Composite optimization for the resource allocation problem

Anastasiya Ivanova[a*], Pavel Dvurechensky[b†],
Alexander Gasnikov[a,c‡] and Dmitry Kamzolov [a§].

[a]Moscow Institute of Physics and Technology, Moscow, Russia;
[b] Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany;
[c]Institute for Information Transmission Problems RAS, Moscow, Russia;

October 1, 2019

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.

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 [1], the mechanisms of decentralized resource allocation gained a lot of attention in economics and operations research since then, see e.g. [3, 6, 10] and references therein. Each producer seeks to minimize its own costs and, in total, all the producers need to produce a certain

1

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.

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 [2]. 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 the 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) [9] or the ellipsoid method (for many-product economy) [6]. These algorithms are effective in the case when the amount of producers are small.

In this paper, following [12], 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 [12], 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 optimal for convex nonsmooth optimization convergence rate $O\left(\dfrac{1}{\sqrt{t}}\right)$, t being the iteration counter. In contrast to [12], 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 $O\left(\dfrac{1}{t}\right)$ and accelerated gradient descent to obtain convergence rate $O\left(\dfrac{1}{t^2}\right)$.

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 one 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

$$(P) \qquad \min_{\substack{\sum\limits_{k=1}^{n} y_k \geqslant C,\ x_k \geqslant y_k; \\ y_k \geqslant 0,\ x_k \geqslant 0,\ k=1,\dots,n,}} f(\mathbf{x}) = \sum_{k=1}^{n} f_k(x_k),$$

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

Remark 1. We point that, from the optimization point of view the variables y_k, $k = 1, ..., n$ 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. 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 one 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, etc.

Introducing dual variables p_k, $k = 1, ..., n$ and using the duality theory, we obtain

$$\min_{\substack{\sum_{k=1}^{n} y_k \geqslant C, \, x_k \geqslant y_k, \, y_k \geqslant 0; \\ x_k \geqslant 0, \, k=1,...,n}} f(\mathbf{x}) = \min_{\substack{\sum_{k=1}^{n} y_k \geqslant C \, y_k \geqslant 0; \\ x_k \geqslant 0, \, k=1,...,n}} \left\{ f(\mathbf{x}) + \sum_{k=1}^{n} \max_{p_k \geqslant 0} p_k(y_k - x_k) \right\}$$

$$= -\min_{p_1, ..., p_n \geqslant 0} \left\{ \sum_{k=1}^{n} \max_{x_k \geqslant 0} (p_k x_k - f_k(x_k)) - \min_{\substack{\sum_{k=1}^{n} y_k \geqslant C; \, y_k \geqslant 0}} \sum_{k=1}^{n} p_k y_k \right\}$$

$$[y_{k^*}^* = C, \text{ where } k^* = \arg\min_k p_k \text{ and } y_k^* = 0 \text{ for } k \neq k^*]$$

$$= -\min_{p_1, ..., p_n \geqslant 0} \left\{ \sum_{k=1}^{n} \max_{x_k \geqslant 0} (p_k x_k - f_k(x_k)) - C \min_{k=1,...,n} p_k \right\}$$

$$= -\min_{p_1, ..., p_n \geqslant 0} \left\{ \sum_{k=1}^{n} \left\{ p_k x_k(p_k) - f_k(x_k(p_k)) \right\} - C \min_{k=1,...,n} p_k \right\},$$

where

$$x_k(p_k) = \underset{x_k \geqslant 0}{\operatorname{argmax}} \left\{ p_k x_k - f_k(x_k) \right\}, \quad k = 1, 2, \ldots, n. \tag{1}$$

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

$$(D) \qquad \varphi(p_1, \ldots, p_n) = \sum_{k=1}^{n} \left\{ p_k x_k(p_k) - f_k(x_k(p_k)) \right\} - C \min_{k=1,...,n} p_k \to \min_{p_1, ..., p_n \geqslant 0}.$$

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, $k = 1, ..., n$ can be reconstructed. We point that primal-dual gradient methods [4, 5, 8] are not applicable here since the dual problem (D) is a composite optimization problem.

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 [9], 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.
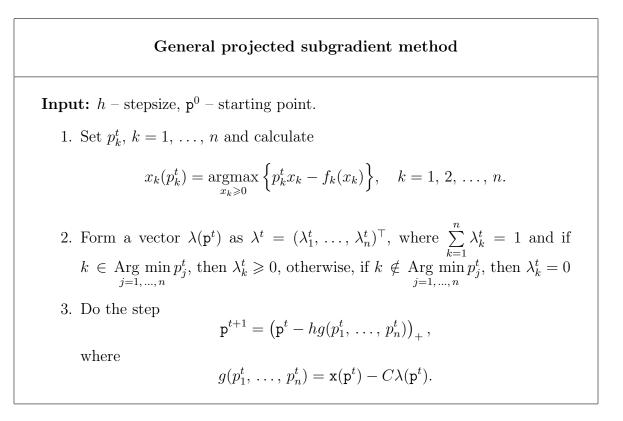
To solve the problem (D), we use the projected subgradient method with the step given by

$$\mathbf{p}^{t+1} = \left[\mathbf{p}^t - hg(\mathbf{p}^t)\right]_+ , \tag{2}$$

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

$$g(p_1, \ldots, p_n) = \mathbf{x}(\mathbf{p}) - C\lambda(\mathbf{p}), \tag{3}$$

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

General projected subgradient method

Input: h – stepsize, $\mathbf{p}^0$ – starting point.

1. Set p_k^t, $k = 1, \ldots, n$ and calculate

$$x_k(p_k^t) = \operatorname*{argmax}_{x_k \geqslant 0} \left\{ p_k^t x_k - f_k(x_k) \right\}, \quad k = 1, 2, \ldots, n.$$

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

3. Do the step
$$\mathbf{p}^{t+1} = \left(\mathbf{p}^t - hg(p_1^t, \ldots, p_n^t)\right)_+ ,$$
where
$$g(p_1^t, \ldots, p_n^t) = \mathbf{x}(\mathbf{p}^t) - C\lambda(\mathbf{p}^t).$$

Next, we discuss the economic interpretation of the steps of the subgradient method. First of all, $C\lambda_k(\mathbf{p}^t)$ can be seen as the volume the Center intends to purchase from the k-th producer on t iteration. Indeed, the dual variable $\mathbf{p}$ corresponds to the vector of prices set by the producers and non-zero components of the vector λ can correspond only to those

producers, who have minimal price. Since $\sum_{k=1:\, k \in \underset{j=1,\,...,\,n}{\text{Arg min}\, p_j}}^{n} \lambda_k^t = 1$ and all λ_k^t are non-negative, λ_k^t for $k \in \underset{j=1,\,...,\,n}{\text{Arg min}\, p_j}$ can be seen as the relative proportions of purchase form the producers having minimal price. Thus, $C\lambda_k(\mathbf{p}^t)$ 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, $\mathbf{p}^0$ – starting point.

1. Set the stepsize $h = \frac{\varepsilon}{nC^2}$.

2. Given the price vector $\mathbf{p}^t$ for the current year, producers calculate the optimal production plan for these prices as

$$x_k(p_k^t) = \underset{x_k \geqslant 0}{\text{argmax}} \left\{ p_k^t x_k - f_k(x_k) \right\}, \quad k = 1,\, 2,\, \ldots,\, n.$$

and communicates this information to the Center.

3. The Center determines the shares of purchases for each producer, i.e. forms a vector $\lambda(\mathbf{p}^t)$ as $\lambda^t = (\lambda_1^t,\, \ldots,\, \lambda_n^t)^\top$, where $\sum_{k=1}^{n} \lambda_k^t = 1$, $\lambda_k^t \geqslant 0$ if $k \in \underset{j=1,\,...,\,n}{\text{Arg min}\, p_j^t}$ and $\lambda_k^t = 0$, if $k \notin \underset{j=1,\,...,\,n}{\text{Arg min}\, p_j^t}$ and sends this vector to all factories.

4. Each factory adjusts the price for the next year as follows

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

(4)

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

Lemma 1. Let the $\mathbf{p}^*$ be a solution to the dual problem (D). Then

$$\|\mathbf{p}^*\|_2 \leqslant \sqrt{n} p_{max}. \tag{5}$$

where

$$p_{max} := \frac{n}{C} \left(\sum_{k=1}^{n} f_k \left(\frac{2C}{n} \right) - \sum_{k=1}^{n} f_k(0) \right). \tag{6}$$

The proof of this lemma is deferred to the Appendix.

Then, we can formulate the following theorem about convergence rate

Theorem 1 ([9]). Let Algorithm (4) be run with starting point $\mathbf{p}^0$ satisfying $0 \leqslant p_k^0 \leqslant p_{max}$, $k = 1, \ldots, n$ for

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

steps. Then

$$f(\mathbf{x}^N) - f(\mathbf{x}^*) \leqslant \varepsilon, \ C - \sum_{k=1}^{n} x_k^N \leqslant \frac{\varepsilon}{3 p_{max}}, \tag{7}$$

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

Note that the number of iterations N to achieve accuracy ε 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

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

where

$$\psi(p_1, \ldots, p_n) = \sum_{k=1}^{n} \left\{ p_k x_k(p_k) - f_k(x_k(p_k)) \right\} = \langle \mathbf{p}, \mathbf{x}(\mathbf{p}) \rangle - f(\mathbf{x}(\mathbf{p})) \tag{8}$$

is convex function. Gradient of this function

$$\nabla \psi(p) = \mathbf{x}(\mathbf{p}), \tag{9}$$

7

satisfies Lipschitz condition (see e.g. [11])

$$\|\nabla\psi(\mathbf{p}^1) - \nabla\psi(\mathbf{p}^2)\|_2 \leqslant L_\psi\|\mathbf{p}^1 - \mathbf{p}^2\|_2, \ \forall\mathbf{p}^1, \mathbf{p}^2 \geqslant 0, \tag{10}$$

where $L_\psi = \dfrac{n}{\mu}$ (see e.g. [9]) and

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

is convex non smooth function. The idea is to use first-order information about ψ and use the function g as a whole, see the main projected gradient step (11). This allows the method to work in accordance with the smoothness properties of ψ, i.e. be faster since ψ 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_ψ – Lipschitz constant of gradient ψ, $\mathbf{p}^0$ – starting point.

1. Find
$$x_k(p_k^t) = \operatorname*{argmax}_{x_k \geqslant 0} \left\{ p_k^t x_k - f_k(x_k) \right\}, \quad k = 1, 2, \ldots, n.$$

2. Do the step
$$\mathbf{p}^{t+1} = \operatorname*{argmin}_{\mathbf{p} \geqslant 0} \left\{ \langle \nabla\psi(\mathbf{p}^t), \mathbf{p} - \mathbf{p}^t \rangle - C \min_{k=1,\ldots,n} p_k + \frac{L_\psi}{2}\|\mathbf{p} - \mathbf{p}^t\|_2^2 \right\}. \tag{11}$$

Lemma 2. Let $\tilde{\mathbf{p}}^{t+1} = \mathbf{p}^t - \dfrac{1}{L_\psi}\mathbf{x}(\mathbf{p}^t)$. Then $\mathbf{p}^{t+1}$ in (11) is defined as follows

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

where if $\sum\limits_{k=1}^{n} \left(-\tilde{p}_k^{t+1}\right)_+ \geqslant \dfrac{C}{L_\psi}$ than $p_{center}^{t+1} = 0$, else $p_{center}^{t+1} > 0$ is a solution of equation

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

The proof of Lemma 2 can be found in Appendix. Note that the step (11) can be rewritten as

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

8

where $\lambda(\mathbf{p}^{t+1})$ is such that $\sum_{k=1}^{n} \lambda_k(p_k^{t+1}) = 1$, $\lambda_k(p_k^{t+1}) \geqslant 0$ if $k \in \underset{j=1,\dots,n}{\mathrm{Arg\,min}}\, p_j^{t+1}$ and $\lambda_k(p_k^{t+1}) = 0$, if $k \notin \underset{j=1,\dots,n}{\mathrm{Arg\,min}}\, p_j^{t+1}$. This equality looks very similar to equality (2).

