center, knowing the prices and production of each producer, forms a prediction $\tilde{p}_k^{t+1} = p_k^t - (1/L_\psi)x_k(p_k^t)$ for the lowest possible producers’ prices vector for the next year. After that, the goal of the Center is to set its purchase price $p_{\text{center}}^{t+1}$ to
satisfy the total demand $C$. To explain, how it is done, let us look at (12), which looks very similar to (2). The key difference is that in (12) $\lambda$ depends on the unknown price vector $p^{t+1}_k$, resulting in the implicit definition of this price vector as a solution to nonlinear equation. But, $C\lambda(p^{t+1})$ is still an estimate of the purchase from each producer for the next year and the Center chooses to buy only from the producers having the lowest price. By Lemma 4.1, the solution of (12) can be written as

$$p^{t+1}_k = \left[ \tilde{p}^{t+1}_k + \frac{C}{L_{\psi}} \lambda(p^{t+1}) \right]_+,$$

where $\lambda_k(p^{t+1}) = \frac{L_{\psi}}{C} \left( p^{t+1}_{center} - \tilde{p}^{t+1}_k \right)_+$. Thus the estimate for the purchase from the $k$th producer is $L_{\psi}(p^{t+1}_{center} - \tilde{p}^{t+1}_k)_+$, i.e. if the predicted price $\tilde{p}^{t+1}_k$ is higher than $p^{t+1}_{center}$, there will be no purchase from this producer. Again, by Lemma 4.1, if for the smallest possible price $p^{t+1}_{center} = 0$, the total purchased amount $L_{\psi} \sum_{k=1}^n (\tilde{p}^{t+1}_k)_+$ is greater than $C$, the Center sets its price $p^{t+1}_{center} = 0$. Otherwise, it determines such price $p^{t+1}_{center}$ that the total estimated purchase $L_{\psi} \sum_{k=1}^n (p^{t+1}_{center} - \tilde{p}^{t+1}_k)_+$ is exactly $C$. Having defined $p^{t+1}_{center}$, the Center informs the producers about this price.

From the perspective of the producer, similarly to Section 3, each producer $k$, knowing the price $p^{t}_k$ for the current year $t$, determines the optimal plan $x_k(p^{t}_k)$ for this price and reports the price and production to the Center. After receiving $p^{t+1}_{center}$ from the Center, each producer adjusts its price $p^{t+1}_k$ for the next year as $p^{t+1}_k = \max(p^{t+1}_{center}, \tilde{p}^{t+1}_k)$. This means that the lowest possible price $\tilde{p}^{t+1}_k$ is compared to the Center’s price. If $p^{t+1}_{center} \geq \tilde{p}^{t+1}_k$, the producer can increase the price up to $p^{t+1}_{center}$ and the Center will buy. On the other hand, if $p^{t+1}_{center} < \tilde{p}^{t+1}_k$, there is no sense for the producer to sell at the price $p^{t+1}_{center}$ and have a loss.

We summarize the strategies of the Center and producers in Algorithm (13).

Note that at equilibrium the Center will purchase from all producers and its price will be constant at all iterations. That is, all producers will be able to set a minimum price for the next year so that the Center buys from them the optimal volume.

---

**Composite gradient method for the resource allocation**

**Input:** $N > 0$ – number of steps, $L_{\psi}$ – Lipschitz constant of gradient $\psi$, $p^0$ – starting point.

(1) Knowing the prices $p^t_k$, $k = 1, \ldots, n$, for the current year $t$, producers calculate the optimal plan for the production according these prices

$$x_k(p^t_k) = \arg\max_{x_k \geq 0} \left\{ p^t_k x_k - f_k(x_k) \right\}, \quad k = 1, 2, \ldots, n.$$
(1) The Center forms a prediction for the lowest possible producers’ prices vector for the next year
\[ \hat{p}^{t+1}_k = p^t_k - \frac{1}{L} x_k(p^t_k), \quad k = 1, 2, \ldots, n. \]

(2) The Center determines the price \( p^{t+1}_{\text{center}} \) at which it will purchase product for the next year \( t + 1 \) and sends this price to all factories.
- If \( \sum_{k=1}^{n} (-\hat{p}^{t+1}_k) + \geq \frac{C}{L} \) then \( p^{t+1}_{\text{center}} = 0 \).
- Else \( p^{t+1}_{\text{center}} > 0 \) and solves equation
\[ \sum_{k=1}^{n} (p^{t+1}_{\text{center}} - \hat{p}^{t+1}_k) + = \frac{C}{L}. \]

(3) Each producer adjusts the price for the next year as follows:
\[ p^{t+1}_k = \max (p^{t+1}_{\text{center}}, \hat{p}^{t+1}_k), \quad k = 1, \ldots, n. \]

Theorem 4.1: Let Algorithm (13) be run for \( N \) steps with starting point \( p^0 \) satisfying \( 0 \leq p^0_k \leq p_{\text{max}}, \quad k = 1, \ldots, n \), where \( p_{\text{max}} \) is given in (6). Then
\[ f(x^N) - f(x^*) \leq f(x^N) + \varphi(p^*) \leq \varphi(p^N) + f(x^N) \leq \frac{82p_{\text{max}}^2 n^2}{N \mu}, \]
\[ \left[ C - \sum_{k=1}^{n} x^N_k \right] + \leq \frac{82p_{\text{max}}^2 n^2}{3N \mu}, \]
where \( p^N = 1/N \sum_{t=1}^{N} b^t \) and \( x^N = 1/N \sum_{t=0}^{N-1} x(p^t) \).

The proof of this theorem is deferred to the Appendix.

Let us make a remark on the complexity of this procedure. Assume that we want to solve problem \( (P) \) with accuracy \( \varepsilon \) in the following sense:
\[ f(x^N) - f(x^*) \leq \varepsilon, \quad C - \sum_{k=1}^{n} x^N_k \leq \frac{\varepsilon}{3p_{\text{max}}}. \] \hspace{1cm} (14)

To solve this problem, we consider the dual problem \( (D) \) and solve it by the composite gradient method starting with \( p^0 \) satisfying \( 0 \leq p^0_k \leq p_{\text{max}}, \quad k = 1, \ldots, yn \), where \( p_{\text{max}} \) is given in (6). Theorem 4.1 states that Algorithm will find the solution no later than after \( N = 82p_{\text{max}}^2 n^2 / \varepsilon \mu \) iterations.
5. Accelerated composite gradient method for the resource allocation problem

In this section, we use accelerated composite gradient method to solve the resource allocation problem. Accelerated algorithm allows to improve the convergence rate in comparison to previous section.

| Accelerated composite gradient descent scheme |
|-----------------------------------------------|
| **Input:** \( N > 0 \) – number of steps, \( L_\psi \) – Lipschitz constant of gradient \( \psi \), \( p^0, y^0, w^0 \) – starting points, \( \alpha_0 = A_0 = 0 \). |
| (1) Find \( \alpha_{t+1} \) as the largest root of the equation |
| \[ A_{t+1} := A_t + \alpha_{t+1} = L_\psi \alpha_{t+1}^2. \] |
| (2) Calculate |
| \[ p^{t+1} = \frac{\alpha_{t+1} y_t + A_t w_t}{A_{t+1}}. \] |
| (3) Calculate |
| \[ y^{t+1} = \argmin_{p \geq 0} \left\{ \alpha_{t+1} \left( \langle \nabla \psi (p^{t+1}), p - p^{t+1} \rangle + \xi (p) \right) + \frac{1}{2} ||p - y_t||^2 \right\}. \] |
| (4) Calculate |
| \[ w^{t+1} = \frac{\alpha_{t+1} y^{t+1} + A_t w_t}{A_{t+1}}. \] |
| (5) Set \( x^{t+1} = \frac{1}{A_{t+1}} \sum_{i=0}^{t+1} \alpha_i x(p^i) = \frac{\alpha_{t+1} x(p^{t+1}) + A_t x^t}{A_{t+1}}. \) |

We can rewrite (17) step as follows:

\[
y^{t+1} = \argmin_{p \geq 0} \left\{ \alpha_{t+1} \left( \langle \nabla \psi (p^{t+1}), p - p^{t+1} \rangle - C \min_{k=1,...,n} p_k \right) + \frac{1}{2} ||p - y_t||^2 \right\}
\]

\[
= \argmin_{p \geq 0} \left\{ \langle x(p^{t+1}), p - p^{t+1} \rangle + C \max_{k=1,...,n} (-p_k) + \frac{1}{2 \alpha_{t+1}} ||p - y_t||^2 \right\}
\]

\[
= \argmin_{p \geq 0} \left\{ \max_{k=1,...,n} (-p_k) + \frac{1}{2 C \alpha_{t+1}} ||p - (y^t - \alpha_{t+1} x(p^{t+1}))||^2 \right\}.
\]

Define \( \tilde{y}^{t+1} = y^t - \alpha_{t+1} x(p^{t+1}) \) and then, using Lemma 4.1, we obtain the following solution:
• If \( \sum_{k=1}^{n} (-\tilde{y}_{k}^{t+1})_+ \geq C_{t+1} \) then \( y_{center}^{t+1} = 0 \) and

\[
y_{k}^{t+1} = \max \left( 0, \tilde{y}_{k}^{t+1} \right), \quad k = 1, \ldots, n.
\]

• Else \( y_{center}^{t+1} > 0 \) is determined from

\[
\sum_{k=1}^{n} (y_{center}^{t+1} - \tilde{y}_{k}^{t+1})_+ = C_{t+1}
\]

and

\[
y_{k}^{t+1} = \max \left( y_{center}^{t+1}, \tilde{y}_{k}^{t+1} \right), \quad k = 1, \ldots, n.
\]

We apply this general method to the resource allocation problem. Let us make some remarks on the interpretation.

Note that in the non-accelerated method we have fixed \( L_{\psi} \), which we interpreted as the expectation of the Center on how the amount of production (of any producer) changes, if the purchase price is changed by one. This was the upper bound for this value. In the accelerated method, this role is played by \( \alpha_{t} \), which is adjusted on each iteration and allows to have faster convergence.

Vector \( w^{t+1} \) can be interpreted as an average of the historical price data \( y^{t}, t \geq 0 \). As we prove in next theorem, \( w^{t+1} \) is a good approximation for the solution of the dual problem. This price can give a hint to each producer on the value of the predicted price for the next period. Thus this average historical price is combined with the prediction price \( \tilde{y}^{t} \) from the previous period to form \( p^{t+1} \). Then the optimal production plan is calculated and, similarly to the non-accelerated method, producers predict the price for next year \( \tilde{y}^{t+1} \) and communicate it to the Center. After getting from the Center \( y_{center}^{t+1} \), each producer adjust the prediction price for the next year.