4.1 Composite gradient method for the resource allocation problem

In this subsection, 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 $\nabla\psi(\mathbf{p}) = \mathbf{x}(\mathbf{p})$, the constant L_ψ determines the relation between prices $\mathbf{p}$ and production $\mathbf{x}(\mathbf{p})$, i.e. says, what is the maximum change in the production if the price changes. From Lemma 2, each year t the Center, knowing the prices and production of each producer, forms a prediction $\tilde{p}_k^{t+1} = p_k^t - \frac{1}{L_\psi} x_k(p_k^t)$ 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_{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) λ depends on the unknown price vector p_k^{t+1}, resulting in the implicit definition of this price vector as a solution to nonlinear equation. But, $C\lambda(\mathbf{p}^{t+1})$ 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 2, the solution of (12) can be written as

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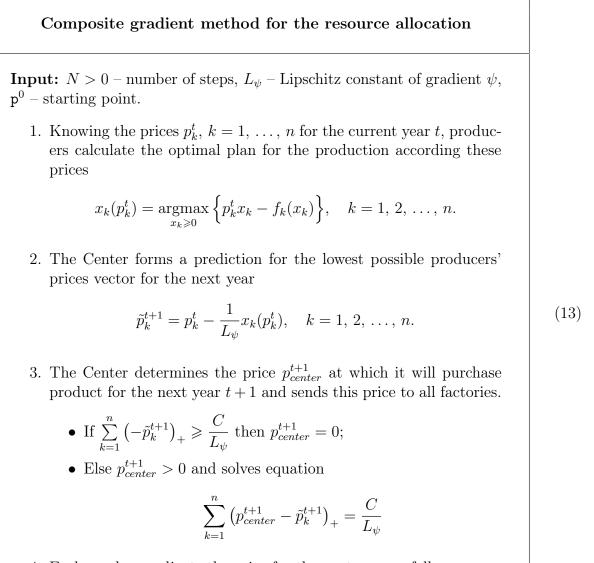
where

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

Thus, the estimate for the purchase from the k-th producer is $L_\psi \left(p_{center}^{t+1} - \tilde{p}_k^{t+1} \right)_+$, i.e. if the predicted price $\tilde{p}_k^{t+1}$ is higher than p_{center}^{t+1}, there will be no purchase from this producer. Again, by Lemma 2, if for the smallest possible price $p_{center}^{t+1} = 0$, the total purchased amount $L_\psi \sum_{k=1}^{n} \left(-\tilde{p}_k^{t+1} \right)_+$ is greater than C, the Center sets its price $p_{center}^{t+1} = 0$. Otherwise, it determines such price p_{center}^{t+1} that the total estimated purchase $L_\psi \sum_{k=1}^{n} \left(p_{center}^{t+1} - \tilde{p}_k^{t+1} \right)_+$ is exactly C. Having defined p_{center}^{t+1}, 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_k^t for the current year t, determines the optimal production plan $x_k(p_k^t)$ for this price and reports the price and production to the Center. After receiving p_{center}^{t+1} from the center, each producer adjusts its price p_k^{t+1} for the next year as $p_k^{t+1} = \max\left(p_{center}^{t+1}, \tilde{p}_k^{t+1} \right)$. This means that the lowest possible price $\tilde{p}_k^{t+1}$ is compared to the Center's price. If $p_{center}^{t+1} \geqslant \tilde{p}_k^{t+1}$, the producer can increase the price up to p_{center}^{t+1} and the Center will buy. On the other hand, if $p_{center}^{t+1} < \tilde{p}_k^{t+1}$, there is no sense for the producer to sell at the price p_{center}^{t+1} 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_ψ – Lipschitz constant of gradient ψ, $\mathbf{p}^0$ – starting point.

1. Knowing the prices p_k^t, $k = 1, \ldots, n$ for the current year t, producers calculate the optimal plan for the production according these prices

$$x_k(p_k^t) = \operatorname*{argmax}_{x_k \geqslant 0} \left\{ p_k^t x_k - f_k(x_k) \right\}, \quad k = 1, 2, \ldots, n.$$

2. The Center forms a prediction for the lowest possible producers' prices vector for the next year

$$\tilde{p}_k^{t+1} = p_k^t - \frac{1}{L_\psi} x_k(p_k^t), \quad k = 1, 2, \ldots, n. \tag{13}$$

3. The Center determines the price p_{center}^{t+1} at which it will purchase product for the next year $t + 1$ and sends this price to all factories.

   - If $\sum_{k=1}^n \left(-\tilde{p}_k^{t+1} \right)_+ \geqslant \dfrac{C}{L_\psi}$ then $p_{center}^{t+1} = 0$;

   - Else $p_{center}^{t+1} > 0$ and solves equation

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

4. Each producer adjusts the price for the next year as follows

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

Theorem 2. Let Algorithm (13) be run for N steps with starting point $\mathbf{p}^0$ satisfying $0 \leqslant p_k^0 \leqslant$

p_{max}, $k = 1, \ldots, n$, where p_{max} is given in (6). Then

$$f(\mathbf{x}^N) - f(\mathbf{x}^*) \leqslant f(\mathbf{x}^N) + \varphi(\mathbf{p}^*) \leqslant \varphi(\mathbf{p}^N) + f(\mathbf{x}^N) \leqslant \frac{82 p_{max}^2 n^2}{N\mu},$$

$$\left[C - \sum_{k=1}^{n} x_k^N \right]_+ \leqslant \frac{82 p_{max} n^2}{3N\mu},$$

where $\mathbf{p}^N = \dfrac{1}{N} \sum\limits_{t=1}^{N} \mathbf{p}^t$ and $\mathbf{x}^N = \dfrac{1}{N} \sum\limits_{t=0}^{N-1} \mathbf{x}(\mathbf{p}^t)$.

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 ε in the following sense

$$f(\mathbf{x}^N) - f(\mathbf{x}^*) \leqslant \varepsilon, \ C - \sum_{k=1}^{n} x_k^N \leqslant \frac{\varepsilon}{3p_{max}}. \tag{14}$$

To solve this problem, we consider the dual problem (D), and solve it by the composite gradient method starting with $\mathbf{p}^0$ satisfying $0 \leqslant p_k^0 \leqslant p_{max}$, $k = 1, \ldots, n$, where p_{max} is given in (6).

Theorem 2 states that Algorithm will find the solution no later than after $N = \dfrac{82 p_{max}^2 n^2}{\varepsilon \mu}$ 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.

We can rewrite 17 step as follows

$$
\begin{aligned}
\mathbf{y}^{t+1} &= \underset{\mathbf{p} \geqslant 0}{\operatorname{argmin}} \left\{ \alpha_{t+1} \left(\langle \nabla \psi(\mathbf{p}^{t+1}), \mathbf{p} - \mathbf{p}^{t+1} \rangle - C \min_{k=1, \ldots, n} p_k \right) + \frac{1}{2} \|\mathbf{p} - \mathbf{y}^t\|_2^2 \right\} \\
&= \underset{\mathbf{p} \geqslant 0}{\operatorname{argmin}} \left\{ \langle \mathbf{x}(\mathbf{p}^{t+1}), \mathbf{p} - \mathbf{p}^{t+1} \rangle + C \max_{k=1, \ldots, n} (-p_k) + \frac{1}{2\alpha_{t+1}} \|\mathbf{p} - \mathbf{y}^t\|_2^2 \right\} \\
&= \underset{\mathbf{p} \geqslant 0}{\operatorname{argmin}} \left\{ \max_{k=1, \ldots, n} (-p_k) + \frac{1}{2C\alpha_{t+1}} \|\mathbf{p} - (\mathbf{y}^t - \alpha_{t+1}\mathbf{x}(\mathbf{p}^{t+1})) \|_2^2 \right\}.
\end{aligned}
$$

Define $\tilde{\mathbf{y}}^{t+1} = \mathbf{y}^t - \alpha_{t+1}\mathbf{x}(\mathbf{p}^{t+1})$ and then, using Lemma 2, we obtain the following solution:

- If $\sum\limits_{k=1}^{n} \left(-\tilde{y}_k^{t+1} \right)_+ \geqslant C\alpha_{t+1}$ then $y_{center}^{t+1} = 0$ and

$$y_k^{t+1} = \max \left(0, \tilde{y}_k^{t+1} \right), \ k = 1, \ldots, n.$$

11

- Else $y_{center}^{t+1} > 0$ is determined from

$$\sum_{k=1}^{n} \left(y_{center}^{t+1} - \tilde{y}_k^{t+1} \right)_+ = C\alpha_{t+1}$$

and

$$y_k^{t+1} = \max \left(y_{center}^{t+1}, \tilde{y}_k^{t+1} \right), \ k = 1, \dots, n.$$

Accelerated composite gradient descent scheme

Input: $N > 0$ – number of steps, L_ψ – Lipschitz constant of gradient ψ, $\mathbf{p}^0$, $\mathbf{y}^0$, $\mathbf{w}^0$ – starting points, $\alpha_0 = A_0 = 0$.

1. Find α_{t+1} as the largest root of the equation

$$A_{t+1} := A_t + \alpha_{t+1} = L_\psi \alpha_{t+1}^2. \tag{15}$$

2. Calculate

$$\mathbf{p}^{t+1} = \frac{\alpha_{t+1}\mathbf{y}^t + A_t \mathbf{w}^t}{A_{t+1}}. \tag{16}$$

3. Calculate

$$\mathbf{y}^{t+1} = \operatorname*{argmin}_{\mathbf{p} \geqslant 0} \left\{ \alpha_{t+1} \left(\langle \nabla\psi(\mathbf{p}^{t+1}), \mathbf{p} - \mathbf{p}^{t+1} \rangle + g(\mathbf{p}) \right) + \frac{1}{2}\|\mathbf{p} - \mathbf{y}^t\|_2^2 \right\}. \tag{17}$$

4. Calculate

$$\mathbf{w}^{t+1} = \frac{\alpha_{t+1}\mathbf{y}^{t+1} + A_t \mathbf{w}^t}{A_{t+1}}. \tag{18}$$

5. Set

$$\mathbf{x}^{t+1} = \frac{1}{A_{t+1}} \sum_{i=0}^{t+1} \alpha_i \mathbf{x}(\mathbf{p}^i) = \frac{\alpha_{t+1} x(\mathbf{p}^{t+1}) + A_t \mathbf{x}^t}{A_{t+1}}.$$

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_ψ, 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 α_t, which is adjusted on each iteration and allows to have faster convergence. Vector $\mathbf{w}^{t+1}$ can be interpreted as an average of the historical price data $\mathbf{y}^t$, $t \geq 0$. As we prove in next theorem, $\mathbf{w}^{t+1}$ 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 $\mathbf{y}^t$ from the previous period to form $\mathbf{p}^{t+1}$. Then the optimal production plan is calculated and, similarly to the non-accelerated method, producers predict the price for next year $\tilde{\mathbf{y}}^{t+1}$ 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.

Theorem 3. Let Algorithm (20) be run for N steps with starting points $\mathbf{p}^0 = \mathbf{y}^0 = \mathbf{w}^0$ satisfying $0 \leqslant p_k^0 \leqslant p_{max}$, $k = 1, \ldots, n$, where p_{max} is given in (6). Then

$$f(\mathbf{x}^N) - f(\mathbf{x}^*) \leqslant f(\mathbf{x}^N) + \varphi(\mathbf{w}^*) \leqslant \varphi(\mathbf{w}^N) + f(\mathbf{x}^N) \leqslant \frac{148n^2 p_{max}^2}{(N+1)^2 \mu},$$

$$\left[C - \sum_{k=1}^n x_k^N \right]_+ \leqslant \frac{148n^2 R p_{max}}{5(N+1)^2 \mu},$$

where $\mathbf{x}^N = \dfrac{1}{A_N} \sum_{t=0}^N \alpha_t \mathbf{x}(\mathbf{p}^t)$

The proof of this theorem mostly follows the steps of the Theorem 2 in the previous section, but we give all steps for the reader's convenience in Appendix.

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.

Firstly, 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 10000 tons of wood. Each local producer has the following cost function:

$$f_k(x_k) = \alpha_k x_k + \frac{\mu}{2} x_k^2, \quad k = 1, \ldots, 100, \tag{19}$$

where α_k is the different local costs of production and transportation for each producer, μ is the costs 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 α 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 fig. 1. The convergence rate for the constraint feasibility is shown in log scale in fig. 2. The convergence rate for the value of the duality gap is shown in log scale in fig. 3.
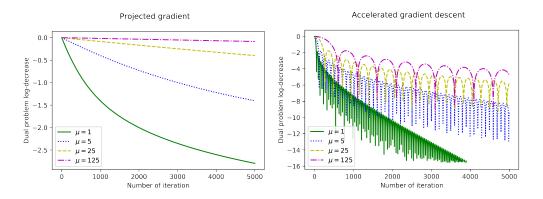
13

Accelerated composite gradient descent for the resource allocation

Input: $N > 0$ – number of steps, L_ψ Lipschitz constant of gradient ψ, $\mathbf{p}^0 = \mathbf{y}^0 = \mathbf{w}^0$ – starting point.

1. In the current year t producers find α_{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 $\mathbf{p}^{t+1}$ as the following convex combination

$$p_k^{t+1} = \frac{\alpha_{t+1} y_k^t + A_t w_k^t}{A_{t+1}}, \ k = 1, \dots, n$$

and calculate the optimal plan for the production according to this price as

$$x_k(p_k^{t+1}) = \underset{x_k \geqslant 0}{\operatorname{argmax}} \left\{ p_k^{t+1} x_k - f_k(x_k) \right\}, \quad k = 1, 2, \dots, n.$$

3. Each producer predicts the price for the next year $t+1$ as

$$\tilde{y}_k^{t+1} = y_k^t - \alpha_{t+1} x_k(\mathbf{p}^{t+1}), \ k = 1, \dots, n$$

and sends this information to the Center.

4. The Center determines the prediction price y_{center}^{t+1} at which it will purchase product for the next year $t+1$ as

   - If $\sum_{k=1}^n \left(-\tilde{y}_k^{t+1} \right)_+ \geqslant C\alpha_{t+1}$ then $y_{center}^{t+1} = 0$

   - Else $y_{center}^{t+1} > 0$ and is determined from

$$\sum_{k=1}^n \left(y_{center}^{t+1} - \tilde{y}_k^{t+1} \right)_+ = C\alpha_{t+1},$$

   and sends this price to all producers.

5. Each producer adjusts the prediction price for the next year as

$$y_k^{t+1} = \max \left(y_{center}^{t+1}, \tilde{y}_k^{t+1} \right), \ k = 1, \dots, n.$$

   and calculates the historical price for for the next year

$$w_k^{t+1} = \frac{\alpha_{t+1} y_k^{t+1} + A_t w_k^t}{A_{t+1}^{14}}, \ k = 1, \dots, n.$$

Figure 1: Projected gradient vs Accelerated gradient descent, 100 factories, cost functions (19), Dual value in log scale, 1 run and mean of 20 run
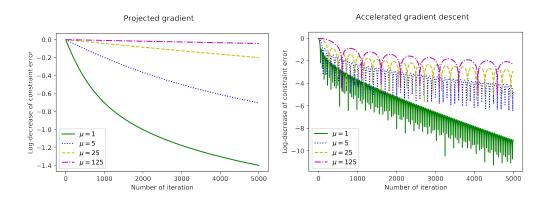

Figure 2: Projected gradient vs Accelerated gradient descent, 100 factories, cost functions (19), Inexactness of constraints in log scale, 1 run and mean of 20 run
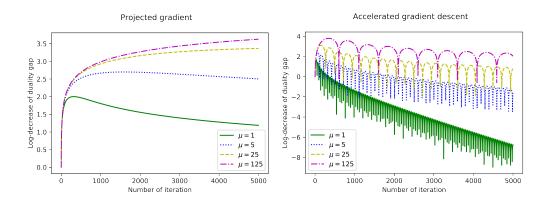

Figure 3: Projected gradient vs Accelerated gradient descent, 100 factories, cost functions (19), Value of duality gap in log scale, 1 run and mean of 20 run

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.

Secondly, we consider, how the results depend on the strong convexity parameter μ. For that, we consider functions (19) with $\alpha \in N[100; 400]$ and $\mu \in \{1, 5, 25, 125\}$. We make 20 runs with random α 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 fig. 4. The convergence rate for the constraint feasibility is shown in log scale in fig. 5. The convergence rate for the value of the duality gap is shown in log scale in fig. 6.

Figure 4: Projected gradient vs Accelerated gradient descent, 100 factories, cost functions (19), Dual value in log scale, mean of 20 run

Figure 5: Projected gradient vs Accelerated gradient descent, 100 factories, cost functions (19), Inexactness of constraints in log scale, mean of 20 run

16

Figure 6: Projected gradient vs Accelerated gradient descent, 100 factories, cost functions (19), Duality gap value in log scale, mean of 20 run

These plots show that the theoretical dependence on μ 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.

References

[1] Kenneth Joseph Arrow and Leonid Hurwicz. *Decentralization and computation in resource allocation.* Stanford University, Department of Economics, 1958.

[2] Jean-Pierre Aubin. L'analyse non-linéaire et ses motivations économiques. *Masson*, 1984.

[3] Donald E Campbell et al. *Resource allocation mechanisms.* Cambridge University Press, 1987.

[4] Pavel Dvurechensky, Darina Dvinskikh, Alexander Gasnikov, César A. Uribe, and Angelia Nedić. Decentralize and randomize: Faster algorithm for Wasserstein barycenters. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, NeurIPS 2018, pages 10783–10793. Curran Associates, Inc., 2018. arXiv:1806.03915.

[5] Pavel Dvurechensky, Alexander Gasnikov, Evgenia Gasnikova, Sergey Matsievsky, Anton Rodomanov, and Inna Usik. Primal-dual method for searching equilibrium in hierarchical congestion population games. In *Supplementary Proceedings of the 9th International Conference on Discrete Optimization and Operations Research and Scientific School (DOOR 2016) Vladivostok, Russia, September 19 - 23, 2016*, pages 584–595, 2016. arXiv:1606.08988.

[6] Eric J Friedman and Shmuel S Oren. The complexity of resource allocation and price mechanisms under bounded rationality. *Economic Theory*, 6(2):225–250, 1995.

[7] Alexander Gasnikov. Universal gradient descent. *arXiv preprint arXiv:1711.00394*, 2017.

[8] S. V. Guminov, Yu. 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(2):125–128, Mar 2019.

[9] Anastasiya Ivanova, Alexander Gasnikov, Evgeni Nurminski, and Evgeniya 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. / Sib. Branch of Russ. Acad. of Sci. –– Novosibirsk, 2019. –– Vol. 22, No 4. –– P. 411–432.*

[10] Ali Kakhbod. *Resource allocation in decentralized systems with strategic agents: an implementation theory approach.* Springer Science & Business Media, 2013.

[11] Yu Nesterov. Smooth minimization of non-smooth functions. *Mathematical programming*, 103(1):127–152, 2005.

[12] Yu Nesterov and Vladimir Shikhman. Dual subgradient method with averaging for optimal resource allocation. *European Journal of Operational Research*, 270(3):907–916, 2018.

7 Appendix

Proof of Lemma 1.

Set $\bar{x}_1 = \ldots = \bar{x}_n = \dfrac{2C}{n}$ and $\bar{y}_1 = \ldots = \bar{y}_n = \dfrac{C}{n}$. And define $\bar{\mathbf{x}} = (\bar{x}_1, \ldots, \bar{x}_n)^\top$ and $\bar{\mathbf{y}} = (\bar{y}_1, \ldots, \bar{y}_n)^\top$. Notice that the point $\left(\bar{\mathbf{x}}^\top, \bar{\mathbf{y}}^\top\right)^\top$ satisfies the Slater's condition as

$$\frac{2C}{n} = \bar{x}_k > \bar{y}_k = \frac{C}{n}, \quad k = 1, \ldots, n.$$

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

$$\sum_{k=1}^{n} f_k(0) = \min_{\substack{\sum\limits_{k=1}^{n} y_k \geqslant C,\, y_k \geqslant 0, \\ x_k \geqslant 0,\, k=1,\ldots,n}} \sum_{k=1}^{n} f_k(x_k)$$

$$= \min_{\substack{\sum\limits_{k=1}^{n} y_k \geqslant C,\, y_k \geqslant 0, \\ x_k \geqslant 0,\, k=1,\ldots,n}} \left\{ \sum_{k=1}^{n} f_k(x_k) + \sum_{k=1}^{n} (y_k - x_k) \underbrace{p_k}_{=0} \right\}$$

$$\leqslant \max_{p \geqslant 0} \min_{\substack{\sum\limits_{k=1}^{n} y_k \geqslant C,\, y_k \geqslant 0, \\ x_k \geqslant 0,\, k=1,\ldots,n}} \left\{ \sum_{k=1}^{n} f_k(x_k) + \sum_{k=1}^{n} (y_k - x_k)\, p_k \right\}$$

$$= \min_{\substack{\sum\limits_{k=1}^{n} y_k \geqslant C,\, y_k \geqslant 0, \\ x_k \geqslant 0,\, k=1,\ldots,n}} \left\{ \sum_{k=1}^{n} f_k(x_k) + \sum_{k=1}^{n} (y_k - x_k)\, p_k^* \right\}$$

$$\leqslant \sum_{k=1}^{n} f_k(\bar{x}_k) + \sum_{k=1}^{n} (\bar{y}_k - \bar{x}_k)\, p_k^*$$

$$\leqslant \sum_{k=1}^{n} f_k(\bar{x}_k) - \frac{C}{n} \sum_{k=1}^{n} p_k^*.$$

Since $p_k^* \geqslant 0$, $k = 1, \ldots, n$ we obtain that

$$\|\mathbf{p}^*\|_1 \leqslant p_{max},$$

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

$$p_k^* \leqslant p_{max}, \quad k = 1, \ldots, n$$

19

from which we obtain the statement of the Lemma. $\qquad\square$

Proof of Lemma 2.

Using (8), we can rewrite the step as

$$
\begin{aligned}
\mathbf{p}^{t+1} &= \operatorname*{argmin}_{\mathbf{p}\geqslant 0}\left\{ C\max_{k=1,\dots,n}(-p_k) + \langle \mathbf{x}(\mathbf{p}^t),\mathbf{p}-\mathbf{p}^t\rangle + \frac{L_\psi}{2}\left(\|\mathbf{p}\|_2^2 - 2\langle \mathbf{p},\mathbf{p}^t\rangle + \|\mathbf{p}^t\|_2^2\right)\right\} \\
&= \operatorname*{argmin}_{\mathbf{p}\geqslant 0}\left\{ C\max_{k=1,\dots,n}(-p_k) + \frac{L_\psi}{2}\left(\|\mathbf{p}\|_2^2 - 2\langle \mathbf{p},\mathbf{p}^t\rangle + \frac{2}{L_\psi}\langle \mathbf{x}(\mathbf{p}^t),\mathbf{p}-\mathbf{p}^t\rangle + \|\mathbf{p}^t\|_2^2\right)\right\} \\
&= \operatorname*{argmin}_{\mathbf{p}\geqslant 0}\left\{ \max_{k=1,\dots,n}(-p_k) + \frac{L_\psi}{2C}\left\|\mathbf{p}-\left(\mathbf{p}^t - \frac{1}{L_\psi}\mathbf{x}(\mathbf{p}^t)\right)\right\|_2^2\right\}.
\end{aligned}
$$

Let us define $\gamma = \dfrac{C}{L_\psi}$. Then to determine $\mathbf{p}^{t+1}$ it is necessary to solve the following problem

$$
\max_{k=1,\dots n}(-p_k) + \frac{1}{2\gamma}\|\mathbf{p}-\tilde{\mathbf{p}}^{t+1}\|_2^2 \to \min_{p_k\geqslant 0,\,k=1,\dots n,}
$$

which we can rewrite as the following equivalent problem

$$
\eta_{t+1} + \frac{1}{2\gamma}\|\mathbf{p}-\tilde{\mathbf{p}}^{t+1}\|_2^2 \to \min_{\eta_{t+1}\leqslant 0,\,\eta_{t+1}\geqslant -p_k,\,k=1,\dots n}
$$

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

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

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

$$
-p_k^* \leqslant \eta_{t+1}^*,\ k=1,\dots,n. \tag{21}
$$

$$
z_k^* \geqslant 0,\ k=1,\dots,n. \tag{22}
$$

$$
z_k^*(p_k^* + \eta_{t+1}^*) = 0,\ k=1,\dots,n. \tag{23}
$$

$$
\frac{1}{\gamma}(p_k^* - \tilde{p}_k^{t+1}) - z_k^* = 0,\ k=1,\dots,n. \tag{24}
$$

$$
\sum_{k=1}^{n} z_k^* = 1 + w^*. \tag{25}
$$

$$
w^*\eta_{t+1}^* = 0. \tag{26}
$$

$$
w^* \geqslant 0. \tag{27}
$$

$$
\eta_{t+1}^* \leqslant 0. \tag{28}
$$

- If $\eta_{t+1}^* = 0$.

– If $\eta^*_{t+1} = -p^*_k = 0$, then from (24) and (22) obtain, that

$$z^*_k = \frac{1}{\gamma}(-\tilde{p}^{t+1}_k) = \frac{1}{\gamma}(-\tilde{p}^{t+1}_k)_+. \tag{29}$$

– If $p^*_k > -\eta^*_{t+1} = 0$ then from (23) and (24)

$$z^*_k = 0, \; p^*_k = \tilde{p}^{t+1}_k.$$

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

$$z^*_k = 0 = \frac{1}{\gamma}(-\tilde{p}^{t+1}_k)_+. \tag{30}$$

Then, using (29), (30), (25) and (27) obtain, that

$$\sum_{k=1}^{n}(-\tilde{p}^{t+1}_k)_+ = \gamma(1 + w^*) \geqslant \gamma.$$

As a result, we obtain that for $\sum\limits_{k=1}^{n}(-\tilde{p}^{t+1}_k)_+ \geqslant \gamma$ the case then $\eta^*_{t+1} = 0$ and solution determines as

$$p^*_k = \max\{0, \tilde{p}^{t+1}_k\}.$$

- If $\eta^*_{t+1} < 0$ then $w^* = 0$.

– If $p^*_k = -\eta^*_{t+1}$ then from (24) obtain

$$z^*_k = \frac{1}{\gamma}(-\eta^*_{t+1} - \tilde{p}^{t+1}_k)_+. \tag{31}$$

– If $p^*_k > -\eta^*_{t+1}$ then from (23) and (24) we obtain that $z^*_k = 0$, $p^*_k = \tilde{p}^{t+1}_k$. Also since in this case $\tilde{p}^{t+1}_k = p^*_k > -\eta^*_{t+1}$, then

$$z^*_k = 0 = \frac{1}{\gamma}(-\eta^*_{t+1} - \tilde{p}^{t+1}_k)_+. \tag{32}$$

Finally using (31), (32) and (25) obtain that

$$\sum_{k=1}^{n}(-\eta^*_{t+1} - \tilde{p}^{t+1}_k)_+ = \gamma.$$

and solution determines as

$$p^*_k = \max\{-\eta^*_{t+1}, \tilde{p}^{t+1}_k\}.$$

21

So, define $p_{center}^{t+1} = -\eta_{t+1}^*$ we finishes the proof. $\qquad\square$

Proof of Theorem 2.