**Accelerated composite gradient descent for the resource allocation**

| Input: \( N > 0 \) – number of steps, \( L_{\psi} \) Lipschitz constant of gradient \( \psi \), \( p^{0} = y^{0} = w^{0} \) – starting point. |
|---|
| (1) In the current year \( t \) producers find \( \alpha_{t+1} \) as the largest root of the equation |
| \( A_{t+1} := A_{t} + \alpha_{t+1} = L_{\psi} \alpha_{t+1}^{2} \). |
| (2) All producers calculate the average price \( p^{t+1} \) as the following convex combination |
| \( p_{k}^{t+1} = \frac{\alpha_{t+1} y_{k}^{t} + A_{t} w_{k}^{t}}{A_{t+1}}, \quad k = 1, \ldots, n \). |
and calculate the optimal plan for the production according to this price as
\[
x_k(p_k^{t+1}) = \arg \min_{x_k \geq 0} \left\{ p_k^{t+1} x_k - f_k(x_k) \right\}, \quad k = 1, 2, \ldots, n.
\]

(1) Each producer predicts the price for the next year \( t + 1 \) as
\[
y^t_{k+1} = y^t_k - \alpha_{t+1} x_k(p^{t+1}), \quad k = 1, \ldots, n
\]
and sends this information to the Center.

(2) The Center determines the prediction price \( y^t_{center} \) at which it will purchase product for the next year \( t + 1 \) as
- If \( \sum_{k=1}^n (-\tilde{y}^t_k) + \geq C\alpha_{t+1} \) then \( y^t_{center} = 0 \)
- Else \( y^t_{center} > 0 \) and is determined from
\[
\sum_{k=1}^n (y^t_{center} - \tilde{y}^t_k)_+ = C\alpha_{t+1},
\]
and sends this price to all producers.

(3) Each producer adjusts the prediction price for the next year as
\[
y^t_k = \max \left( y^t_{center}, \tilde{y}^t_k \right), \quad k = 1, \ldots, n.
\]
and calculates the historical price for the next year
\[
w^t_k = \frac{\alpha_{t+1} y^t_k + A_t w^t_k}{A_{t+1}}, \quad k = 1, \ldots, n.
\]

**Theorem 5.1:** Let Algorithm (20) be run for \( N \) steps with starting points \( p^0 = y^0 = w^0 \) satisfying \( 0 \leq p^0_k \leq p_{max}, \quad k = 1, \ldots, n \), where \( p_{max} \) is given in (6). Then
\[
f(x^N) - f(x^*) \leq f(x^N) + \varphi(w^*) \leq \varphi(w^N) + f(x^N) \leq \frac{148n^2 p_{max}^2}{(N + 1)^2 \mu},
\]
\[
\left[ C - \sum_{k=1}^n x^N_k \right]_+ \leq \frac{148n^2 R p_{max}}{5(N + 1)^2 \mu},
\]
where \( x^N = (1/A_N) \sum_{t=0}^N \alpha_t x(p^t) \)

The proof of this theorem mostly follows the steps of Theorem 4.1 in the previous section, but we give all steps for the reader’s convenience in Appendix.
Figure 1. Projected gradient versus accelerated gradient descent, 100 factories, cost functions (21), dual value in log scale, 1 run and mean of 20 runs.

Figure 2. Projected gradient versus accelerated gradient descent, 100 factories, cost functions (21), inexactness of constraints in log scale, 1 run and mean of 20 runs.

6. Numerical experiments

In this section, we compare and verify the algorithms presented in the paper. All the algorithms were implemented in the programming language Python.

First, we compare composite projected gradient method and accelerated composite gradient descent on a simple problem. We assume that there are 100 local producers of some long-living product, for example, wood. All this producers sell it to pulp and paper company (Center). Center needs to buy 10,000 tons of wood. Each local producer has the following cost function:

\[ f_k(x_k) = \alpha_k x_k + \frac{\mu}{2} x_k^2, \quad k = 1, \ldots, 100, \tag{21} \]

where \( \alpha_k \) is the different local cost of production and transportation for each producer, \( \mu \) is the cost of production that depends on technology and risks, that's why the same for producers. We consider this problem with random \( \alpha_k \in N[100; 400], \mu = 2. \) Next in graphics the left side will be experiments for one start with random \( \alpha \) and the right side will be the mean of 20 random starts. The practical convergence rate for the objective value of the dual problem (D) is shown in log scale in Figure 1. The convergence rate for the constraint feasibility is shown in log scale in Figure 2. The convergence rate for the value of the duality gap is shown in log scale in Figure 3.
Figure 3. Projected gradient versus accelerated gradient descent, 100 factories, cost functions (21), value of duality gap in log scale, 1 run and mean of 20 runs.

Figure 4. Projected gradient versus accelerated gradient descent, 100 factories, cost functions (21), dual value in log scale, mean of 20 runs.

Figure 5. Projected gradient versus accelerated gradient descent, 100 factories, cost functions (21), inexactness of constraints in log scale, mean of 20 runs.

From the plots, we can see that both methods work well. Projected gradient is monotone and converges in accordance with theoretical convergence rates. Accelerated gradient descent is non-monotone, faster than projected gradient and converges in accordance with theoretical convergence rates.
Second, we consider, how the results depend on the strong convexity parameter $\mu$. For that, we consider functions (21) with $\alpha \in \mathbb{N}[100; 400]$ and $\mu \in \{1, 5, 25, 125\}$. We make 20 runs with random $\alpha$ and show the mean of this runs. Next in graphics, the left side will be experiments for projected gradient and the right side will be experiments for accelerated gradient descent. The practical convergence rate for the objective value of the dual problem (D) is shown in log scale in Figure 4. The convergence rate for the constraint feasibility is shown in log scale in Figure 5. The convergence rate for the value of the duality gap is shown in log scale in Figure 6.

These plots show that the theoretical dependence on $\mu$ in both theorems is correct and occurs in practice.

**Conclusion**

In this paper, we considered the resource allocation problem as a convex optimization problem. To solve this problem, we use and interpret gradient and accelerated gradient descent applied to the special variant of dual problem. We obtain the convergence rate both for the primal iterates and the dual iterates. We obtain faster convergence rates than the ones known in the literature (for interpretable methods) and generalize our methods for the case of multiple producers and products. We point that a close approach, but with a sort of stochastic gradient methods can be used for a close problem of resource allocation in communication networks with large number of users [14,19,27].

**Acknowledgements**

We would like to thank Yu. Nesterov for fruitful discussions.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Funding**

The work of A.V. Gasnikov and P.E. Dyurechensky in the first part of the paper was supported by Russian Foundation for Fundamental Investigations (RFBR) grant 18-29-03071 mk. The work of
A.V. Gasnikov in the second part of the paper partially supported by Russian Foundation for Basic Research (RFBR) 18-31-20005 mol-a-ved. The work of A.V. Gasnikov and A.S. Ivanova in Appendix was prepared within the framework of the HSE University Basic Research Program and funded by the Russian Academic Excellence Project ‘5–100’.

**Notes on contributors**

**Anastasia Ivanova** got BSc degree from the Faculty of Control and Applied Mathematics of Moscow Institute of Physics and Technology (MIPT) in 2018. Since 2018 a master student in MIPT.

**Pavel Dvurechensky** got BSc and MSc degrees from the Faculty of Control and Applied Mathematics of Moscow Institute of Physics and Technology in 2008 and 2010, respectively. He obtained a PhD from the same university in 2014. Since 2015 he has been a research associate at Weierstrass Institute for Applied Analysis and Stochastics in Berlin.

**Alexander V. Gasnikov** received the BSc, MSc, PhD, Dr habil. degrees from the Department of Control and Applied Mathematics of Moscow Institute of Physics and Technology (Russia) in 2004, 2006, 2007, and 2016, respectively. Since 2011, he has been an Associate Professor with the Department of Mathematical Foundations of Control, MIPT. Also, since 2015, he has been the Lead Researcher with the Institute for Information Transmission Problems of Russian Academy of Science, Moscow, Russia, and an Associate Professor with the Faculty of Computer Science High school of economics.

**Dmitry Kamzolov** got BSc and MSc degrees from the Faculty of Control and Applied Mathematics of Moscow Institute of Physics and Technology (MIPT) in 2016 and 2018, respectively. Also, he obtained an MSc in Operations Research, Combinatorics and Optimization from the University of Grenoble Alpes. Since 2018 he has been pursuing a PhD in MIPT.

**ORCID**

Pavel Dvurechensky [http://orcid.org/0000-0003-1201-2343](http://orcid.org/0000-0003-1201-2343)

Alexander Gasnikov [http://orcid.org/0000-0002-7386-039X](http://orcid.org/0000-0002-7386-039X)

and Dmitry Kamzolov [http://orcid.org/0000-0001-8488-9692](http://orcid.org/0000-0001-8488-9692)

**References**

[1] A.S. Anikin, A.V. Gasnikov, P.E. Dvurechensky, A.I. Tyurin and A.V. Chernov, *Dual approaches to the minimization of strongly convex functionals with a simple structure under affine constraints*, Comput. Math. Math. Phys. 57 (2017), pp. 1262–1276.

[2] K.J. Arrow and L. Hurwicz, *Decentralization and computation in resource allocation*, Stanford University, Department of Economics; 1958.

[3] J.P. Aubin, *L’analyse non-linéaire et ses motivations économiques*, Masson, Paris, 1984.

[4] D.R. Baimurzina, A.V. Gasnikov, E.V. Gasnikova, P.E. Dvurechensky, E.I. Ershov, M.B. Kuben-taeva and A.A. Lagunovskaya, *Universal method of searching for equilibria and stochastic equilibria in transportation networks*, Comput. Math. Math. Phys. 59 (2019), pp. 19–33. arXiv:1701.02473.

[5] A. Bayandina, P. Dvurechensky, A. Gasnikov, F. Stonyakin and A. Titov, *Mirror descent and convex optimization problems with non-smooth inequality constraints*, in *Large-Scale and Distributed Optimization*, P. Giselsson and A. Rantzer, eds., chap. 8, Springer International Publishing, 2018, pp. 181–215, arXiv:1710.06612.

[6] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn. 3 (2011), pp. 1–122. [http://dx.doi.org/10.1561/2200000016](http://dx.doi.org/10.1561/2200000016).

[7] D.E. Campbell, *Resource allocation mechanisms*, Cambridge University Press, Cambridge, 1987.
A. Chernov, P. Dvurechensky and A. Gasnikov, *Fast Primal-Dual Gradient Method for Strongly Convex Minimization Problems with Linear Constraints*, Discrete Optimization and Operations Research: 9th International Conference, DOOR 2016, Vladivostok, Russia, September 19–23, 2016, Proceedings, Springer International Publishing, 2016, pp. 391–403.