Let's consider the step

$$\mathbf{p}^{t+1} = \underset{\mathbf{p} \geqslant 0}{\operatorname{argmin}} \underbrace{\left\{ \langle \nabla\psi(\mathbf{p}^t), \mathbf{p} - \mathbf{p}^t \rangle - C \min_{k=1,\ldots,n} p_k + \frac{L_\psi}{2} \|\mathbf{p} - \mathbf{p}^t\|_2^2 \right\}}_{\Phi(\mathbf{p},\mathbf{p}^t)}.$$

The function $\Phi(\mathbf{p}, \mathbf{p}^t)$ is convex with respect to the variable $\mathbf{p}$, then from the definition of $\mathbf{p}^{t+1}$ follows that

$$\langle \nabla_{\mathbf{p}^{t+1}}\Phi(\mathbf{p}^{t+1}, \mathbf{p}^t), \mathbf{p} - \mathbf{p}^{t+1} \rangle \geqslant 0, \ \forall \mathbf{p} \geqslant 0. \tag{33}$$

Define function

$$\tilde{\Phi}(\mathbf{p}, \mathbf{p}^t) = \langle \nabla\psi(\mathbf{p}^t), \mathbf{p} - \mathbf{p}^t \rangle - C \min_{k=1,\ldots,n} p_k = \langle \nabla\psi(\mathbf{p}^t), \mathbf{p} - \mathbf{p}^t \rangle + g(\mathbf{p}),$$

which is convex with respect to the variable $\mathbf{p}$. From (33) we obtain

$$0 \leqslant \langle \nabla_{\mathbf{p}^{t+1}}\Phi(\mathbf{p}^{t+1}, \mathbf{p}^t), \mathbf{p} - \mathbf{p}^{t+1} \rangle = \langle \nabla_{\mathbf{p}^{t+1}}\tilde{\Phi}(\mathbf{p}^{t+1}, \mathbf{p}^t) + L_\psi(\mathbf{p}^{t+1} - \mathbf{p}^t), \mathbf{p} - \mathbf{p}^{t+1} \rangle$$

$$\overset{conv-ty}{\leqslant} \tilde{\Phi}(\mathbf{p}, \mathbf{p}^t) - \tilde{\Phi}(\mathbf{p}^{t+1}, \mathbf{p}^t) + L_\psi \langle \mathbf{p}^{t+1} - \mathbf{p}^t, \mathbf{p} - \mathbf{p}^{t+1} \rangle. \tag{34}$$

Notice, that

$$L_\psi \langle \mathbf{p}^{t+1} - \mathbf{p}^t, \mathbf{p} - \mathbf{p}^{t+1} \rangle = L_\psi \langle \mathbf{p}^{t+1}, \mathbf{p} \rangle + L_\psi \langle \mathbf{p}^t, \mathbf{p}^{t+1} \rangle - L_\psi \langle \mathbf{p}^{t+1}, \mathbf{p}^{t+1} \rangle - L_\psi \langle \mathbf{p}^t, \mathbf{p} \rangle$$

$$= \frac{L}{2} \langle \mathbf{p}, \mathbf{p} \rangle - L_\psi \langle \mathbf{p}^t, \mathbf{p} \rangle + \frac{L_\psi}{2} \langle \mathbf{p}^t, \mathbf{p}^t \rangle - \frac{L_\psi}{2} \langle \mathbf{p}, \mathbf{p} \rangle + L_\psi \langle \mathbf{p}, \mathbf{p}^{t+1} \rangle$$

$$- \frac{L_\psi}{2} \langle \mathbf{p}^{t+1}, \mathbf{p}^{t+1} \rangle - \frac{L_\psi}{2} \langle \mathbf{p}^{t+1}, \mathbf{p}^{t+1} \rangle - \frac{L_\psi}{2} \langle \mathbf{p}^t, \mathbf{p}^t \rangle + L_\psi \langle \mathbf{p}^t, \mathbf{p}^{t+1} \rangle$$

$$= \frac{L_\psi}{2} \|\mathbf{p} - \mathbf{p}^t\|_2^2 - \frac{L_\psi}{2} \|\mathbf{p} - \mathbf{p}^{t+1}\|_2^2 - \frac{L_\psi}{2} \|\mathbf{p}^{t+1} - \mathbf{p}^t\|_2^2. \tag{35}$$

Substituting this inequality to (34), we obtain

$$0 \leqslant \tilde{\Phi}(\mathbf{p}, \mathbf{p}^t) - \tilde{\Phi}(\mathbf{p}^{t+1}, \mathbf{p}^t) + \frac{L_\psi}{2} \|\mathbf{p} - \mathbf{p}^t\|_2^2 - \frac{L_\psi}{2} \|\mathbf{p} - \mathbf{p}^{t+1}\|_2^2 - \frac{L_\psi}{2} \|\mathbf{p}^{t+1} - \mathbf{p}^t\|_2^2$$

$$= \langle \nabla\psi(\mathbf{p}^t), \mathbf{p} - \mathbf{p}^t \rangle + g(\mathbf{p}) - \langle \nabla\psi(\mathbf{p}^t), \mathbf{p}^{t+1} - \mathbf{p}^t \rangle - g(\mathbf{p}^{t+1}) + \frac{L_\psi}{2} \|\mathbf{p} - \mathbf{p}^t\|_2^2$$

$$- \frac{L_\psi}{2} \|\mathbf{p} - \mathbf{p}^{t+1}\|_2^2 - \frac{L_\psi}{2} \|\mathbf{p}^{t+1} - \mathbf{p}^t\|_2^2$$

Finally, since (10) we can rewrite as

$$\psi(\mathbf{p}^1) \leqslant \psi(\mathbf{p}^2) + \langle \nabla\psi(\mathbf{p}^2), \mathbf{p}^1 - \mathbf{p}^2 \rangle + \frac{L_\psi}{2} \|\mathbf{p}^1 - \mathbf{p}^2\|_2^2, \ \forall \mathbf{p}^1, \mathbf{p}^2 \geqslant 0 \tag{36}$$

22

we obtain that

$$\psi(\mathbf{p}^{t+1}) \leqslant \psi(\mathbf{p}^t) + \langle \nabla \psi(\mathbf{p}^t), \mathbf{p} - \mathbf{p}^t \rangle + g(\mathbf{p}) - g(\mathbf{p}^{t+1}) + \frac{L_\psi}{2}\|\mathbf{p} - \mathbf{p}^t\|_2^2 - \frac{L_\psi}{2}\|\mathbf{p} - \mathbf{p}^{t+1}\|_2^2$$

or

$$\varphi(\mathbf{p}^{t+1}) \leqslant \varphi(\mathbf{p}^t) + \langle \nabla \psi(\mathbf{p}^t), \mathbf{p} - \mathbf{p}^t \rangle - g(\mathbf{p}^t) + g(\mathbf{p}) + \frac{L_\psi}{2}\|\mathbf{p} - \mathbf{p}^t\|_2^2 - \frac{L_\psi}{2}\|\mathbf{p} - \mathbf{p}^{t+1}\|_2^2.$$

Summing these inequalities from $t = 0$ to $t = N - 1$ we get, using the convexity of φ and that it is true for all $\mathbf{p} \geqslant 0$

$$\varphi(\mathbf{p}^N) \leqslant \frac{1}{N} \min_{\mathbf{p} \geqslant 0} \left\{ \sum_{t=0}^{N-1} \left[\varphi(\mathbf{p}^t) + \langle \nabla \psi(\mathbf{p}^t), \mathbf{p} - \mathbf{p}^t \rangle - g(\mathbf{p}^t) \right] + g(\mathbf{p}) + \frac{L_\psi}{2}\|\mathbf{p} - \mathbf{p}^0\|_2^2 \right\}, \quad (37)$$

where $\mathbf{p}^N = \frac{1}{N}\sum_{t=1}^{N} \mathbf{p}^t$. Let us define the starting point $\mathbf{p}^0$ satisfying $0 \leqslant p_k^0 \leqslant p_{max}$, $k = 1, \ldots, n$, where p_{max} is given in (6). Then we obtain that

$$\|\mathbf{p}^0\|_2 \leqslant \sqrt{n}p_{max}.$$

Let us introduce a set

$$B_{3R}^+(0) = \{\mathbf{p} \; : \; \mathbf{p} \geqslant 0, \; \|\mathbf{p} - \mathbf{p}^0\|_2 \leqslant 3R\},$$

where

$$\|\mathbf{p}^0 - \mathbf{p}^*\|_2 + \|\mathbf{p}^0\|_2 \leqslant 2\|\mathbf{p}^0\|_2 + \|\mathbf{p}^*\|_2 = 3p_{max}\sqrt{n} = R, \quad (38)$$

herewith all the obtaining $\mathbf{p}^t$ will consist in $B_{2R}^+(0)$:

$$\|\mathbf{p}^t\|_2 \leqslant 2R, \quad (39)$$

since (second paragraph [7])

$$\|\mathbf{p}^t\|_2 = \|\mathbf{p}^t - \mathbf{p}^0\|_2 + \|\mathbf{p}^0\|_2 \leqslant \|\mathbf{p}^t - \mathbf{p}^*\|_2 + \|\mathbf{p}^* - \mathbf{p}^0\|_2 + \|\mathbf{p}^0\|_2$$
$$\leqslant 2\|\mathbf{p}^* - \mathbf{p}^0\|_2 + \|\mathbf{p}^0\|_2 = 2\|\mathbf{p}^*\|_2 + 3\|\mathbf{p}^0\|_2 = 5p_{max}\sqrt{n} \leqslant 2R.$$

Since that we can rewrite (37) as

$$\varphi(\mathbf{p}^N) \leqslant \frac{1}{N} \min_{\mathbf{p} \geqslant 0} \left\{ \sum_{t=0}^{N-1} \left[\varphi(\mathbf{p}^t) + \langle \nabla \psi(\mathbf{p}^t), \mathbf{p} - \mathbf{p}^t \rangle - g(\mathbf{p}^t) \right] + g(\mathbf{p}) + \frac{L_\psi}{2}\|\mathbf{p} - \mathbf{p}^0\|_2^2 \right\}$$
$$\leqslant \frac{1}{N} \min_{\mathbf{p} \in B_{3R}^+(0)} \left\{ \sum_{t=0}^{N-1} \left[\psi(\mathbf{p}^t) + \langle \nabla \psi(\mathbf{p}^t), \mathbf{p} - \mathbf{p}^t \rangle \right] + g(\mathbf{p}) + \frac{L_\psi}{2}\|\mathbf{p} - \mathbf{p}^0\|_2^2 \right\}.$$

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, $p^* \in B_R^+(p^0)$. Define

$$\tilde{p} = \underset{p \in B_{3R}^+(0)}{\operatorname{argmin}} \left\{ \sum_{t=0}^{N-1} \left[\psi(p^t) + \langle \nabla \psi(p^t), p - p^t \rangle \right] + g(p) \right\},$$

then, since

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

we obtain

$$\varphi(p^N) \leqslant \frac{1}{N} \min_{p \in B_{3R}^+(0)} \left\{ \sum_{t=0}^{N-1} \left[\psi(p^t) + \langle \nabla \psi(p^t), p - p^t \rangle \right] + g(p) + \frac{L_\psi}{2} \|p - p^0\|_2^2 \right\}$$

$$\leqslant \frac{1}{N} \min_{p \in B_{3R}^+(0)} \left\{ \sum_{t=0}^{N-1} \left[\psi(p^t) + \langle \nabla \psi(p^t), p - p^t \rangle \right] + g(p) \right\} + \frac{L_\psi}{2N} \|\tilde{p} - p^0\|_2^2$$

$$\leqslant \frac{1}{N} \min_{p \in B_{3R}^+(0)} \left\{ \sum_{t=0}^{N-1} \left[\psi(p^t) + \langle \nabla \psi(p^t), p - p^t \rangle \right] + g(p) \right\} + \frac{82 L_\psi R^2}{9N}.$$

Notice, that

$$-\frac{1}{N} \min_{p \in B_{3R}^+(0)} \left\{ \sum_{t=0}^{N-1} \left[\psi(p^t) + \langle \nabla \psi(p^t), p - p^t \rangle \right] + g(p) \right\}$$

$$= \frac{1}{N} \max_{p \in B_{3R}^+(0)} \left\{ \sum_{t=0}^{N-1} \left[\langle x(p^t), p^t - p \rangle + \sum_{k=1}^{n} f_k(x_k(p^t)) - \langle x(p^t), p^t \rangle \right] - g(p) \right\}$$

$$= \frac{1}{N} \sum_{t=0}^{N-1} f(x(p^t)) + \frac{1}{N} \max_{p \in B_{3R}^+(0)} \left\{ \sum_{t=0}^{N-1} \langle -x(p^t), p \rangle - g(p) \right\}$$

$$\overset{conv-ty}{\geqslant} f(x^N) + \frac{1}{N} \max_{p \in B_{3R}^+(0)} \left\{ \sum_{t=0}^{N-1} \langle -x(p^t), p \rangle - g(p) \right\}$$

$$= f(x^N) + \max_{p \in B_{3R}^+(0)} \left\{ -\langle x^N, p \rangle + C \min_{k=1,\dots,n} p_k \right\},$$

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

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

we obtain

$$\max_{\mathbf{p} \in B_{3R}^+(0)} \left\{ -\langle \mathbf{x}^N, \mathbf{p} \rangle + C \min_{k=1,\dots,n} p_k \right\} = \max_{\substack{\mathbf{p} \in B_{3R}^+(0) \\ p_1 = \dots = p_n = p}} \left\{ -\langle \mathbf{x}^N, \mathbf{p} \rangle + C \min_{k=1,\dots,n} p_k \right\}$$

$$= \max_{\substack{\mathbf{p} \in B_{3R}^+(0) \\ p_1 = \dots = p_n = p}} \left\{ Cp - \sum_{k=1}^{n} x_k^N p \right\} = \frac{3R}{\sqrt{n}} \left[C - \sum_{k=1}^{n} x_k^N \right]_+ .$$

Thus, we obtain the following estimation

$$\varphi(\mathbf{p}^N) + f(\mathbf{x}^N) + \frac{3R}{\sqrt{n}} \left[C - \sum_{k=1}^{n} x_k^N \right]_+ \leqslant \frac{82 L_\psi R^2}{9N}. \tag{40}$$

Considering the weak duality $-f(\mathbf{x}^*) \leqslant \varphi(\mathbf{p}^*)$, obtain

$$f(\mathbf{x}^N) - f(\mathbf{x}^*) \leqslant f(\mathbf{x}^N) + \varphi(\mathbf{p}^*) \leqslant \varphi(\mathbf{p}^N) + f(\mathbf{x}^N)$$

$$\leqslant \varphi(\mathbf{p}^N) + f(\mathbf{x}^N) + \frac{3R}{\sqrt{n}} \left[C - \sum_{k=1}^{n} x_k^N \right]_+ \leqslant \frac{82 L_\psi R^2}{9N} = \varepsilon.$$

Using (39), (40) and (D) we get

$$\frac{R}{\sqrt{n}} \left[C - \sum_{k=1}^{n} x_k^N \right]_+ \leqslant \overbrace{\underbrace{\langle \mathbf{x}^N, \mathbf{p}^N \rangle - C \min_{k=1,\dots,n} p_k^N - f(\mathbf{x}^N)}_{\geqslant -\frac{2R}{\sqrt{n}} \left[C - \sum_{k=1}^{n} x_k^N \right]_+}}^{\leqslant \varphi(\mathbf{p}^N)} + f(\mathbf{x}^N) + \frac{3R}{\sqrt{n}} \left[C - \sum_{k=1}^{n} x_k^N \right]_+ \leqslant \varepsilon. $$

$$\tag{41}$$

And since $R = 3 p_{max} \sqrt{n}$ and $L_\psi = \dfrac{n}{\mu}$ we obtain the statement of the theorem.

$\square$

Proof of Theorem 3.

Let's consider the step

$$\mathbf{y}^{t+1} = \underset{\mathbf{p} \geqslant 0}{\arg\min} \underbrace{\left\{ \alpha_{t+1} \left(\langle \nabla \psi(\mathbf{p}^{t+1}), \mathbf{p} - \mathbf{p}^{t+1} \rangle - C \min_{k=1,\dots,n} p_k \right) + \frac{1}{2} \|\mathbf{p} - \mathbf{y}^t\|_2^2 \right\}}_{\Psi(\mathbf{p}, \mathbf{p}^{t+1})}.$$

The function $\Psi(\mathbf{p}, \mathbf{p}^t)$ is convex with respect to the variable $\mathbf{p}$, then from the definition of $\mathbf{y}^{t+1}$ follows that

$$\langle \nabla_{\mathbf{y}^{t+1}} \Psi(\mathbf{y}^{t+1}, \mathbf{p}^{t+1}), \mathbf{p} - \mathbf{y}^{t+1} \rangle \geqslant 0, \ \forall \mathbf{p} \geqslant 0. \tag{42}$$

Define function

$$\tilde{\Psi}(\mathbf{p},\mathbf{p}^t) = \left\langle \nabla\psi(\mathbf{p}^t),\mathbf{p}-\mathbf{p}^t\right\rangle - C\min_{k=1,\ldots,n}p_k = \left\langle \nabla\psi(\mathbf{p}^t),\mathbf{p}-\mathbf{p}^t\right\rangle + g(\mathbf{p}),$$

which is convex with respect to the variable $\mathbf{p}$. From (42) we obtain

$$0 \leqslant \left\langle \nabla_{\mathbf{y}^{t+1}}\Psi(\mathbf{y}^{t+1},\mathbf{p}^{t+1}),\mathbf{p}-\mathbf{y}^{t+1}\right\rangle = \left\langle \alpha_{t+1}\nabla_{\mathbf{y}^{t+1}}\tilde{\Psi}(\mathbf{y}^{t+1},\mathbf{p}^{t+1}) + (\mathbf{y}^{t+1}-\mathbf{y}^t),\mathbf{p}-\mathbf{y}^{t+1}\right\rangle$$

$$\overset{conv-ty}{\leqslant} \alpha_{t+1}\tilde{\Psi}(\mathbf{p},\mathbf{p}^{t+1}) - \alpha_{t+1}\tilde{\Psi}(\mathbf{y}^{t+1},\mathbf{p}^{t+1}) + \left\langle \mathbf{y}^{t+1}-\mathbf{y}^t,\mathbf{p}-\mathbf{y}^{t+1}\right\rangle. \tag{43}$$

Using (35), we obtain

$$0 \leqslant \alpha_{t+1}\tilde{\Psi}(\mathbf{p},\mathbf{p}^{t+1}) - \alpha_{t+1}\tilde{\Phi}(\mathbf{y}^{t+1},\mathbf{p}^{t+1}) + \frac{1}{2}\|\mathbf{p}-\mathbf{y}^t\|_2^2 - \frac{1}{2}\|\mathbf{p}-\mathbf{y}^{t+1}\|_2^2 - \frac{1}{2}\|\mathbf{y}^{t+1}-\mathbf{y}^t\|_2^2$$

$$= \alpha_{t+1}\left\langle \nabla\psi(\mathbf{p}^{t+1}),\mathbf{p}-\mathbf{p}^{t+1}\right\rangle + \alpha_{t+1}g(\mathbf{p}) - \alpha_{t+1}\left(\left\langle \nabla\psi(\mathbf{p}^{t+1}),\mathbf{y}^{t+1}-\mathbf{p}^{t+1}\right\rangle + g(\mathbf{y}^{t+1})\right)$$

$$+ \frac{1}{2}\|\mathbf{p}-\mathbf{y}^t\|_2^2 - \frac{1}{2}\|\mathbf{p}-\mathbf{y}^{t+1}\|_2^2 - \frac{1}{2}\|\mathbf{y}^{t+1}-\mathbf{y}^t\|_2^2$$

Notice, that

$$-\alpha_{t+1}\left(\left\langle \nabla\psi(\mathbf{p}^{t+1}),\mathbf{y}^{t+1}-\mathbf{p}^{t+1}\right\rangle + g(\mathbf{y}^{t+1})\right) - \frac{1}{2}\|\mathbf{y}^{t+1}-\mathbf{y}^t\|_2^2$$

$$= \alpha_{t+1}\left(\left\langle \nabla\psi(\mathbf{p}^{t+1}),\mathbf{y}^{t+1}-\mathbf{y}^t\right\rangle + \left\langle \nabla\psi(\mathbf{p}^{t+1}),\mathbf{y}^t-\mathbf{p}^{t+1}\right\rangle - g(\mathbf{y}^{t+1})\right) - \frac{1}{2}\|\mathbf{y}^{t+1}-\mathbf{y}^t\|_2^2$$

$$\overset{(16),\ (18)}{=} A_t\left\langle \nabla\psi(\mathbf{p}^{t+1}),\mathbf{w}^t-\mathbf{p}^{t+1}\right\rangle - \alpha_{t+1}g(\mathbf{y}^{t+1}) + A_{t+1}\left\langle \nabla\psi(\mathbf{p}^{t+1}),\mathbf{p}^{t+1}-\mathbf{w}^{t+1}\right\rangle$$

$$- \frac{A_{t+1}^2}{2\alpha_{t+1}^2}\|\mathbf{p}^{t+1}-\mathbf{w}^{t+1}\|_2^2$$

$$\overset{(15)}{\leqslant} A_t\left\langle \nabla\psi(\mathbf{p}^{t+1}),\mathbf{w}^t-\mathbf{p}^{t+1}\right\rangle - \alpha_{t+1}g(\mathbf{y}^{t+1})$$

$$+ A_{t+1}\left(\left\langle \nabla\psi(\mathbf{p}^{t+1}),\mathbf{p}^{t+1}-\mathbf{w}^{t+1}\right\rangle - \frac{L}{2}\|\mathbf{p}^{t+1}-\mathbf{w}^{t+1}\|_2^2\right)$$

$$\overset{(36)}{\leqslant} A_t\psi(\mathbf{w}^t) - A_t\psi(\mathbf{p}^{t+1}) - \alpha_{t+1}g(\mathbf{y}^{t+1}) + A_{t+1}\psi(\mathbf{p}^{t+1}) - A_{t+1}\psi(\mathbf{w}^{t+1})$$

$$= A_t\psi(\mathbf{w}^t) - \alpha_{t+1}g(\mathbf{y}^{t+1}) + \alpha_{t+1}\psi(\mathbf{p}^{t+1}) - A_{t+1}\psi(\mathbf{w}^{t+1}) + A_tg(\mathbf{w}^t) - A_tg(\mathbf{w}^t)$$

$$\overset{(18),conv-ty}{\leqslant} A_t\psi(\mathbf{w}^t) + \alpha_{t+1}\psi(\mathbf{p}^{t+1}) - A_{t+1}\psi(\mathbf{w}^{t+1}) + A_tg(\mathbf{w}^t) - A_{t+1}g(\mathbf{w}^{t+1})$$

$$= A_t\varphi(\mathbf{w}^t) + \alpha_{t+1}\psi(\mathbf{p}^{t+1}) - A_{t+1}\varphi(\mathbf{w}^{t+1})$$

Finally, using this, we obtain that

$$A_{t+1}\varphi(\mathbf{w}^{t+1}) - A_t\varphi(\mathbf{w}^t) \leqslant \alpha_{t+1}\left(\psi(\mathbf{p}^{t+1}) + \left\langle \nabla\psi(\mathbf{p}^{t+1}),\mathbf{p}-\mathbf{p}^{t+1}\right\rangle + g(\mathbf{p})\right)$$

$$+ \frac{1}{2}\|\mathbf{p}-\mathbf{y}^t\|_2^2 - \frac{1}{2}\|\mathbf{p}-\mathbf{y}^{t+1}\|_2^2.$$

Sum all the inequalities for $t = 0, \dots, N-1$. Then, for any $\mathbf{p} \geqslant 0$

$$A_N \varphi(\mathbf{w}^N) - A_0 \varphi(\mathbf{w}^0) \leqslant \sum_{t=0}^{N-1} \alpha_{t+1} \left(\psi(\mathbf{p}^{t+1}) + \left\langle \nabla \psi(\mathbf{p}^{t+1}), \mathbf{p} - \mathbf{p}^{t+1} \right\rangle + g(\mathbf{p}) \right)$$
$$+ \frac{1}{2} \|\mathbf{p} - \mathbf{y}^0\|_2^2 - \frac{1}{2} \|\mathbf{p} - \mathbf{y}^N\|_2^2.$$

Whence, since $C_0 = \alpha_0 = 0$

$$A_N \varphi(\mathbf{w}^N) \leqslant \sum_{t=0}^{N} \alpha_t \left(\psi(\mathbf{p}^t) + \left\langle \nabla \psi(\mathbf{p}^t), \mathbf{p} - \mathbf{p}^t \right\rangle + g(\mathbf{p}) \right) + \frac{1}{2} \|\mathbf{p} - \mathbf{y}^0\|_2^2.$$

Taking in the right hand side the minimum in $\mathbf{p} \geqslant 0$ we obtain

$$A_N \varphi(\mathbf{w}^N) \leqslant \min_{\mathbf{p} \geqslant 0} \left\{ \sum_{t=0}^{N} \alpha_t \left(\psi(\mathbf{p}^t) + \left\langle \nabla \psi(\mathbf{p}^t), \mathbf{p} - \mathbf{p}^t \right\rangle + g(\mathbf{p}) \right) + \frac{1}{2} \|\mathbf{p} - \mathbf{y}^0\|_2^2 \right\}. \qquad (44)$$

Put the vector $\mathbf{y}^0 = (y_1^0, \dots, y_n^0)^\top$, where $\mathbf{y}^0$ — vector of initial prices. And its components such as

$$0 \leqslant y_k^0 \leqslant p_{max}, \quad k = 1, \dots, n,$$

then we obtain that

$$\|\mathbf{y}^0\|_2 \leqslant \sqrt{n} p_{max}.$$

Let us introduce a set $B_{2R}^+(0) = \{ \mathbf{p} : \mathbf{p} \geqslant 0, \|\mathbf{p}\|_2 \leqslant 2R \}$, where R determines from (38). Since (39) we obtain

$$A_N \varphi(\mathbf{w}^N) \leqslant \min_{\mathbf{p} \geqslant 0} \left\{ \sum_{t=0}^{N} \alpha_t \left(\psi(\mathbf{p}^t) + \left\langle \nabla \psi(\mathbf{p}^t), \mathbf{p} - \mathbf{p}^t \right\rangle + g(\mathbf{p}) \right) + \frac{1}{2} \|\mathbf{p} - \mathbf{y}^0\|_2^2 \right\}$$
$$\leqslant \min_{\mathbf{p} \in B_{2R}^+(0)} \left\{ \sum_{t=0}^{N} \alpha_t \left(\psi(\mathbf{p}^t) + \left\langle \nabla \psi(\mathbf{p}^t), \mathbf{p} - \mathbf{p}^t \right\rangle + g(\mathbf{p}) \right) + \frac{1}{2} \|\mathbf{p} - \mathbf{y}^0\|_2^2 \right\}.$$

And since

$$\|\mathbf{p} - \mathbf{y}^0\|_2^2 \leqslant 2\|\mathbf{y}^0\|_2^2 + 2\|\mathbf{p}\|_2^2 \leqslant 8R^2 + \frac{2}{9}R^2 = \frac{74}{9}R^2$$

obtain the following

$$A_N \varphi(\mathbf{w}^N) \leqslant \min_{\mathbf{p} \in B_{2R}^+(0)} \left\{ \sum_{t=0}^{N} \alpha_t \left(\psi(\mathbf{p}^t) + \left\langle \nabla \psi(\mathbf{p}^t), \mathbf{p} - \mathbf{p}^t \right\rangle + g(\mathbf{p}) \right) \right\} + \frac{37}{9}R^2.$$