J.C. Duchi, A. Agarwal and M.J. Wainwright, *Dual averaging for distributed optimization: Convergence analysis and network scaling*, IEEE. Trans. Automat. Contr. 57 (2012), pp. 592–606.

D. Dvinskikh, E. Gorbunov, A. Gasnikov, P. Dvurechensky and C.A. Uribe, *On Primal–Dual Approach for Distributed Stochastic Convex Optimization over Networks*, 2019 IEEE Conference on Decision and Control (CDC), arXiv:1903.09844 2019.

P. Dvurechensky, A. Gasnikov and A. Kroshnin, *Computational Optimal Transport: Complexity by Accelerated Gradient Descent Is Better Than by Sinkhorn’s Algorithm*, Proceedings of the 35th International Conference on Machine Learning, Proceedings of Machine Learning Research, Vol. 80, arXiv:1802.04367, 2018, pp. 1367–1376.

P. Dvurechensky, A. Gasnikov, E. Gasnikova, S. Matsievsky, A. Rodomanov and I. Usik, *Primal-Dual Method for Searching Equilibrium in Hierarchical Congestion Population Games*, Supplementary Proceedings of the 9th International Conference on Discrete Optimization and Operations Research and Scientific School (DOOR 2016) Vladivostok, Russia, September 19–23, 2016, arXiv:1606.08988, 2016, pp. 584–595.

P. Dvurechensky, D. Dvinskikh, A. Gasnikov, C.A. Uribe and A. Nedić, *Decentralize and Randomize: Faster Algorithm for Wasserstein Barycenters*, Advances in Neural Information Processing Systems 31, NeurIPS 2018, arXiv:1806.03915, Curran Associates, Inc., 2018, pp. 10783–10793.

P.E. Dvurechensky, A.V. Gasnikov and A.A. Lagunovskaya, *Parallel algorithms and probability of large deviation for stochastic convex optimization problems*, Numer. Anal. Appl. 11 (2018), pp. 33–37. arXiv:1701.01830.

E.J. Friedman and S.S. Oren, *The complexity of resource allocation and price mechanisms under bounded rationality*, Economic Theory 6 (1995), pp. 225–250.

A. Gasnikov, *Universal gradient descent*, arXiv preprint arXiv:1711.00394 (2017).

S.V. Guminov, Y.E. Nesterov, P.E. Dvurechensky and A.V. Gasnikov, *Accelerated primal-dual gradient descent with linesearch for convex, nonconvex, and nonsmooth optimization problems*, Doklady Mathematics 99 (2019), pp. 125–128.

A. Ivanova, A. Gasnikov, E. Nurminski and E. Vorontsova, *Walrasian equilibrium and centralized distributed optimization from the point of view of modern convex optimization methods on the example of resource allocation problem*, Siberian J. Num. Math. 22 (2019), pp. 411–432. Sib. Branch of Russ. Acad. of Sci. Novosibirsk.

A. Ivanova, D. Pasechnyuk, P. Dvurechensky, A. Gasnikov and E. Vorontsova, *Numerical methods for the resource allocation problem in networks*, Comput. Math. Math. Phys. (2020), arXiv:1909.13321.

D. Jakovetić, J. Xavier and J.M.F. Moura, *Fast distributed gradient methods*, IEEE. Trans. Automat. Contr. 59 (2014), pp. 1131–1146.

A. Kakhbod, *Resource allocation in decentralized systems with strategic agents: an implementation theory approach*, Springer Science & Business Media, New York, 2013.

A. Kroshnin, N. Tupitsa, D. Dvinskikh, P. Dvurechensky, A. Gasnikov and C. Uribe, *On the Complexity of Approximating Wasserstein Barycenters*, Proceedings of the 36th International Conference on Machine Learning, Proceedings of Machine Learning Research, Vol. 97, 09–15 Jun, arXiv:1901.08686, PMLR, Long Beach, CA, 2019, pp. 3530–3540.

A. Nedic and A. Ozdaglar, *Distributed subgradient methods for multi-agent optimization*, IEEE. Trans. Automat. Contr. 54 (2009), pp. 48–61.

Y. Nesterov, *Smooth minimization of non-smooth functions*, Math. Program. 103 (2005), pp. 127–152.

Y. Nesterov and V. Shikhman, *Dual subgradient method with averaging for optimal resource allocation*, Eur. J. Oper. Res. 270 (2018), pp. 907–916.

B. Polyak, *A general method of solving extremum problems*, Soviet Math. Doklady 8 (1967), pp. 593–597.
Appendix

Proof of Lemma 3.1: Set \( \tilde{x}_1 = \cdots = \tilde{x}_n = 2C/n \) and \( \tilde{y}_1 = \cdots = \tilde{y}_n = C/n \). And define \( \tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_n)^\top \) and \( \tilde{y} = (\tilde{y}_1, \ldots, \tilde{y}_n)^\top \). Notice that the point \( (\tilde{x}^\top, \tilde{y}^\top)^\top \) satisfies the Slater’s condition as

\[
\frac{2C}{n} = \tilde{x}_k > \tilde{y}_k = \frac{C}{n}, \quad k = 1, \ldots, n.
\]

Because the cost functions \( f_k(x_k), k = 1, \ldots, n \) are increasing due to the economic interpretation, we obtain

\[
\sum_{k=1}^{n} f_k(0) = \min_{\sum_{k=1}^{n} y_k \geq C, \ y_k \geq 0, \ x_k \geq 0, \ k = 1, \ldots, n} \sum_{k=1}^{n} f_k(x_k) = \min_{\sum_{k=1}^{n} y_k \geq C, \ y_k \geq 0, \ x_k \geq 0, \ k = 1, \ldots, n} \left\{ \sum_{k=1}^{n} f_k(x_k) + \sum_{k=1}^{n} (y_k - x_k) p_k \right\} = 0
\]

\[
\leq \max_{p \geq 0} \sum_{k=1}^{n} f_k(x_k) + \sum_{k=1}^{n} (y_k - x_k) p_k
\]

\[
= \min_{\sum_{k=1}^{n} y_k \geq C, \ y_k \geq 0, \ x_k \geq 0, \ k = 1, \ldots, n} \left\{ \sum_{k=1}^{n} f_k(x_k) + \sum_{k=1}^{n} (y_k - x_k) p_k^* \right\}
\]

\[
\leq \sum_{k=1}^{n} f_k(\tilde{x}_k) + \sum_{k=1}^{n} (\tilde{y}_k - \tilde{x}_k) p_k^* \leq \sum_{k=1}^{n} f_k(\tilde{x}_k) - \frac{C}{n} \sum_{k=1}^{n} p_k^*.
\]

Since \( p_k^* \geq 0, k = 1, \ldots, n \), we obtain that

\[
\|p^*\|_1 \leq p_{max},
\]

where

\[
p_{max} = \frac{n}{C} \left( \sum_{k=1}^{n} f_k \left( \frac{2C}{n} \right) - \sum_{k=1}^{n} f_k(0) \right).
\]

And since that we obtain the upper bound for each component

\[
p_k^* \leq p_{max}, \quad k = 1, \ldots, n,
\]

from which we obtain the statement of the Lemma.

Proof of Lemma 4.1: Using (8), we can rewrite the step as

\[
p^{t+1} = \arg \min_{p \geq 0} \left\{ C \max_{k=1, \ldots, n} (-p_k) + \langle x(p^t), p - p^t \rangle + \frac{L}{2} \left( \|p\|^2_2 - 2\langle p, p^t \rangle + \|p^t\|^2_2 \right) \right\}
\]
Let us define \( \gamma = C/L_\psi \). Then to determine \( \tilde{p}^{t+1} \) it is necessary to solve the following problem:

\[
\max_{k=1, ..., n} (-p_k) + \frac{1}{2\gamma} \|p - \tilde{p}^{t+1}\|_2^2 \quad \rightarrow \quad \min_{k=1, ..., n} \quad \eta_{t+1} \leq 0, \eta_{t+1} \geq -p_k, \quad k = 1, ..., n,
\]

which we can rewrite as the following equivalent problem:

\[
\eta_{t+1} + \frac{1}{2\gamma} \|p - \tilde{p}^{t+1}\|_2^2 \quad \rightarrow \quad \min_{\eta_{t+1} \leq 0, \eta_{t+1} \geq -p_k, \quad k = 1, ..., n},
\]

where \( \eta_{t+1} \in \mathbb{R} \). The Lagrangian is

\[
L(p, \eta_{t+1}, z, w) = \eta_{t+1} + \frac{1}{2\gamma} \|p - \tilde{p}^{t+1}\|_2^2 + \sum_{k=1}^{n} z_k(-p_k - \eta_{t+1}) + w\eta_{t+1}
\]

with dual variable \( z \in \mathbb{R}^+_n \) and \( w \in \mathbb{R}_+ \). Let \( (p^*, \eta^*_{t+1}, z^*, w^*) \) be a solution, then the optimality conditions are

\[
\begin{align*}
-p_k^* & \leq \eta_{t+1}^*, \quad k = 1, \ldots, n. \quad \text{(A1)} \\
z_k^* & \geq 0, \quad k = 1, \ldots, n. \quad \text{(A2)} \\
z_k^*(p_k^* + \eta_{t+1}^*) & = 0, \quad k = 1, \ldots, n. \quad \text{(A3)} \\
 \frac{1}{\gamma}(p_k^* - \tilde{p}_k^{t+1}) - z_k^* & = 0, \quad k = 1, \ldots, n. \quad \text{(A4)} \\
\sum_{k=1}^{n} z_k^* & = 1 + w^*. \quad \text{(A5)} \\
w^*\eta_{t+1}^* & = 0. \quad \text{(A6)} \\
w^* & \geq 0. \quad \text{(A7)} \\
\eta_{t+1}^* & \leq 0. \quad \text{(A8)}
\end{align*}
\]

- If \( \eta_{t+1}^* = 0 \).
  1. If \( \eta_{t+1}^* = -p_k^* = 0 \), then from (A4) and (A2) obtain, that
     \[
     z_k^* = \frac{1}{\gamma}(-\tilde{p}_k^{t+1}) = \frac{1}{\gamma}(-\tilde{p}_k^{t+1})_+.
     \]
  2. If \( p_k^* > -\eta_{t+1}^* = 0 \), then from (A3) and (A4)
     \[
     z_k^* = 0, \quad p_k^* = \tilde{p}_k^{t+1}.
     \]

Also since in this case \( p_k^* > 0 \), then \( \tilde{p}_k^{t+1} > 0 \). And using this we obtain

\[
z_k^* = 0 = \frac{1}{\gamma}(-\tilde{p}_k^{t+1})_+.
\]

Then, using (A9), (A10), (A5) and (A7) obtain that

\[
\sum_{k=1}^{n}(-\tilde{p}_k^{t+1})_+ = \gamma (1 + w^*) \geq \gamma.
\]
As a result, we obtain that for \( \sum_{k=1}^{n} (-\tilde{p}_k^{t+1})_+ \geq \gamma \) the case then \( \eta_{t+1}^* = 0 \) and solution determines as
\[
p_k^* = \max\{0, \tilde{p}_k^{t+1}\}.
\]

- If \( \eta_{t+1}^* < 0 \), then \( w^* = 0 \).
- If \( p_k^* = -\eta_{t+1}^* \), then from (A4) obtain
\[
z_k^* = \frac{1}{\gamma}(-\eta_{t+1}^* - \tilde{p}_k^{t+1})_+.
\]

(1) If \( p_k^* > -\eta_{t+1}^* \), then from (A3) and (A4) we obtain that \( z_k^* = 0, p_k^* = \tilde{p}_k^{t+1} \). Also since in this case \( \tilde{p}_k^{t+1} = p_k^* > -\eta_{t+1}^* \), then
\[
z_k^* = \frac{1}{\gamma}(-\eta_{t+1}^* - \tilde{p}_k^{t+1})_+.
\]

Finally using (A11), (A12) and (A5) obtain that
\[
\sum_{k=1}^{n} (-\eta_{t+1}^* - \tilde{p}_k^{t+1})_+ = \gamma
\]
and solution determines as
\[
p_k^* = \max\{-\eta_{t+1}^*, \tilde{p}_k^{t+1}\}.
\]

So, define \( p_{\text{center}}^{t+1} = -\eta_{t+1}^* \) we finishes the proof.