Notice, that

$$-\min_{\mathbf{p}\in B_{2R}^+(\mathbf{0})}\left\{\sum_{t=0}^N \alpha_t\left(\psi(\mathbf{p}^t)+\langle\nabla\psi(\mathbf{p}^t),\mathbf{p}-\mathbf{p}^t\rangle+g(\mathbf{p})\right)\right\}$$

$$=\max_{\mathbf{p}\in B_{2R}^+(\mathbf{0})}\left\{\sum_{t=0}^N \alpha_t\left(\langle\mathbf{x}(\mathbf{p}^t),\mathbf{p}^t-\mathbf{p}\rangle-g(\mathbf{p})-\langle\mathbf{x}(\mathbf{p}^t),\mathbf{p}^t\rangle+\sum_{k=1}^n f_k(x_k(\mathbf{p}^t))\right)\right\}$$

$$=\sum_{t=0}^N \alpha_t f(\mathbf{x}(\mathbf{p}^t))+\max_{\mathbf{p}\in B_{2R}^+(\mathbf{0})}\left\{\sum_{t=0}^N \alpha_t\left(\langle-\mathbf{x}(\mathbf{p}^t),\mathbf{p}\rangle-g(\mathbf{p})\right)\right\}$$

$$\overset{conv-ty}{\geqslant} A_t f(\mathbf{x}^N)+A_t\max_{\mathbf{p}\in B_{2R}^+(\mathbf{0})}\left\{-\langle\mathbf{x}^N,\mathbf{p}\rangle+C\min_{k=1,\dots,n}p_k\right\}$$

$$=A_t f(\mathbf{x}^N)+A_t\max_{\substack{\mathbf{p}\in B_{2R}^+(\mathbf{0})\\p_1=\dots=p_n=p}}\left\{-\langle\mathbf{x}^N,\mathbf{p}\rangle+C\min_{k=1,\dots,n}p_k\right\}$$

$$=A_t f(\mathbf{x}^N)+A_t\max_{\substack{\mathbf{p}\in B_{2R}^+(\mathbf{0})\\p_1=\dots=p_n=p}}\left\{Cp-\sum_{k=1}^n x_k^N p\right\}$$

$$=A_t f(\mathbf{x}^N)+A_t\frac{2R}{\sqrt{n}}\left[C-\sum_{k=1}^n x_k^N\right]_+,$$

where $\mathbf{x}^N=\dfrac{1}{A_N}\sum_{t=0}^N \alpha_t\mathbf{x}(\mathbf{p}^t)$. Thus, we obtain the following estimation

$$\varphi(\mathbf{w}^N)+f(\mathbf{x}^N)+\frac{2R}{\sqrt{n}}\left[C-\sum_{k=1}^n x_k^N\right]_+\leqslant\frac{37R^2}{9A_N}. \tag{45}$$

Since $\mathbf{w}^*$ is an optimal solution of Problem (D), we have, for any $\mathbf{x}\geqslant 0$

$$f(\mathbf{x}^*)\leqslant f(\mathbf{x})-\langle\mathbf{x},\mathbf{w}^*\rangle+C\min_{k=1,\dots,n}w_k^*.$$

Using the estimation (5), we get

$$f(\mathbf{x}^N)-f(\mathbf{x}^*)\geqslant-\frac{R}{3\sqrt{n}}\left[C-\sum_{k=1}^n x_k^N\right]_+.$$

Considering the weak duality $f(\mathbf{x}^*)\geqslant-\varphi(\mathbf{w}^*)$, obtain

$$f(\mathbf{x}^N)+\varphi(\mathbf{w}^N)=f(\mathbf{x}^N)+f(\mathbf{x}^*)-f(\mathbf{x}^*)+\varphi(\mathbf{w}^*)-\varphi(\mathbf{w}^*)+\varphi(\mathbf{w}^N)$$

$$\geqslant f(\mathbf{x}^N)-f(\mathbf{x}^*)\geqslant-\frac{R}{3\sqrt{n}}\left[C-\sum_{k=1}^n x_k^N\right]_+.$$

This and (45) give

$$\left[C - \sum_{k=1}^{n} x_k^N \right]_+ \leqslant \frac{37R\sqrt{n}}{15A_N}. \tag{46}$$

And also due to the weak duality, we obtain

$$f(\mathbf{x}^N) - f(\mathbf{x}^*) \leqslant \frac{37R^2}{9A_N}. \tag{47}$$

Let us show that the lower bound for the sequence A_t, $t \geqslant 0$ is

$$A_t \geqslant \frac{(t+1)^2}{4L_\psi}, \quad \forall t \geqslant 1 \tag{48}$$

where L_ψ is the Lipschitz constant for the gradient of ψ.

For $t = 1$, since $\alpha_0 = 0$ and $A_1 = \alpha_0 + \alpha_1 = \alpha_1$, we have from (15)

$$A_1 = \alpha_1 = \frac{1}{L_\psi} \geqslant \frac{1}{L_\psi}.$$

Let us now assume that (48) holds for some $t \geqslant 1$ and prove that it holds for $t+1$. From (15) we have a quadratic equation for α_{t+1}

$$L_\psi \alpha_{t+1}^2 - \alpha_{t+1} - A_t = 0.$$

Since we need to take the largest root, we obtain,

$$\alpha_{t+1} = \frac{1 + \sqrt{1 + 4L_\psi A_t}}{2L_\psi} = \frac{1}{2L_\psi} + \sqrt{\frac{1}{4L_\psi^2} + \frac{A_t}{L_\psi}} \geqslant \frac{1}{2L_\psi} + \sqrt{\frac{A_t}{L_\psi}}$$

$$\geqslant \frac{1}{2L_\psi} + \frac{1}{\sqrt{L_\psi}} \frac{t+1}{2\sqrt{L_\psi}} = \frac{t+2}{2L_\psi},$$

where we used the induction assumption that (48) holds for t. Using the obtained inequality, from (15) and (48) for t, we get

$$A_{t+1} = A_t + \alpha_{t+1} \geqslant \frac{(t+1)^2}{4L_\psi} + \frac{t+2}{2L_\psi} \geqslant \frac{(t+2)^2}{4L_\psi}.$$

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 (46) and (47).

$\square$

7.1 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 $f_k(\mathbf{x}_k)$, $k = 1, \ldots, n$ representing the total cost of production of the volume $\mathbf{x}_k \in \mathbb{R}^m$ which is the number of tons of products produced by the producer k in one 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 $\mathbf{x}_{j\bullet}$ the j row of matrix and $\mathbf{x}_{\bullet k}$ 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

$$(P_1) \qquad \sum_{k=1}^{n} f_k(\mathbf{x}_{\bullet k}) \to \min_{\substack{\sum_{k=1}^{n} y_{jk} \geqslant c_j,\, j=1,\ldots,m; \\ \mathbf{x}_{\bullet k} \geqslant \mathbf{y}_{\bullet k}, \mathbf{y}_{\bullet k} \geqslant 0,\, k=1,\ldots,n,}}$$

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

Introducing dual variables $\mathbf{p}_{\bullet k}$, $k = 1, \ldots, n$ and using the duality, we obtain

$$\min_{\substack{\sum_{k=1}^{n} y_{jk} \geqslant c_j,\, j=1,\ldots,m; \\ \mathbf{x}_{\bullet k} \geqslant \mathbf{y}_{\bullet k}, \mathbf{y}_{\bullet k} \geqslant 0,\, k=1,\ldots,n,}} \sum_{k=1}^{n} f_k(\mathbf{x}_{\bullet k}) = \min_{\substack{\sum_{k=1}^{n} y_{jk} \geqslant c_j,\, j=1,\ldots,m; \\ \mathbf{y}_{\bullet k} \geqslant 0,\, k=1,\ldots,n}} \left\{ \sum_{k=1}^{n} f_k(\mathbf{x}_{\bullet k}) + \sum_{k=1}^{n} \max_{\mathbf{p}_{\bullet k} \geqslant 0} \mathbf{p}_{\bullet k}^{\top}(\mathbf{y}_{\bullet k} - \mathbf{x}_{\bullet k}) \right\}$$

$$= - \min_{\mathbf{p}_{\bullet 1}, \ldots, \mathbf{p}_{\bullet n} \geqslant 0} \left\{ \sum_{k=1}^{n} \max_{\mathbf{x}_{\bullet k} \geqslant 0}(\mathbf{p}_{\bullet k}^{\top} \mathbf{x}_{\bullet k} - f_k(\mathbf{x}_{\bullet k})) - \min_{\substack{\sum_{k=1}^{n} y_{jk} \geqslant c_j;\, \mathbf{y}_{\bullet k} \geqslant 0}} \sum_{k=1}^{n} \mathbf{p}_{\bullet k}^{\top} \mathbf{y}_{\bullet k} \right\}$$

$$= - \min_{\mathbf{p}_{\bullet 1}, \ldots, \mathbf{p}_{\bullet n} \geqslant 0} \left\{ \sum_{k=1}^{n} \left\{ \mathbf{p}_{k}^{\top} \mathbf{x}_{\bullet k}(\mathbf{p}_{\bullet k}) - f_k(\mathbf{x}_{\bullet k}(\mathbf{p}_{\bullet k})) \right\} - \sum_{j=1}^{m} c_j \min_{k=1,\ldots,n} p_{jk} \right\},$$

where

$$\mathbf{x}_{\bullet k}(\mathbf{p}_{\bullet k}) = \operatorname*{argmax}_{\mathbf{x}_{\bullet k} \geqslant 0} \left\{ \mathbf{p}_{\bullet k}^{\top} \mathbf{x}_k - f_k(\mathbf{x}_{\bullet k}) \right\}, \quad k = 1, 2, \ldots, n. \tag{49}$$

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

$$(D_1) \qquad \varphi(\mathbf{p}_{\bullet 1}, \ldots, \mathbf{p}_{\bullet n}) = \sum_{k=1}^{n} \left\{ \mathbf{p}_{\bullet k}^{\top} \mathbf{x}_{\bullet k}(\mathbf{p}_{\bullet k}) - f_k(\mathbf{x}_{\bullet k}(\mathbf{p}_{\bullet k})) \right\} - \sum_{j=1}^{m} c_j \min_{k=1,\ldots,n} p_{jk} \to \min_{\mathbf{p}_{\bullet 1}, \ldots, \mathbf{p}_{\bullet n} \geqslant 0}.$$

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

$$\varphi(\mathbf{p}_{\bullet 1}, \ldots, \mathbf{p}_{\bullet n}) = \psi(\mathbf{p}_{\bullet 1}, \ldots, \mathbf{p}_{\bullet n}) + g(\mathbf{p}_{\bullet 1}, \ldots, \mathbf{p}_{\bullet n}),$$

where

$$\psi(\mathsf{p}_{\bullet 1}, \ldots, \mathsf{p}_{\bullet n}) = \sum_{k=1}^{n} \left\{ \mathsf{p}_{\bullet k}^{\top} \mathsf{x}_{\bullet k}(\mathsf{p}_{\bullet k}) - f_k(\mathsf{x}_{\bullet k}(\mathsf{p}_{\bullet k})) \right\} \qquad (50)$$

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

$$g(\mathsf{p}_{\bullet 1}, \ldots, \mathsf{p}_{\bullet n}) = -\sum_{j=1}^{m} c_j \min_{k=1, \ldots, n} p_{jk}.$$

We define $\bar{\mathsf{x}}_{\bullet k} = \left(\dfrac{2c_1}{n}, \ldots, \dfrac{2c_m}{n} \right)^{\top}$, $k = 1, \ldots, n$ and $\bar{\mathsf{y}}_{\bullet k} = \left(\dfrac{c_1}{n}, \ldots, \dfrac{c_m}{n} \right)^{\top}$, $k = 1, \ldots, n$, and similarly to Lemma 1, using the Slater's condition, we obtain the following Lemma.

Lemma 3. Let the $\mathsf{p}_{j\bullet}^*$, $j = 1, \ldots, m$ be a solution to the dual problem (D_1). Then it satisfies the inequality

$$\|\mathsf{p}_{j\bullet}^*\|_2 \leqslant \sqrt{n}\bar{p}_{max}.$$

where

$$\bar{p}_{max} := \frac{n}{\min\limits_{j=1,\ldots,m} c_j} \left(\sum_{k=1}^{n} f_k\left(\bar{\mathsf{x}}_{\bullet k}\right) - \sum_{k=1}^{n} f_k(0) \right). \qquad (51)$$

7.1.1 Composite gradient method

In this subsection to solve the resource allocation problem in vector case we use composite gradient method from the Section 4. Let us define vectors $\bar{\mathsf{y}} = \left(\mathsf{y}_{\bullet 1}^{\top}, \ldots, \mathsf{y}_{\bullet n}^{\top} \right) \in \mathbb{R}^{mn}$, $\bar{\mathsf{p}} = \left(\mathsf{p}_{\bullet 1}^{\top}, \ldots, \mathsf{p}_{\bullet n}^{\top} \right) \in \mathbb{R}^{mn}$ and $\bar{\mathsf{x}}(\bar{\mathsf{p}}) = \left(\mathsf{x}_{\bullet 1}(\mathsf{p}_{\bullet 1})^{\top}, \ldots, \mathsf{x}_{\bullet n}(\mathsf{p}_{\bullet n})^{\top} \right) \in \mathbb{R}^{mn}$, then the step of the method can be rewritten as follows

$$\bar{\mathsf{y}}^{t+1} = \operatorname*{argmin}_{\bar{\mathsf{p}} \geqslant 0} \left\{ \left\langle \bar{\mathsf{x}}(\bar{\mathsf{p}}^{t+1}), \bar{\mathsf{p}} - \tilde{\mathsf{p}}^{t+1} \right\rangle + g(\mathsf{p}_{\bullet 1}, \ldots, \mathsf{p}_{\bullet n}) + \frac{L_\psi}{2} \|\bar{\mathsf{p}} - \bar{\mathsf{y}}^t\|_2^2 \right\}$$

$$= \operatorname*{argmin}_{\bar{\mathsf{p}} \geqslant 0} \left\{ \sum_{k=1}^{n} \left\langle \mathsf{x}_{\bullet k}(\mathsf{p}_{\bullet k}^{t+1}), \mathsf{p}_{\bullet k} - \mathsf{p}_{\bullet k}^{t+1} \right\rangle - \sum_{j=1}^{m} c_j \min_{k=1, \ldots, n} p_{jk} + \frac{L_\psi}{2} \sum_{k=1}^{n} \|\mathsf{p}_{\bullet k} - \mathsf{y}_{\bullet k}^t\|_2^2 \right\}$$

$$= \operatorname*{argmin}_{\bar{\mathsf{p}} \geqslant 0} \left\{ \sum_{j=1}^{m} \left(\left\langle \mathsf{x}_{j\bullet}(\mathsf{p}_{j\bullet}^{t+1}), \mathsf{p}_{j\bullet} - \mathsf{p}_{j\bullet}^{t+1} \right\rangle - c_j \min_{k=1, \ldots, n} p_{jk} + \frac{L_\psi}{2} \|\mathsf{p}_{j\bullet} - \mathsf{y}_{j\bullet}^t\|_2^2 \right) \right\},$$

where $\mathsf{p}_{j\bullet} = (p_{j1}, \ldots, p_{jn})^{\top}$, 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

$$\mathsf{y}_{j\bullet}^{t+1} = \operatorname*{argmin}_{\mathsf{p}_{j\bullet} \geqslant 0} \left\{ \left\langle \mathsf{x}_{j\bullet}(\mathsf{p}_{j\bullet}^{t+1}), \mathsf{p}_{j\bullet} - \mathsf{p}_{j\bullet}^{t+1} \right\rangle - c_j \min_{k=1, \ldots, n} p_{jk} + \frac{\bar{L}_\psi}{2} \|\mathsf{p}_{j\bullet} - \mathsf{y}_{j\bullet}^t\|_2^2 \right\},$$

where $p_{j\bullet}$ – is price vector of all producers for product j and $x_{j\bullet}(p_{j\bullet})$ – is vector of optimal plans for the production of product j for all producers. Define $\tilde{p}_{j\bullet}^{t+1} = p_{j\bullet}^{t} - \frac{1}{L_\psi} x_{j\bullet}(p_{j\bullet}^{t+1})$ and then, using Lemma 2, we obtain the following solution:

- If $\sum_{k=1}^{n} \left(-\tilde{p}_{jk}^{t+1}\right)_{+} \geqslant \frac{c_j}{L_\psi}$ then $p_{j.center}^{t+1} = 0$.

- Else $p_{j.center}^{t+1} > 0$ is determined from

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

and the solution is determined as

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

Theorem 4. Let Algorithm (52) be run for N steps with starting points $p_{j\bullet}^0$, $j = 1, \ldots, m$ satisfying $0 \leqslant p_{jk}^0 \leqslant \bar{p}_{max}$, $k = 1, \ldots, n$, where $\bar{p}_{max}$ is given in (51). Then

$$\sum_{k=1}^{n} f_k(x_{\bullet k}^N) - \sum_{k=1}^{n} f(x_{\bullet k}^*) \leqslant \sum_{k=1}^{n} f_k(x_{\bullet k}^N) + \varphi(\bar{p}^*) \leqslant \varphi(\bar{p}^N) + \sum_{k=1}^{n} f_k(x_{\bullet k}^N) \leqslant \frac{82\bar{p}_{max}^2 n^2 m}{N\mu},$$

$$\sum_{j=1}^{m} \left[c_j - \sum_{k=1}^{n} x_{jk}^N\right]_{+} \leqslant \frac{82\bar{p}_{max} n^2 m}{3N\mu},$$

where $p_{\bullet k}^N = \frac{1}{N}\sum_{t=1}^{N} p_{\bullet k}^t$, $x_{\bullet k}^N = \frac{1}{N}\sum_{t=0}^{N-1} x_{\bullet k}(p_{\bullet k}^t)$ and $\bar{p}^N = (p_{\bullet 1}^N, \ldots, p_{\bullet n}^N)$

Proof of Theorem 4. Similarly to the proof of the theorem 2 we can obtain the equation (37), but in this case we can write dawn this equation as

$$\varphi(\bar{p}^N) \leqslant \frac{1}{N} \min_{\bar{p} \geqslant 0} \left\{\sum_{t=0}^{N}\sum_{j=1}^{m} \left\langle x_{j\bullet}(p_{j\bullet}^{t+1}), p_{j\bullet}\right\rangle - \sum_{k=1}^{n} f_k(x_{\bullet k}) + g(\bar{p}) + \frac{\bar{L}_\psi}{2}\sum_{j=1}^{m} \|p_{j\bullet} - p_{j\bullet}^0\|_2^2\right\},$$

where $\bar{p}^N = (p_1^N, \ldots, p_n^N)$. Let us introduce a set $B_{3\bar{R}}^{+}(0) = \{p : p \geqslant 0, \|p\|_2 \leqslant 2\bar{R}\}$, where $\bar{R} = 3\sqrt{n}\bar{p}_{max}$ determines similarly to (38). Since we also have equation similarly to (39) we obtain that $p_{j\bullet}^t \in B_{3\bar{R}}^{+}(0)$, $\forall t$. So, we obtain

$$\varphi(\bar{p}^N) \leqslant \frac{1}{N} \min_{\bar{p} \geqslant 0} \left\{\sum_{t=0}^{N}\sum_{j=1}^{m} \left\langle x_{j\bullet}(p_{j\bullet}^{t+1}), p_{j\bullet}\right\rangle - \sum_{k=1}^{n} f_k(x_{\bullet k}) + g(\bar{p}) + \frac{\bar{L}_\psi}{2}\sum_{j=1}^{m} \|p_{j\bullet} - p_{j\bullet}^0\|_2^2\right\}$$

$$= \frac{1}{N} \min_{\bar{p}_{j\bullet} \in B_{3\bar{R}}^{+}(0)} \left\{\sum_{t=0}^{N}\sum_{j=1}^{m} \left\langle x_{j\bullet}(p_{j\bullet}^{t+1}), p_{j\bullet}\right\rangle - \sum_{k=1}^{n} f_k(x_{\bullet k}) + g(\bar{p}) + \frac{\bar{L}_\psi}{2}\sum_{j=1}^{m} \|p_{j\bullet} - p_{j\bullet}^0\|_2^2\right\}$$

And since
$$\|\mathrm{p}_{j\bullet} - \mathrm{p}_{j\bullet}^0\|_2^2 \leqslant 2\|\mathrm{p}_{j\bullet}^0\|_2^2 + 2\|\mathrm{p}_{j\bullet}\|_2^2 \leqslant 18\bar{R}^2 + \frac{2}{9}\bar{R}^2 = \frac{164}{9}\bar{R}^2$$
obtain the following
$$\varphi(\bar{\mathrm{p}}^N) \leqslant \frac{1}{N} \min_{\bar{\mathrm{p}}_{j\bullet} \in B_{3\bar{R}}^+(0)} \left\{ \sum_{t=0}^{N} \sum_{j=1}^{m} \left\langle \mathrm{x}_{j\bullet}(\mathrm{p}_{j\bullet}^{t+1}), \mathrm{p}_{j\bullet} \right\rangle - \sum_{k=1}^{n} f_k(\mathrm{x}_{\bullet k}) + g(\bar{\mathrm{p}}) \right\} + \frac{82m\bar{L}_\psi}{9N}\bar{R}^2.$$

Notice, that

$$- \frac{1}{N} \min_{\bar{\mathrm{p}}_{j\bullet} \in B_{3\bar{R}}^+(0)} \left\{ \sum_{t=0}^{N} \left(\sum_{j=1}^{m} \left\langle \mathrm{x}_{j\bullet}(\mathrm{p}_{j\bullet}^{t+1}), \mathrm{p}_{j\bullet} \right\rangle - \sum_{k=1}^{n} f_k(\mathrm{x}_{\bullet k}) + g(\bar{\mathrm{p}}) \right) \right\}$$
$$= \frac{1}{N} \sum_{t=0}^{N} \sum_{k=1}^{n} f_k(\mathrm{x}_{\bullet k}) + \frac{1}{N} \max_{\bar{\mathrm{p}}_{j\bullet} \in B_{3\bar{R}}^+(0)} \left\{ \sum_{t=0}^{N} - \sum_{j=1}^{m} \left\langle \mathrm{x}_{j\bullet}(\mathrm{p}_{j\bullet}^{t+1}), \mathrm{p}_{j\bullet} \right\rangle + \sum_{j=1}^{m} c_j \min_{k=1,\ldots,n} p_{jk} \right\}$$
$$\overset{conv-ty}{\geqslant} \sum_{k=1}^{n} f_k(\mathrm{x}_{\bullet k}^N) + \max_{\substack{\mathrm{p}_{j\bullet} \in B_{3\bar{R}}^+(0) \\ p_{j1}=\ldots=p_{jn}=p_j}} \left\{ \sum_{j=1}^{m} \left(-\left\langle \mathrm{x}_{j\bullet}^N, \mathrm{p}_{j\bullet} \right\rangle + c_j \min_{k=1,\ldots,n} p_{jk} \right) \right\}$$
$$= \sum_{k=1}^{n} f_k(\mathrm{x}_{\bullet k}^N) + \max_{\substack{\mathrm{p}_{j\bullet} \in B_{3\bar{R}}^+(0) \\ p_{j1}=\ldots=p_{jn}=p_j}} \left\{ \sum_{j=1}^{m} \left(-\sum_{k=1}^{n} x_{jk}^N p_j + c_j p_j \right) \right\}$$
$$= \sum_{k=1}^{n} f_k(\mathrm{x}_{\bullet k}^N) + \frac{3\bar{R}}{\sqrt{n}} \sum_{j=1}^{m} \left[c_j - \sum_{k=1}^{n} x_{jk}^N \right]_+ ,$$

where $\mathrm{x}_{\bullet k}^N = \frac{1}{N} \sum_{t=0}^{N} \mathrm{x}_{\bullet k}(\mathrm{p}_{\bullet k}^t)$. Thus, we obtain the following estimation

$$\varphi(\bar{\mathrm{w}}^N) + \sum_{k=1}^{n} f_k(\mathrm{x}_{\bullet k}^N) + \frac{3\bar{R}}{\sqrt{n}} \sum_{j=1}^{m} \left[c_j - \sum_{k=1}^{n} x_{jk}^N \right]_+ \leqslant \frac{82\bar{L}_\psi \bar{R}^2 m}{9N}.$$

And similarly to (41) we obtain

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

And due to the weak duality, we obtain

$$\sum_{k=1}^{n} f_k(\mathrm{x}_{\bullet k}^N) - \sum_{k=1}^{n} f_k(\mathrm{x}_{\bullet k}^*) \leqslant \frac{82\bar{L}_\psi \bar{R}^2 m}{9N}.$$

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

33

□

Composite gradient method for the resource allocation (vector case)

Input: $N > 0$ – number of steps, L_ψ – Lipschitz constant of gradient ψ, $\mathrm{p}_{j\bullet}^0$, $j = 1, \ldots, m$ – starting point.

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

$$\mathrm{x}_{\bullet k}(\mathrm{p}_{\bullet k}^{t+1}) = \operatorname*{argmax}_{\mathrm{x}_k \geqslant 0} \left\{ \sum_{j=1}^{m} p_{jk}^{t+1} x_{jk} - f_k(\mathrm{x}_k) \right\}, \quad k = 1, 2, \ldots, n.$$

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

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

and send this information to the Center.