**Proof of Theorem 4.1:** Let's consider the step
\[
p^{t+1} = \arg\min_{p \geq 0} \left\{ \langle \nabla \psi(p'), p - p' \rangle - C \min_{k=1,...,m} p_k + \frac{L_{\psi}}{2} \|p - p'\|^2 \right\}
\]
\[
\Phi(p,p')
\]
The function \( \Phi(p,p') \) is convex with respect to the variable \( p \), then from the definition of \( p^{t+1} \) follows that
\[
\langle \nabla_{p^{t+1}} \Phi(p^{t+1}, p'), p - p^{t+1} \rangle \geq 0, \quad \forall p \geq 0.
\]
Define function
\[
\tilde{\Phi}(p,p') = \langle \nabla \psi(p'), p - p' \rangle - C \min_{k=1,...,m} p_k = \langle \nabla \psi(p'), p - p' \rangle + \xi(p),
\]
which is convex with respect to the variable \( p \). From (A13), we obtain
\[
0 \leq \langle \nabla_{p^{t+1}} \Phi(p^{t+1}, p'), p - p^{t+1} \rangle = \left\{ \nabla_{p^{t+1}} \tilde{\Phi}(p^{t+1}, p') + L_{\psi}(p^{t+1} - p'), p - p^{t+1} \right\}
\]
\[
\leq \tilde{\Phi}(p, p') - \tilde{\Phi}(p^{t+1}, p') + L_{\psi}(p^{t+1} - p', p - p^{t+1}).
\]
Note that
\[
L_{\psi}(p^{t+1} - p', p - p^{t+1}) = L_{\psi}(p^{t+1}, p) + L_{\psi}(p', p^{t+1}) - L_{\psi}(p^{t+1}, p^{t+1}) - L_{\psi}(p', p)
\]
\[
= \frac{L_{\psi}}{2} (p, p) - L_{\psi}(p', p) + \frac{L_{\psi}}{2} (p', p') - \frac{L_{\psi}}{2} (p, p) + L_{\psi}(p, p^{t+1})
\]
\[
- \frac{L_{\psi}}{2} (p^{t+1}, p^{t+1}) - \frac{L_{\psi}}{2} (p^{t+1}, p^{t+1}) - \frac{L_{\psi}}{2} (p, p') + L_{\psi}(p, p^{t+1})
\]
\[
= \frac{L_{\psi}}{2} \|p - p'\|^2 - \frac{L_{\psi}}{2} \|p - p^{t+1}\|^2 - \frac{L_{\psi}}{2} \|p^{t+1} - p'\|^2.
\]

(A14)
Substituting this inequality to (A14), we obtain
\[
0 \leq \tilde{\Phi}(p, p') - \tilde{\Phi}(p^{t+1}, p') + \frac{L\psi}{2} \|p - p'\|^2 - \frac{L\psi}{2} \|p - p^{t+1}\|^2 - \frac{L\psi}{2} \|p^{t+1} - p'\|^2 \\
= \langle \nabla \psi(p'), p - p' \rangle + \xi(p) - \langle \nabla \psi(p'), p^{t+1} - p' \rangle - \xi(p^{t+1}) + \frac{L\psi}{2} \|p - p'\|^2 \\
- \frac{L\psi}{2} \|p - p^{t+1}\|^2 - \frac{L\psi}{2} \|p^{t+1} - p'\|^2.
\]

Finally, since (10) we can rewrite as
\[
\psi(p^1) \leq \psi(p^2) + \langle \nabla \psi(p^2), p^1 - p^2 \rangle + \frac{L\psi}{2} \|p^1 - p^2\|^2, \forall p^1, p^2 \geq 0, \quad (A16)
\]
we obtain that
\[
\psi(p^{t+1}) \leq \psi(p^t) + \langle \nabla \psi(p^t), p - p^t \rangle + \xi(p) - \langle \nabla \psi(p^t), p^{t+1} - p^t \rangle - \xi(p^{t+1}) + \frac{L\psi}{2} \|p - p^t\|^2 - \frac{L\psi}{2} \|p - p^{t+1}\|^2
\]
or
\[
\varphi(p^{t+1}) \leq \varphi(p^t) + \langle \nabla \psi(p^t), p - p^t \rangle - \xi(p^t) + \xi(p) + \frac{L\psi}{2} \|p - p^t\|^2 - \frac{L\psi}{2} \|p - p^{t+1}\|^2.
\]
Summing these inequalities from \( t = 0 \) to \( t = N-1 \) we get, using the convexity of \( \psi \) and that it is true for all \( p \geq 0 \)
\[
\varphi(p^N) \leq \frac{1}{N} \min_{p \geq 0} \left\{ \sum_{t=0}^{N-1} \left[ \varphi(p^t) + \langle \nabla \psi(p^t), p - p^t \rangle - \xi(p^t) + \xi(p) + \frac{L\psi}{2} \|p - p^0\|^2 \right] \right\}, \quad (A17)
\]
where \( p^N = 1/N \sum_{i=1}^{N} p^i \). Let us define the starting point \( p^0 \) satisfying \( 0 \leq p^0_k \leq p_{\text{max}}, k = 1, \ldots, n \), where \( p_{\text{max}} \) is given in (6). Then we obtain that
\[
\|p^0\|_2 \leq \sqrt{n}p_{\text{max}}.
\]
Let us introduce a set
\[
B_{2R}^+(0) = \{ p : p \geq 0, \|p - p^0\|_2 \leq 3R \},
\]
where
\[
\|p^0 - p^*\|_2 + \|p^0\|_2 \leq 2\|p^0\|_2 + \|p^*\|_2 = 3p_{\text{max}}\sqrt{n} = R, \quad (A18)
\]
herewith all the obtaining \( p^t \) will consist in \( B_{2R}^+(0) : \)
\[
\|p^t\|_2 \leq 2R, \quad (A19)
\]
since (second paragraph [16])
\[
\|p^t\|_2 = \|p^t - p^0\|_2 + \|p^0\|_2 \leq \|p^t - p^*\|_2 + \|p^* - p^0\|_2 + \|p^0\|_2 \leq 2\|p^* - p^0\|_2 + \|p^0\|_2 = \|p^*\|_2 + 3\|p^0\|_2 = 5p_{\text{max}}\sqrt{n} \leq 2R.
\]
Since that we can rewrite (A17) as
\[
\varphi(p^N) \leq \frac{1}{N} \min_{p \geq 0} \left\{ \sum_{t=0}^{N-1} \left[ \varphi(p^t) + \langle \nabla \psi(p^t), p - p^t \rangle - \xi(p^t) + \xi(p) + \frac{L\psi}{2} \|p - p^0\|^2 \right] \right\}
\]
\[
\leq \frac{1}{N} \min_{p \in B_{2R}^+(0)} \left\{ \sum_{t=0}^{N-1} \left[ \psi(p^t) + \langle \nabla \psi(p^t), p - p^t \rangle + \xi(p) + \frac{L\psi}{2} \|p - p^0\|^2 \right] \right\}.
\]
Notice that the minimum is taken over the set $B_{3R}^+ (0) = B_{3R}^+ (p^0)$ and not over the set $B_{3R}^+ (p^*)$, since the $p^*$ is not known and, according to the definition of $R$, $p^* \in B_{3R}^+ (p^0)$. Define

$$\tilde{p} = \arg \min_{p \in B_{3R}^+ (0)} \left\{ \sum_{t=0}^{N-1} \left[ \psi(p') + \{\nabla\psi(p'), p - p'\} \right] + \xi(p) \right\},$$

then since

$$\| \tilde{p} - p^0 \|_2^2 \leq 2\| \tilde{p} \|_2^2 + 2\| p^0 \|_2^2 \leq 18R^2 + \frac{2}{9}R^2 = \frac{164}{9}R^2,$$

we obtain

$$\varphi(p^N) \leq \frac{1}{N} \min_{p \in B_{3R}^+ (0)} \left\{ \sum_{t=0}^{N-1} \left[ \psi(p') + \{\nabla\psi(p'), p - p'\} \right] + \xi(p) + \frac{L_\psi}{2} \| p - p^0 \|_2^2 \right\} \leq \frac{1}{N} \min_{p \in B_{3R}^+ (0)} \left\{ \sum_{t=0}^{N-1} \left[ \psi(p') + \{\nabla\psi(p'), p - p'\} \right] + \xi(p) \right\} + \frac{L_\psi}{2N} \| \tilde{p} - p^0 \|_2^2 \leq \frac{1}{N} \min_{p \in B_{3R}^+ (0)} \left\{ \sum_{t=0}^{N-1} \left[ \psi(p') + \{\nabla\psi(p'), p - p'\} \right] + \xi(p) \right\} + \frac{82L_\psi R^2}{9N}.$$  

Notice that

$$- \frac{1}{N} \min_{p \in B_{3R}^+ (0)} \left\{ \sum_{t=0}^{N-1} \left[ \psi(p') + \{\nabla\psi(p'), p - p'\} \right] + \xi(p) \right\} \leq \frac{1}{N} \max_{p \in B_{3R}^+ (0)} \left\{ \sum_{t=0}^{N-1} \left[ x(p), p' - p \right] + \sum_{k=1}^{n} f_k(x_k(p')) - \left\{ x(p'), p' \right\} \right\} - \xi(p) \leq \frac{1}{N} \sum_{t=0}^{N-1} f(x(p')) + \frac{1}{N} \max_{p \in B_{3R}^+ (0)} \left\{ \sum_{t=0}^{N-1} \left\{ -x(p'), p \right\} - \xi(p) \right\} \converges \iff f(x^N) + \frac{1}{N} \max_{p \in B_{3R}^+ (0)} \left\{ \sum_{t=0}^{N-1} \left\{ -x(p'), p \right\} - \xi(p) \right\} = f(x^N) + \max_{p \in B_{3R}^+ (0)} \left\{ -\left\{ x^N, p \right\} + C \min_{k=1, \ldots, n} p_k \right\},$$

where $x^N = 1/N \sum_{t=0}^{N-1} x(p')$. Since

$$C \min_{k=1, \ldots, n} p_k - \sum_{k=1}^{n} x_k^N \min_{k=1, \ldots, n} p_k \geq C \min_{k=1, \ldots, n} p_k - \sum_{k=1}^{n} x_k^N p_k,$$

we obtain

$$\max_{p \in B_{3R}^+ (0)} \left\{ -\left\{ x^N, p \right\} + C \min_{k=1, \ldots, n} p_k \right\} = \max_{p \in B_{3R}^+ (0)} \left\{ -\left\{ x^N, p \right\} + C \min_{k=1, \ldots, n} p_k \right\} = \max_{p \in B_{3R}^+ (0)} \left\{ C p - \sum_{k=1}^{n} x_k^N p_k \right\} = \frac{3R}{\sqrt{n}} \left[ C - \sum_{k=1}^{n} x_k^N \right].$$
Thus we obtain the following estimation:
\[
\varphi(p^N) + f(x^N) + \frac{3R}{\sqrt{n}} \left[ C - \sum_{k=1}^{n} x_k^N \right]_+ \leq \frac{82L_\psi R^2}{9N}. \tag{A20}
\]

Considering the weak duality \(-f(x^*) \leq \varphi(p^*)\), obtain
\[
f(x^N) - f(x^*) \leq f(x^N) + \varphi(p^*) \leq \varphi(p^N) + f(x^N)
\]
\[
\leq \varphi(p^N) + f(x^N) + \frac{3R}{\sqrt{n}} \left[ C - \sum_{k=1}^{n} x_k^N \right]_+ \leq \frac{82L_\psi R^2}{9N} = \epsilon.
\]

Using \((A19), (A20)\) and \((D)\), we get
\[
\frac{R}{\sqrt{n}} \left[ C - \sum_{k=1}^{n} x_k^N \right]_+ \leq \left\{ \langle \nabla \psi(p^{t+1}), p - p^{t+1} \rangle - \min_{k=1, \ldots, n} p_k \right\} + \frac{1}{2} \| p - y^t \|_2.
\]
\[
\langle \nabla \psi(p^{t+1}), p - p^{t+1} \rangle - \min_{k=1, \ldots, n} p_k \leq \frac{2R}{\sqrt{n}} [C - \sum_{k=1}^{n} x_k^N]_+.
\]
\[
And since \(R = 3p_{\max}\sqrt{n} \) and \(L_\psi = n/\mu\), we obtain the statement of the theorem. \(\blacksquare\)

**Proof of Theorem 5.1:** Let's consider the step
\[
y^{t+1} = \arg\min_{p \geq 0} \left\{ \alpha_{t+1} \left( \langle \nabla \psi(p^{t+1}), p - p^{t+1} \rangle - \min_{k=1, \ldots, n} p_k \right) + \frac{1}{2} \| p - y^t \|_2 \right\}.
\]
The function \(\Psi(p, p^t)\) is convex with respect to the variable \(p\), then from the definition of \(y^{t+1}\) follows that
\[
\langle \nabla \psi(y^{t+1}, p^{t+1}), p - y^{t+1} \rangle \geq 0, \quad \forall p \geq 0. \tag{A22}
\]
Define function
\[
\tilde{\Psi}(p, p^t) = \langle \nabla \psi(p^t), p - p^{t+1} \rangle - \min_{k=1, \ldots, n} p_k = \langle \nabla \psi(p^t), p - p^{t+1} \rangle + \xi(p),
\]
which is convex with respect to the variable \(p\). From \((A22)\), we obtain
\[
0 \leq \langle \nabla \psi(y^{t+1}, p^{t+1}), p - y^{t+1} \rangle = \langle \alpha_{t+1} \nabla \psi(y^{t+1}, p^{t+1}) + (y^{t+1} - y^t), p - y^{t+1} \rangle
\]
\[
\leq \alpha_{t+1} \nabla \tilde{\psi}(p, p^{t+1}) - \alpha_{t+1} \tilde{\psi}(y^{t+1}, p^{t+1}) + \langle y^{t+1} - y^t, p - y^{t+1} \rangle. \tag{A23}
\]
Using \((A15)\), we obtain
\[
0 \leq \alpha_{t+1} \tilde{\psi}(p, p^{t+1}) - \alpha_{t+1} \Phi(y^{t+1}, p^{t+1}) + \frac{1}{2} \| p - y^t \|_2^2 - \frac{1}{2} \| p - y^{t+1} \|_2^2 - \frac{1}{2} \| y^{t+1} - y^t \|_2^2
\]
\[
= \alpha_{t+1} \langle \nabla \psi(p^{t+1}), p - p^{t+1} \rangle + \alpha_{t+1} \xi(p) - \alpha_{t+1} \left( \langle \nabla \psi(p^{t+1}), y^{t+1} - p^{t+1} \rangle + \xi(y^{t+1}) \right)
\]
\[
+ \frac{1}{2} \| p - y^t \|_2^2 - \frac{1}{2} \| p - y^{t+1} \|_2^2 - \frac{1}{2} \| y^{t+1} - y^t \|_2^2.
\]
Notice that
\[
- \alpha_{t+1} \left( \langle \nabla \psi(p^{t+1}), y^{t+1} - p^{t+1} \rangle + \xi(y^{t+1}) \right) - \frac{1}{2} \| y^{t+1} - y^t \|_2^2
\]
\[
= \alpha_{t+1} \left( \langle \nabla \psi(p^{t+1}), y^{t+1} - y^t \rangle + \langle \nabla \psi(p^{t+1}), y^t - p^{t+1} \rangle - \xi(y^{t+1}) \right) - \frac{1}{2} \| y^{t+1} - y^t \|_2^2.
(16)\(A_t\left(\nabla:\varphi^t,\varphi - \varphi^t\right) - \alpha t_1\xi(y^{t+1}) + A_t_1\left(\nabla:\varphi^t,\varphi - \varphi^t\right)
\]

\[\frac{A^2}{2\alpha t_1}||\varphi^{t+1} - \varphi^t||_2^2\]

(15) \[A_t\left(\nabla:\varphi^t,\varphi - \varphi^t\right) - \alpha t_1\xi(y^{t+1})
\]

\[+ A_t_1\left(\nabla:\varphi^t,\varphi - \varphi^t\right) - \frac{L}{2}||\varphi^{t+1} - \varphi^t||_2^2\]

Finally, using this, we obtain that

\[A_t_1\varphi^{t+1} - A_t\varphi^t \leq \alpha t_1\left(\psi(\varphi^{t+1}) + \nabla:\varphi^t,\varphi - \varphi^t + \xi(\varphi)\right).
\]

Sum all the inequalities for \(t = 0, \ldots, N - 1\). Then, for any \(p \geq 0\)

\[A_N\varphi(w^N) - A_0\varphi(w^0) \leq \sum_{t=0}^{N-1} \alpha t_1\left(\psi(\varphi^{t+1}) + \nabla:\varphi^t,\varphi - \varphi^t + \xi(\varphi)\right).
\]

Whence, since \(C_0 = \alpha_0 = 0\)

\[A_N\varphi(w^N) \leq \sum_{t=0}^{N} \alpha t_1\left(\psi(\varphi^t) + \nabla:\varphi^t,\varphi - \varphi^t + \xi(\varphi)\right) + \frac{1}{2}||\varphi - y^0||_2^2.
\]

Taking in the right-hand side the minimum in \(p \geq 0\) we obtain

\[A_N\varphi(w^N) \leq \min_{p \geq 0} \left\{ \sum_{t=0}^{N} \alpha t_1\left(\psi(\varphi^t) + \nabla:\varphi^t,\varphi - \varphi^t + \xi(\varphi)\right) + \frac{1}{2}||\varphi - y^0||_2^2 \right\}.
\]

Put the vector \(y^0 = (y^0_1, \ldots, y^0_n)\)^T, where \(\varphi^0\) – vector of initial prices. And its components such as

\[0 < y^0_k < p_{max}, \text{ } k = 1, \ldots, n,
\]

then we obtain that

\[||y^0||_2 \leq \sqrt{n}p_{max}.
\]

Let us introduce a set \(B_{2R}(0) = \{p : p \geq 0, ||p||_2 < 2R\}\), where \(R\) determines from (A18).

Since (A19), we obtain

\[A_N\varphi(w^N) \leq \min_{p \in B_{2R}(0)} \left\{ \sum_{t=0}^{N} \alpha t_1\left(\psi(\varphi^t) + \nabla:\varphi^t,\varphi - \varphi^t + \xi(\varphi)\right) + \frac{1}{2}||\varphi - y^0||_2^2 \right\}.
\]
And since
\[ \|p - y^0\|_2^2 \leq 2\|y^0\|_2^2 + 2\|p\|_2^2 \leq 8R^2 + \frac{2}{9}R^2 = \frac{74}{9}R^2, \]
onobtain the following:
\[ A_N\varphi(w^N) \leq \min_{p \in \mathcal{B}_{2k}(0)} \left\{ \sum_{t=0}^{N} \alpha_t \left( \psi(p^t) + \langle \nabla \psi(p^t), p - p^t \rangle + \xi(p) \right) \right\} + \frac{37}{9}R^2. \quad (A25) \]

Notice that
\[ - \min_{p \in \mathcal{B}_{2k}(0)} \left\{ \sum_{t=0}^{N} \alpha_t \left( \psi(p^t) + \langle \nabla \psi(p^t), p - p^t \rangle + \xi(p) \right) \right\} \]
\[ = \max_{p \in \mathcal{B}_{2k}(0)} \left\{ \sum_{t=0}^{N} \alpha_t \left( \langle x(p^t), p^t - p \rangle - \xi(p) - \langle x(p^t), p^t \rangle + \sum_{k=1}^{n} f_k(x_k(p^t)) \right) \right\} \]
\[ = \sum_{t=0}^{N} \alpha_t f(x(p^t)) + \max_{p \in \mathcal{B}_{2k}(0)} \left\{ \sum_{t=0}^{N} \alpha_t \left( -\langle x(p^t), p \rangle - \xi(p) \right) \right\} \]
\[ \geq A_N f(x^N) + A_t \max_{p \in \mathcal{B}_{2k}(0)} \left\{ -\langle x^N, p \rangle + C \min_{k=1, \ldots, n} p_k \right\} \]
\[ = A_N f(x^N) + A_t \max_{p \in \mathcal{B}_{2k}(0)} \left\{ C_p - \sum_{k=1}^{n} x_k^N \right\} \]
\[ = A_N f(x^N) + A_t \frac{2R}{\sqrt{n}} \left[ C - \frac{2}{\sqrt{n}} \sum_{k=1}^{n} x_k^N \right] +, \]
where \( x^N = 1/A_N \sum_{t=0}^{N} \alpha_t x(p^t) \). Thus we obtain the following estimation:
\[ \varphi(w^N) + f(x^N) + \frac{2R}{\sqrt{n}} \left[ C - \frac{2}{\sqrt{n}} \sum_{k=1}^{n} x_k^N \right] + \leq \frac{37R^2}{9A_N}. \quad (A26) \]

Since \( w^* \) is an optimal solution of Problem (D), we have, for any \( x \geq 0 \)
\[ f(x^*) \leq f(x) - \langle x, w^* \rangle + C \min_{k=1, \ldots, n} w_k^*. \]

Using the estimation (5), we get
\[ f(x^N) - f(x^*) \geq -\frac{R}{3\sqrt{n}} \left[ C - \sum_{k=1}^{n} x_k^N \right] +. \]

Considering the weak duality \( f(x^*) \geq -\varphi(w^*) \), obtain
\[ f(x^N) + \varphi(w^N) = f(x^N) + f(x^*) - f(x^*) + \varphi(w^*) - \varphi(w^*) + \varphi(w^N) \]
\[ \geq f(x^N) - f(x^*) \geq -\frac{R}{3\sqrt{n}} \left[ C - \sum_{k=1}^{n} x_k^N \right] +. \]
This and (A26) give
\[
\left[ C - \sum_{k=1}^{n} x_k^N \right]_+ \leq \frac{37R\sqrt{n}}{15A_N}. \tag{A27}
\]

And also due to the weak duality, we obtain
\[
f(x^N) - f(x^*) \leq \frac{37R^2}{9A_N}. \tag{A28}
\]

Let us show that the lower bound for the sequence \( A_t, t \geq 0 \) is
\[
A_t \geq \frac{(t+1)^2}{4L_\psi}, \quad \forall t \geq 1, \tag{A29}
\]
where \( L_\psi \) is the Lipschitz constant for the gradient of \( \psi \).

For \( t = 1 \), since \( \alpha_0 = 0 \) and \( A_1 = \alpha_0 + \alpha_1 = \alpha_1 \), we have from (15)
\[
A_1 = \alpha_1 = \frac{1}{L_\psi} \geq \frac{1}{L_\psi}.
\]

Let us now assume that (A29) holds for some \( t \geq 1 \) and prove that it holds for \( t + 1 \). From (15), we have a quadratic equation for \( \alpha_{t+1} \)
\[
L_\psi \alpha_{t+1}^2 - \alpha_{t+1} - A_t = 0.
\]

Since we need to take the largest root, we obtain
\[
\alpha_{t+1} = \frac{1 + \sqrt{1 + 4L_\psi A_t}}{2L_\psi} = \frac{1}{2L_\psi} + \sqrt{\frac{1}{4L_\psi^2} + \frac{A_t}{L_\psi}} \geq \frac{1}{2L_\psi} + \sqrt{\frac{A_t}{L_\psi}}
\]
where we used the induction assumption that (A29) holds for \( t \). Using the obtained inequality, from (15) and (A29) for \( t \), we get
\[
A_{t+1} = A_t + \alpha_{t+1} \geq \frac{(t+1)^2}{4L_\psi} + \frac{t + 2}{2L_\psi} \geq \frac{(t+2)^2}{4L_\psi}.
\]

So, using this estimation and that \( R = 3p_{max}\sqrt{n} \) and \( L_\psi = \frac{n}{\mu} \), we obtain the statement of the theorem from (A27) and (A28).

\[\Box\]

### A.1 The resource allocation problem in vector case

In this section, we consider the vector case of resource allocation problem. Let us consider the same problem as the problem from Section 2. But now each producer produces \( m \) different products, having its own cost function \( f_k(x_k), k = 1, \ldots, n \) representing the total cost of production of the volume \( x_k \in \mathbb{R}^m \) which is the number of tons of products produced by the producer \( k \) in 1 year. So, we have production matrix \( X \in \mathbb{R}^{m \times n} \), where each row represents the vector of production for one product by each producer and each column represents the production for one producer of each product. Let us define \( x_{j \bullet} \) the \( j \) row of matrix and \( x_{\bullet k} \) the \( k \) column of matrix. The Center buys product from the producers and chooses its strategy in such a way that the total production volume per year by all producers is not less \( c_m \) tons of product \( m \). To do so, the Center needs to find \( y_{jk} \) – the volume of product \( j \) which is purchased from the producer \( k \). And in this case we can write down
the following resource allocation problem:

\[(P_1) \quad \sum_{k=1}^{n} f_k(x_{\k}) \rightarrow \min_{\sum_{k=1}^{n} y_k \geq c_j, j=1,\ldots,m;\quad x_{\k} \geq y_{\k} y_{\k} \geq 0, k=1,\ldots,n.} \]

where the cost functions \(f_k(x_{\k}) \) \(k = 1, \ldots, n\) are increasing for each variable \(x_{j,k}\) and \(\mu\)-strongly convex.

Introducing dual variables \(p_{\k}, k = 1, \ldots, n\) and using the duality, we obtain

\[\min_{\sum_{k=1}^{n} y_k \geq c_j, j=1,\ldots,m;\quad y_{\k} \geq 0, k=1,\ldots,n.} \left\{ \sum_{k=1}^{n} f_k(x_{\k}) + \sum_{k=1}^{n} \max_{p_{\k} \geq 0} p_{\k}^T (y_{\k} - x_{\k}) \right\} \]

\[= - \min_{\sum_{k=1}^{n} y_k \geq c_j, j=1,\ldots,m;\quad y_{\k} \geq 0, k=1,\ldots,n.} \left\{ \sum_{k=1}^{n} \max_{p_{\k} \geq 0} \left( p_{\k}^T x_{\k} - f_k(x_{\k}) \right) - \min_{\sum_{k=1}^{n} y_k \geq c_j;\quad y_{\k} \geq 0} \sum_{k=1}^{n} p_{\k}^T y_{\k} \right\} \]

\[= - \min_{\sum_{k=1}^{n} y_k \geq c_j, j=1,\ldots,m;\quad y_{\k} \geq 0, k=1,\ldots,n.} \left\{ \sum_{k=1}^{n} \left( p_{\k}^T x_{\k} (p_{\k}) - f_k(x_{\k}(p_{\k})) \right) - \sum_{j=1}^{m} c_j \min_{k=1,\ldots,n} p_{jk} \right\}, \]

where

\[x_{\k}(p_{\k}) = \arg\max_{x_{\k} \geq 0} \left\{ p_{\k}^T x_{\k} - f_k(x_{\k}) \right\}, \quad k = 1, 2, \ldots, n. \quad (A30)\]

Then the dual problem (up to a sign) has the following form:

\[\varphi(p_1, \ldots, p_n) = \sum_{k=1}^{n} \left\{ p_{\k}^T x_{\k}(p_{\k}) - f_k(x_{\k}(p_{\k})) \right\} - \sum_{j=1}^{m} c_j \min_{k=1,\ldots,n} p_{jk} \rightarrow \min_{\sum_{k=1}^{n} y_k \geq c_j;\quad y_{\k} \geq 0} \sum_{k=1}^{n} y_{\k} \geq 0. \quad (D1)\]

Let us consider this problem as the composite optimization problem. We can rewrite the dual problem as

\[\varphi(p_1, \ldots, p_n) = \psi(p_1, \ldots, p_n) + \xi(p_1, \ldots, p_n),\]

where

\[\psi(p_1, \ldots, p_n) = \sum_{k=1}^{n} \left\{ p_{\k}^T x_{\k}(p_{\k}) - f_k(x_{\k}(p_{\k})) \right\} \quad (A31)\]

is convex function, which gradient satisfying Lipschitz condition with constant \(L_\psi = \frac{n}{\mu}\). And convex non-smooth composite function

\[\xi(p_1, \ldots, p_n) = - \sum_{j=1}^{m} c_j \min_{k=1,\ldots,n} p_{jk}.\]

We define \(\tilde{x}_{\k} = (2c_1/n, \ldots, 2c_m/n)^T, k = 1, \ldots, n\) and \(\tilde{y}_{\k} = (c_1/n, \ldots, c_m/n)^T, k = 1, \ldots, n\), and similarly to Lemma 3.1, using the Slater's condition, we obtain the following Lemma.
Lemma A.1: Let the $\mathbf{p}_j^*, j = 1, \ldots, m$ be a solution to the dual problem (D1). Then it satisfies the inequality

$$\|\mathbf{p}_j^*\|_2 \leq \sqrt{n} \tilde{p}_{\text{max}},$$

where

$$\tilde{p}_{\text{max}} := \min_{\mathbf{p} \geq 0} \frac{n}{\sum c_j} \left( \sum_{k=1}^n f_k (\tilde{x}_k) - \sum_{k=1}^n f_k (0) \right).$$  \hfill (A32)

### A.1.1 Composite gradient method

In this section to solve the resource allocation problem in vector case, we use composite gradient method from Section 4. Let us define vectors $\tilde{\mathbf{y}} = (y_{\mathbf{1}}^T, \ldots, y_{\mathbf{n}}^T) \in \mathbb{R}^{mn}, \tilde{\mathbf{p}} = (\mathbf{p}_{\mathbf{1}}^T, \ldots, \mathbf{p}_{\mathbf{n}}^T) \in \mathbb{R}^{mn}$ and $\tilde{x}(\tilde{\mathbf{p}}) = (x_{\mathbf{1}}(\mathbf{p}_{\mathbf{1}})^T, \ldots, x_{\mathbf{n}}(\mathbf{p}_{\mathbf{n}})^T) \in \mathbb{R}^{mn}$, then the step of the method can be rewritten as follows:

$$\tilde{\mathbf{y}}^{t+1} = \arg\min_{\tilde{\mathbf{p}} \geq 0} \left\{ \langle \tilde{x}(\tilde{\mathbf{p}}^{t+1}), \tilde{\mathbf{p}} - \tilde{\mathbf{y}}^{t+1} \rangle + \frac{L_p}{2} \| \tilde{x} - \tilde{\mathbf{y}}^{t} \|_2^2 \right\}$$

$$= \arg\min_{\tilde{\mathbf{p}} \geq 0} \left\{ \sum_{k=1}^n \langle x_k(\mathbf{p}_{k}^{t+1}), \mathbf{p}_k - \mathbf{p}_k^{t+1} \rangle - \sum_{j=1}^m c_j \min_{k=1, \ldots, n} p_{jk} + \frac{L_p}{2} \| \mathbf{p}_k - y_k^{t+1} \|_2^2 \right\}$$

$$= \arg\min_{\tilde{\mathbf{p}} \geq 0} \left\{ \sum_{j=1}^m \left( \langle x_j(\mathbf{p}_j^{t+1}), \mathbf{p}_j - \mathbf{p}_j^{t+1} \rangle - c_j \min_{k=1, \ldots, n} p_{jk} + \frac{L_p}{2} \| \mathbf{p}_j - y_j^{t+1} \|_2^2 \right) \right\},$$

where $\mathbf{p}_j = (p_{j1}, \ldots, p_{jm})^T$, i.e. from the summation of the producers, we moved to the summation of products. Note that we can divide the step into $m$ independent problems for each product, so for product $j$, we have

$$y_j^{t+1} = \arg\min_{\mathbf{p}_j \geq 0} \left\{ \langle x_j(\mathbf{p}_j^{t+1}), \mathbf{p}_j - \mathbf{p}_j^{t+1} \rangle - c_j \min_{k=1, \ldots, n} p_{jk} + \frac{L_p}{2} \| \mathbf{p}_j - y_j^{t+1} \|_2^2 \right\},$$

where $\mathbf{p}_j$ is price vector of all producers for product $j$ and $x_j(\mathbf{p}_j)$ is vector of optimal plans for the production of product $j$ for all producers. Define $\tilde{p}_j^{t+1} = p_j^t - 1/L_p x_j(\mathbf{p}_j^{t+1})$ and then, using Lemma 4.1, we obtain the following solution:

- If $\sum_{k=1}^n (-p_{jk}^{t+1})_+ \geq c_j/L_p$, then $p_{j,\text{center}}^{t+1} = 0$.
- Else $p_{j,\text{center}}^{t+1} > 0$ is determined from

$$\sum_{k=1}^n \left( p_{j,\text{center}}^{t+1} - \tilde{p}_{jk}^{t+1} \right)_+ = \frac{c_j}{L_p},$$

and the solution is determined as

$$p_{jk}^{t+1} = \max\left(p_{j,\text{center}}^{t+1}, \tilde{p}_{jk}^{t+1}\right), \quad k = 1, \ldots, n.$$

Theorem A.1: Let Algorithm (A34) be run for $N$ steps with starting points $\mathbf{p}_j^0, j = 1, \ldots, m$, satisfying $0 \leq p_{jk}^0 \leq \tilde{p}_{\text{max}}, k = 1, \ldots, n$, where $\tilde{p}_{\text{max}}$ is given in (A32). Then

$$\sum_{k=1}^n f_k(\mathbf{x}_k^N) - \sum_{k=1}^n f_k(\mathbf{x}_k) \leq \sum_{k=1}^n f_k(\tilde{x}_k) + \varphi(\tilde{\mathbf{p}}^\top) \leq \varphi(\tilde{\mathbf{p}}) + \sum_{k=1}^n f_k(\mathbf{x}_k) \leq \frac{82 \tilde{p}_{\text{max}}^2 n^2 m}{N \mu},$$

where $\mu$ is the Lipschitz constant of $\varphi$. 

$\mathbf{D}$
\[
\sum_{j=1}^{m} \left[ c_j - \sum_{k=1}^{n} x_{jk}^N \right] \leq \frac{82 \tilde{p}_{\text{max}} n^2 m}{3N \mu},
\]

where \( p_{\text{st}}^N = 1/N \sum_{t=1}^{N} p_{\text{tst}}^N, \quad x_{\text{st}}^N = 1/N \sum_{t=0}^{N-1} x_{\text{st}}(p_{\text{st}}^N) \) and \( \tilde{p}^N = (p_{\text{st}}^N, \ldots, p_{\text{en}}^N) \).

**Proof of Theorem A.1:** Similarly to the proof of Theorem 4.1, we can obtain Equation (A17), but in this case we can write down this equation as

\[
\varphi(\tilde{p}^N) \leq \frac{1}{N} \min_{\tilde{p} \geq 0} \left\{ \sum_{t=0}^{N} \sum_{j=1}^{m} \left( x_{\text{st}}(p_{\text{tst}}^N), p_{\text{st}}^N \right) - \sum_{k=1}^{n} f_k(x_{\text{st}}) + \xi(\tilde{p}) + \frac{\tilde{L}_\psi}{2} \sum_{j=1}^{m} \|p_{\text{st}} - p_{\text{st}}^0\|^2_2 \right\},
\]

where \( \tilde{p}^N = (p_{\text{st}}^N, \ldots, p_{\text{en}}^N) \). Let us introduce a set \( B_{3R}^+ (0) = \{ p : p \geq 0, \|p\|_2 \leq 2\tilde{R} \} \), where \( \tilde{R} = 3\sqrt{n}p_{\text{max}} \) determines similarly to (A18). Since we also have equation similarly to (A19) we obtain that \( p_{\text{st}}^0 \in B_{3R}^+ (0), \forall t \). So, we obtain

\[
\varphi(\tilde{p}^N) \leq \frac{1}{N} \min_{\tilde{p}_{\text{st}} \in B_{3R}^+ (0)} \left\{ \sum_{t=0}^{N} \sum_{j=1}^{m} \left( x_{\text{st}}(p_{\text{tst}}^N), p_{\text{st}}^N \right) - \sum_{k=1}^{n} f_k(x_{\text{st}}) + \xi(\tilde{p}) + \frac{\tilde{L}_\psi}{2} \sum_{j=1}^{m} \|p_{\text{st}} - p_{\text{st}}^0\|^2_2 \right\}.
\]

And since

\[
\|p_{\text{st}} - p_{\text{st}}^0\|^2_2 \leq 2\|p_{\text{st}}^0\|^2_2 + 2\|p_{\text{st}}\|^2_2 \leq 18\tilde{R}^2 + \frac{2}{9} \tilde{R}^2 = \frac{164}{9} \tilde{R}^2
\]

obtain the following:

\[
\varphi(\tilde{p}^N) \leq \frac{1}{N} \min_{\tilde{p}_{\text{st}} \in B_{3R}^+ (0)} \left\{ \sum_{t=0}^{N} \sum_{j=1}^{m} \left( x_{\text{st}}(p_{\text{tst}}^N), p_{\text{st}}^N \right) - \sum_{k=1}^{n} f_k(x_{\text{st}}) + \xi(\tilde{p}) \right\} + \frac{82m\tilde{L}_\psi}{9N} \tilde{R}^2.
\]

Notice that

\[
-\frac{1}{N} \min_{\tilde{p}_{\text{st}} \in B_{3R}^+ (0)} \left\{ \sum_{t=0}^{N} \sum_{j=1}^{m} \left( x_{\text{st}}(p_{\text{tst}}^N), p_{\text{st}}^N \right) - \sum_{k=1}^{n} f_k(x_{\text{st}}) + \xi(\tilde{p}) \right\}
\]

\[
= \frac{1}{N} \sum_{t=0}^{N} \sum_{k=1}^{n} f_k(x_{\text{st}}) + \frac{1}{N} \max_{\tilde{p}_{\text{st}} \in B_{3R}^+ (0)} \left\{ \sum_{t=0}^{N} \sum_{j=1}^{m} \left( x_{\text{st}}(p_{\text{tst}}^N), p_{\text{st}}^N \right) + \sum_{j=1}^{m} c_j \min_{k=1, \ldots, n} p_{\text{jk}} \right\}
\]

\[
\geq \sum_{k=1}^{n} f_k(x_{\text{st}}^N) + \max_{\tilde{p}_{\text{st}} \in B_{3R}^+ (0)} \left\{ \sum_{j=1}^{m} \left( -x_{\text{jst}}^N, p_{\text{st}}^N \right) + \sum_{j=1}^{m} c_j \min_{k=1, \ldots, n} p_{\text{jk}} \right\}
\]

\[
= \sum_{k=1}^{n} f_k(x_{\text{st}}^N) + \max_{\tilde{p}_{\text{st}} \in B_{3R}^+ (0)} \left\{ \sum_{j=1}^{m} \left( -x_{\text{jst}}^N, p_{\text{st}}^N \right) + \sum_{j=1}^{m} c_j p_{\text{jk}} \right\}
\]

\[
= \sum_{k=1}^{n} f_k(x_{\text{st}}^N) + \frac{3\tilde{R}}{\sqrt{N}} \sum_{j=1}^{m} \left[ c_j - \sum_{k=1}^{n} x_{jk}^N \right],
\]

where \( \sum_{t=0}^{N} \sum_{j=1}^{m} x_{tj}^N = 1/N \sum_{t=0}^{N-1} x_{tk}(p_{\text{st}}^N) \) and \( \tilde{p}^N = (p_{\text{st}}^N, \ldots, p_{\text{en}}^N) \).
where $\mathbf{x}^N_k = 1/N \sum_{t=0}^N \mathbf{x}_k(p_{t,k}^*)$. Thus we obtain the following estimation:

$$
\varphi(\bar{\omega}^N) + \frac{1}{N} \sum_{k=1}^n f_k(x^N_k) + \frac{3\bar{R}}{\sqrt{n}} \sum_{j=1}^m \left[ c_j - \sum_{k=1}^n x^N_{jk} \right]_+ \leq \frac{82\bar{L}_\psi \bar{R}^2 m \mu}{9N}.
$$

And similarly to (A21), we obtain

$$
\sqrt{n} = \sum_{j=1}^m \left[ c_j - \sum_{k=1}^n x^N_{jk} \right]_+ \leq \frac{82\bar{L}_\psi \bar{R}^2 m \mu}{9N}.
$$

And due to the weak duality, we obtain

$$
\sum_{k=1}^n f_k(x^N_k) - \sum_{k=1}^n f_k(x^*_k) \leq \frac{82L_\psi \bar{R}^2 m \mu}{9N}.
$$

And since $\bar{R} = 3\sqrt{n}p_{max}$ and $L_\psi \leq \frac{n\mu}{\bar{p}}$, we obtain the statement of the theorem. ■

### Composite gradient method for the resource allocation (vector case)

**Input**: $N > 0$ – number of steps, $L_\psi$ – Lipschitz constant of gradient $\psi$, $p_{0,j}$, $j = 1, \ldots, ym$ – starting point.

1. Knowing the prices $p_{t,k}^*$, $k = 1, \ldots, yn$ for the current year $t$, producers calculate the optimal plan for the production according these prices as

$$
\mathbf{x}_k(p_{t,k}^*) = \arg\max_{x_k \geq 0} \left\{ \sum_{j=1}^m p_{t,j}^* x_{jk} - f_k(x_k) \right\}, \quad k = 1, 2, \ldots, n.
$$

2. Each factory predicts the price for the next year $t + 1$ for product $j = 1, \ldots, m$ as

$$
\tilde{p}_{t+1}^{j,k} = p_{t,j}^* - \frac{1}{L_\psi} x_{jk}(p_{t,j}^*), \quad k = 1, 2, \ldots, n
$$

and send this information to the Center.

3. The Center determines the price $p_{t+1}^{j,\text{center}}$, at which it will purchase product $j$ for the next year $t + 1$ as

- If $\sum_{k=1}^n \left( -\tilde{p}_{t+1}^{j,k} \right)_+ \geq \frac{C}{L_\psi}$ then $p_{t+1}^{j,\text{center}} = 0$;
- Else $p_{t+1}^{j,\text{center}} > 0$ is determined from

$$
\sum_{k=1}^n \left( p_{t+1}^{j,\text{center}} - \tilde{p}_{t+1}^{j,k} \right)_+ = \frac{C}{L_\psi},
$$

and sends this price to all producers.

4. Each producer adjusts the price for product $j = 1, \ldots, m$ for the next year as follows

$$
p_{t+1,j}^{jk} = \max \left( p_{t+1}^{j,\text{center}}, \tilde{p}_{t+1}^{j,k} \right), \quad k = 1, \ldots, n.
A.1.2 Accelerated composite gradient method
In this section, to solve the resource allocation problem in vector case, we use accelerated composite gradient method from the Section 5. Similarly to the previous section, the step of the method can be rewritten as

\[
\tilde{y}^{t+1} = \arg\min_{\bar{p} \geq 0} \left\{ \alpha_{t+1} \left( \left( y^{t+1}, \bar{p} - \tilde{p}^{t+1} \right) + \xi (p_{1}, \ldots, y_{n}) \right) + \frac{1}{2} \| \bar{p} - \tilde{y}^{t} \|^{2} \right\}
\]

\[
= \arg\min_{\bar{p} \geq 0} \left\{ \alpha_{t+1} \left( \sum_{k=1}^{n} (x_{k}^{t+1}, p_{k} - \tilde{p}_{k}^{t+1}) - \sum_{j=1}^{m} c_{j} \min_{k=1, \ldots, n} p_{jk} \right) + \frac{1}{2} \sum_{k=1}^{n} \| p_{k} - y_{k}^{t} \|^{2} \right\}
\]

\[
= \arg\min_{\bar{p} \geq 0} \left\{ \sum_{j=1}^{m} \left( \alpha_{t+1} \left( \left( x_{j}^{t+1}, p_{j} - \tilde{p}_{j}^{t+1} \right) - c_{j} \min_{k=1, \ldots, n} p_{jk} \right) + \frac{1}{2} \| p_{j} - y_{j}^{t} \|^{2} \right) \right\},
\]

where \( p_{j} \) is the price vector of all producers for product \( j \) and \( x_{j}^{t+1} \) is the vector of optimal plans for the production of product \( j \) for all producers. Define \( \tilde{y}_{j}^{t+1} = y_{j}^{t} - \alpha_{t+1} x_{j}^{t+1} \) and then, using Lemma 4.1, we obtain the following solution:

- If \( \sum_{k=1}^{m} (y_{jk}^{t+1}) > 0 \) then \( y_{j,center}^{t+1} = 0 \).
- Else \( y_{j,center}^{t+1} > 0 \) is determined from

\[
\sum_{k=1}^{n} \left( y_{j,center}^{t+1} - y_{jk}^{t+1} \right) = c_{j} \alpha_{t+1}
\]

and the solution is determined as

\[
y_{jk}^{t+1} = \max \left( y_{j,center}^{t+1}, y_{jk}^{t+1} \right), \quad k = 1, \ldots, n.
\]

### Accelerated composite gradient descent for resource allocation (vector case)

**Input:** \( N > 0 \) – number of steps, \( L_{\psi} \) – Lipschitz constant of gradient \( \psi \), \( p_{1}^{0} = y_{1}^{0} = w_{1}^{0}, \quad j = 1, \ldots, m \) – starting points.

1. In the current year producers find \( \alpha_{t+1} \) as the largest root of the equation

\[
A_{t+1} := A_{t} + \alpha_{t+1} = L_{\psi} \alpha_{t+1}^{2}.
\]
**Theorem A.2:** Let Algorithm (A35) be run for $N$ steps with starting points $p_{j*}^0 = y_{j*}^0 = w_{j*}^0$, $j = 1, \ldots, m$, satisfying $0 \leq p_{j*}^0 \leq \bar{p}_{\text{max}}$, $k = 1, \ldots, n$, where $\bar{p}_{\text{max}}$ is given in (A32). Then

\[
\sum_{k=1}^{n} f_k(x_{j*}^{N_k}) - \sum_{k=1}^{n} f(x_{j*}^{N_k}) \leq \sum_{k=1}^{n} f_k(x_{j*}^{N_k}) + \varphi(\bar{\omega}^n) \leq \varphi(\bar{\omega}^N) + \sum_{k=1}^{n} f_k(x_{j*}^{N_k}) \leq \frac{148\bar{p}_{\text{max}}^2 n^2 m}{(N+1)^2 \mu},
\]

\[
\sum_{j=1}^{m} \left[ c_j - \sum_{k=1}^{n} x_{j*}^{N_k} \right]_+ \leq \frac{148 \bar{p}_{\text{max}} m n^2}{5(N+1)^2 \mu},
\]

where $x_{j*}^{N_k} = 1/A_N \sum_{t=0}^{N} \alpha_t x_{j*}(p_{j*}^{t+1})$ and $\bar{\omega}^N = (w_1^N, \ldots, w_{n*}^N)$.

**Proof of Theorem A.2:** Similarly to the proof of Theorem 5.1, we can obtain Equation (A24), but in this case we can write down this equation as

\[
A_N \varphi(\bar{\omega}^N) \leq \min_{\tilde{p} > 0} \left\{ \sum_{t=0}^{N} \alpha_t \left( \sum_{j=1}^{m} x_{j*}(p_{j*}^{t+1}) - \sum_{k=1}^{n} f_k(x_{j*}^{N_k}) + \xi(p) \right) + \frac{1}{2} \sum_{j=1}^{m} \|p_j - y_{j*}^0\|_2^2 \right\},
\]
where $\bar{w}^N = (w_{1}^N, \ldots, y_{m}^N)$. Since we also have equations similarly to (A19) we obtain that $\bar{p}^t'_{i}, w^t_{n}, y^t_{n} \in B^+_R(0), \forall t$, where $B^+_R(0) = \{p : p \geq 0, \|p\|_2 \leq 2\bar{R} \}, \bar{R} = 3\sqrt{m \rho_{\max}}$. So, we obtain

$A_N \varphi(\bar{w}^N) \leq \min_{\tilde{p}_j \in B_{2\bar{R}}^+(0)} \left\{ \sum_{t=0}^{N} \alpha_t \left( \sum_{j=1}^{n} \left( x_j^N(\bar{p}^t_{j+1}), \tilde{p}_j \right) - \sum_{k=1}^{n} f_k(x_{k}) + \xi(\tilde{p}) \right) \right\} + \frac{1}{2} \sum_{j=1}^{m} \| \bar{p}_j - y^0_j \|_2^2$

And since

$\| \bar{p}_j - y^0_j \|_2^2 \leq 2\| y^0_j \|_2^2 + 2\| y^0_j \|_2^2 \leq 8\bar{R}^2 + \frac{2\bar{R}^2}{9} = \frac{74}{9}\bar{R}^2$

obtain the following:

$A_N \varphi(\bar{w}^N) \leq \min_{\tilde{p}_j \in B_{2\bar{R}}^+(0)} \left\{ \sum_{t=0}^{N} \alpha_t \left( \sum_{j=1}^{n} \left( x_j^N(\bar{p}^t_{j+1}), \tilde{p}_j \right) - \sum_{k=1}^{n} f_k(x_{k}) + \xi(\tilde{p}) \right) \right\} + \frac{37m}{9} \bar{R}^2.$

Notice that

$\max_{\tilde{p}_j \in B_{2\bar{R}}^+(0)} \left\{ \sum_{t=0}^{N} \alpha_t \left( \sum_{j=1}^{n} \left( x_j^N(\bar{p}^t_{j+1}), \tilde{p}_j \right) - \sum_{k=1}^{n} f_k(x_{k}) + \xi(\tilde{p}) \right) \right\}

\geq A_t \sum_{k=1}^{n} f_k(x_{k}) + A_t \max_{\tilde{p}_j \in B_{2\bar{R}}^+(0)} \left\{ \sum_{t=0}^{N} \alpha_t \left( -\sum_{j=1}^{m} \left( x_j^N(\bar{p}^t_{j+1}), \tilde{p}_j \right) + \sum_{j=1}^{m} c_j \min_{k=1, \ldots, y^n} \bar{p}_{jk} \right) \right\}

= A_t \sum_{k=1}^{n} f_k(x_{k}) + A_t \max_{\tilde{p}_j \in B_{2\bar{R}}^+(0)} \left\{ \sum_{j=1}^{m} \left( -\sum_{k=1}^{n} x_{jk}^N + c_j \bar{p}_j \right) \right\}

= A_t \sum_{k=1}^{n} f_k(x_{k}) + A_t \sum_{j=1}^{m} \frac{2\bar{R}}{\sqrt{n}} \sum_{j=1}^{m} \left[ c_j - \sum_{k=1}^{n} x_{jk}^N \right]_+,

where $x_{jk}^N = 1/A_N \sum_{t=0}^{N} \alpha_t x_{k}(p_{jk}^t)$. Thus we obtain the following estimation:

$\varphi(\bar{w}^N) + \sum_{k=1}^{n} f_k(x_{k}) + \frac{2\bar{R}}{\sqrt{n}} \sum_{j=1}^{m} \left[ c_j - \sum_{k=1}^{n} x_{jk}^N \right]_+ \leq \frac{37R^2m}{9A_N}.$

And similarly to (A27), we obtain

$\sum_{j=1}^{m} \left[ c_j - \sum_{k=1}^{n} x_{jk}^N \right]_+ \leq \frac{37m}{15A_N}.$
And due to the weak duality, we obtain
\[
\sum_{k=1}^{n} f_k(x^N_k) - \sum_{k=1}^{n} f_k(x^*_k) \leq \frac{37\bar{R}^2 m}{9A_N}.
\]
And using the estimation (A29) for the $A_t, t \geq 0$, we obtain the statement on the theorem.