(52)

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

- If $\sum_{k=1}^{n} \left(-\tilde{p}_{jk}^{t+1} \right)_+ \geqslant \dfrac{C}{L_\psi}$ then $p_{j.center}^{t+1} = 0$;

- Else $p_{j.center}^{t+1} > 0$ is determined from

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

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

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

7.1.2 Accelerated composite gradient method

In this subsection, to solve the resource allocation problem in vector case, we use accelerated composite gradient method from the Section 5. Similarly to the previous subsection the step of the method we can be rewritten as

$$\bar{\mathbf{y}}^{t+1} = \underset{\bar{\mathbf{p}} \geqslant 0}{\operatorname{argmin}} \left\{ \alpha_{t+1} \left(\langle \bar{\mathbf{x}}(\bar{\mathbf{p}}^{t+1}), \bar{\mathbf{p}} - \tilde{\mathbf{p}}^{t+1} \rangle + g(\mathbf{p}_{\bullet 1}, \ldots, \mathbf{p}_{\bullet n}) \right) + \frac{1}{2} \|\bar{\mathbf{p}} - \bar{\mathbf{y}}^t\|_2^2 \right\}$$

$$= \underset{\bar{\mathbf{p}} \geqslant 0}{\operatorname{argmin}} \left\{ \alpha_{t+1} \left(\sum_{k=1}^{n} \langle \mathbf{x}_{\bullet k}(\mathbf{p}_{\bullet k}^{t+1}), \mathbf{p}_{\bullet k} - \mathbf{p}_{\bullet k}^{t+1} \rangle - \sum_{j=1}^{m} c_j \min_{k=1, \ldots, n} p_{jk} \right) + \frac{1}{2} \sum_{k=1}^{n} \|\mathbf{p}_{\bullet k} - \mathbf{y}_{\bullet k}^t\|_2^2 \right\}$$

$$= \underset{\bar{\mathbf{p}} \geqslant 0}{\operatorname{argmin}} \left\{ \sum_{j=1}^{m} \left(\alpha_{t+1} \left(\langle \mathbf{x}_{j\bullet}(\mathbf{p}_{j\bullet}^{t+1}), \mathbf{p}_{j\bullet} - \mathbf{p}_{j\bullet}^{t+1} \rangle - c_j \min_{k=1, \ldots, n} p_{jk} \right) + \frac{1}{2} \|\mathbf{p}_{j\bullet} - \mathbf{y}_{j\bullet}^t\|_2^2 \right) \right\},$$

where $\mathbf{p}_{j\bullet} = (p_{j1}, \ldots, p_{jn})^\top$, 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

$$\mathbf{y}_{j\bullet}^{t+1} = \underset{\mathbf{p}_{j\bullet} \geqslant 0}{\operatorname{argmin}} \left\{ \alpha_{t+1} \left(\langle \mathbf{x}_{j\bullet}(\mathbf{p}_{j\bullet}^{t+1}), \mathbf{p}_{j\bullet} - \mathbf{p}_{j\bullet}^{t+1} \rangle - c_j \min_{k=1, \ldots, n} p_{jk} \right) + \frac{1}{2} \|\mathbf{p}_{j\bullet} - \mathbf{y}_{j\bullet}^t\|_2^2 \right\},$$

where $\mathbf{p}_{j\bullet}$ – is price vector of all producers for product j and $\mathbf{x}_{j\bullet}(\mathbf{p}_{j\bullet})$ – is vector of optimal plans for the production of product j for all producers. Define $\tilde{\mathbf{y}}_{j\bullet}^{t+1} = \mathbf{y}_{j\bullet}^t - \alpha_{t+1}\mathbf{x}_{j\bullet}(\mathbf{p}_{j\bullet}^{t+1})$ and then, using Lemma 2, we obtain the following solution:

- If $\sum_{k=1}^{n} \left(-\tilde{y}_{jk}^{t+1} \right)_+ \geqslant c_j \alpha_{t+1}$ then $y_{j.center}^{t+1} = 0$.

- Else $y_{j.center}^{t+1} > 0$ is determines from

$$\sum_{k=1}^{n} \left(y_{j.center}^{t+1} - \tilde{y}_{jk}^{t+1} \right)_+ = c_j \alpha_{t+1}$$

and the solution is determined as

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

Accelerated composite gradient descent for resource allocation (vector case)

Input: $N > 0$ – number of steps, L_ψ – Lipschitz constant of gradient ψ, $\mathbf{p}_{j\bullet}^0 = \mathbf{y}_{j\bullet}^0 = \mathbf{w}_{j\bullet}^0$, $j = 1, \ldots, m$ – starting points.

1. In the current year producers find α_{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 $\mathbf{p}_{j\bullet}^{t+1}$ for product $j = 1, \ldots, m$ as
$$p_{jk}^{t+1} = \frac{\alpha_{t+1} y_{jk}^t + A_t w_{jk}^t}{A_{t+1}}, \ \ k = 1, \ldots, n$$

   and calculate the optimal plan for the production as
$$\mathbf{x}_{\bullet k}(\mathbf{p}_{\bullet k}^{t+1}) = \operatorname*{argmax}_{\mathbf{x}_k \geqslant 0} \left\{ \sum_{j=1}^m p_{jk}^{t+1} x_{jk} - f_k(\mathbf{x}_k) \right\}, \quad k = 1, 2, \ldots, n.$$

3. Each producer predict the price for the next year $t + 1$ for product $j = 1, \ldots, m$ as
$$\tilde{y}_{jk}^{t+1} = y_{jk}^t - \alpha_{t+1} x_{jk}(\mathbf{p}_k^{t+1}), \ k = 1, \ldots, n$$

   and send this information to the Center.

4. The Center determines the prediction prices $y_{j.center}^{t+1}$ for each product $j = 1, \ldots, m$ for the next year as

   - If $\sum_{k=1}^n \left(-\tilde{y}_{jk}^{t+1} \right)_+ \geqslant c_j \alpha_{t+1}$ then $y_{j.center}^{t+1} = 0$

   - Else $y_{j.center}^{t+1} > 0$ and determines from
$$\sum_{k=1}^n \left(y_{j.center}^{t+1} - \tilde{y}_{jk}^{t+1} \right)_+ = c_j \alpha_{t+1}.$$

5. Each producer adjusts the prediction price for the next year as follows
$$y_{jk}^{t+1} = \max \left(y_{j.center}^{t+1}, \tilde{y}_{jk}^{t+1} \right), \ k = 1, \ldots, n.$$

   and calculate the historical price for product $j = 1, \ldots, m$ for the next year

36

$$w_{jk}^{t+1} = \frac{\alpha_{t+1} y_{jk}^{t+1} + A_t w_{jk}^t}{A_{t+1}}, \ k = 1, \ldots, n.$$

(53)

Theorem 5. Let Algorithm (53) be run for N steps with starting points $\mathtt{p}_{j\bullet}^0 = \mathtt{y}_{j\bullet}^0 = \mathtt{w}_{j\bullet}^0$, $j = 1, \ldots, m$ satisfying $0 \leqslant p_{jk}^0 \leqslant \bar{p}_{max}$, $k = 1, \ldots, n$, where $\bar{p}_{max}$ is given in (51). Then

$$\sum_{k=1}^{n} f_k(\mathtt{x}_{\bullet k}^N) - \sum_{k=1}^{n} f(\mathtt{x}_{\bullet k}^*) \leqslant \sum_{k=1}^{n} f_k(\mathtt{x}_{\bullet k}^N) + \varphi(\bar{\mathtt{w}}^*) \leqslant \varphi(\bar{\mathtt{w}}^N) + \sum_{k=1}^{n} f_k(\mathtt{x}_{\bullet k}^N) \leqslant \frac{148 \bar{p}_{max}^2 n^2 m}{(N+1)^2 \mu},$$

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

where $\mathtt{x}_{\bullet k}^N = \dfrac{1}{A_N} \sum_{t=0}^{N} \alpha_t \mathtt{x}_{\bullet k}(\mathtt{p}_{\bullet k}^t)$ and $\bar{\mathtt{w}}^N = (\mathtt{w}_{\bullet 1}^N, \ldots, \mathtt{w}_{\bullet n}^N)$.

Proof of Theorem 5. Similarly to the proof of the theorem 3 we can obtain the equation (44), but in this case we can write dawn this equation as

$$A_N \varphi(\bar{\mathtt{w}}^N) \leqslant \min_{\bar{\mathtt{p}} \geqslant 0} \left\{ \sum_{t=0}^{N} \alpha_t \left(\sum_{j=1}^{m} \left\langle \mathtt{x}_{j\bullet}(\mathtt{p}_{j\bullet}^{t+1}), \mathtt{p}_{j\bullet} \right\rangle - \sum_{k=1}^{n} f_k(\mathtt{x}_{\bullet k}) + g(\mathtt{p}) \right) + \frac{1}{2} \sum_{j=1}^{m} \|\mathtt{p}_{j\bullet} - \mathtt{y}_{j\bullet}^0\|_2^2 \right\},$$

where $\bar{\mathtt{w}}^N = (\mathtt{w}_1^N, \ldots, \mathtt{w}_n^N)$. Since we also have equations similarly to (39) we obtain that $\mathtt{p}_{j\bullet}^t, \mathtt{w}_{j\bullet}^t, \mathtt{y}_{j\bullet}^t \in B_{2\bar{R}}^+(0)$, $\forall t$, where $B_{2\bar{R}}^+(0) = \{\mathtt{p} : \mathtt{p} \geqslant 0, \|\mathtt{p}\|_2 \leqslant 2\bar{R}\}$, $\bar{R} = 3\sqrt{n}\bar{p}_{max}$. So, we obtain

$$A_N \varphi(\bar{\mathtt{w}}^N) \leqslant \min_{\bar{\mathtt{p}} \geqslant 0} \left\{ \sum_{t=0}^{N} \alpha_t \left(\sum_{j=1}^{m} \left\langle \mathtt{x}_{j\bullet}(\mathtt{p}_{j\bullet}^{t+1}), \mathtt{p}_{j\bullet} \right\rangle - \sum_{k=1}^{n} f_k(\mathtt{x}_{\bullet k}) + g(\mathtt{p}) \right) + \frac{1}{2} \sum_{j=1}^{m} \|\mathtt{p}_{j\bullet} - \mathtt{y}_{j\bullet}^0\|_2^2 \right\}$$

$$= \min_{\bar{\mathtt{p}}_{j\bullet} \in B_{2\bar{R}}^+(0)} \left\{ \sum_{t=0}^{N} \alpha_t \left(\sum_{j=1}^{m} \left\langle \mathtt{x}_{j\bullet}(\mathtt{p}_{j\bullet}^{t+1}), \mathtt{p}_{j\bullet} \right\rangle - \sum_{k=1}^{n} f_k(\mathtt{x}_{\bullet k}) + g(\mathtt{p}) \right) + \frac{1}{2} \sum_{j=1}^{m} \|\mathtt{p}_{j\bullet} - \mathtt{y}_{j\bullet}^0\|_2^2 \right\}$$

And since
$$\|\mathtt{p}_{j\bullet} - \mathtt{y}_{j\bullet}^0\|_2^2 \leqslant 2\|\mathtt{y}_{j\bullet}^0\|_2^2 + 2\|\mathtt{p}_{j\bullet}\|_2^2 \leqslant 8\bar{R}^2 + \frac{2}{9}\bar{R}^2 = \frac{74}{9}\bar{R}^2$$

obtain the following

$$A_N \varphi(\bar{\mathtt{w}}^N) \leqslant \min_{\bar{\mathtt{p}}_{j\bullet} \in B_{2\bar{R}}^+(0)} \left\{ \sum_{t=0}^{N} \alpha_t \left(\sum_{j=1}^{m} \left\langle \mathtt{x}_{j\bullet}(\mathtt{p}_{j\bullet}^{t+1}), \mathtt{p}_{j\bullet} \right\rangle - \sum_{k=1}^{n} f_k(\mathtt{x}_{\bullet k}) + g(\mathtt{p}) \right) \right\} + \frac{37m}{9}\bar{R}^2.$$

Notice, that

$$
-\min_{\bar{\mathtt{p}}_{j\bullet}\in B^+_{2\bar{R}}(0)} \left\{ \sum_{t=0}^{N} \alpha_t \left(\sum_{j=1}^{m} \langle \mathtt{x}_{j\bullet}(\mathtt{p}_{j\bullet}^{t+1}), \mathtt{p}_{j\bullet} \rangle - \sum_{k=1}^{n} f_k(\mathtt{x}_{\bullet k}) + g(\mathtt{p}) \right) \right\}
$$

$$
= \sum_{t=0}^{N}\sum_{k=1}^{n} f_k(\mathtt{x}_{\bullet k}) + \max_{\bar{\mathtt{p}}_{j\bullet}\in B^+_{2\bar{R}}(0)} \left\{ \sum_{t=0}^{N} \alpha_t \left(-\sum_{j=1}^{m} \langle \mathtt{x}_{j\bullet}(\mathtt{p}_{j\bullet}^{t+1}), \mathtt{p}_{j\bullet} \rangle + \sum_{j=1}^{m} c_j \min_{k=1,\ldots,n} p_{jk} \right) \right\}
$$

$$
\overset{conv-ty}{\geqslant} A_t \sum_{k=1}^{n} f_k(\mathtt{x}_{\bullet k}^N) + A_t \max_{\substack{\mathtt{p}_{j\bullet}\in B^+_{2\bar{R}}(0) \\ p_{j1}=\ldots=p_{jn}=p_j}} \left\{ \sum_{j=1}^{m} \left(-\langle \mathtt{x}_{j\bullet}^N, \mathtt{p}_{j\bullet} \rangle + c_j \min_{k=1,\ldots,n} p_{jk} \right) \right\}
$$

$$
= A_t \sum_{k=1}^{n} f_k(\mathtt{x}_{\bullet k}^N) + A_t \max_{\substack{\mathtt{p}_{j\bullet}\in B^+_{2\bar{R}}(0) \\ p_{j1}=\ldots=p_{jn}=p_j}} \left\{ \sum_{j=1}^{m} \left(-\sum_{k=1}^{n} x_{jk}^N p_j + c_j p_j \right) \right\}
$$

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

where $\mathtt{x}_{\bullet k}^N = \frac{1}{A_N} \sum_{t=0}^{N} \alpha_t \mathtt{x}_{\bullet k}(\mathtt{p}_{\bullet k}^t)$. Thus, we obtain the following estimation

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

And similarly to (46) we obtain

$$
\sum_{j=1}^{m} \left[c_j - \sum_{k=1}^{n} x_{jk}^N \right]_+ \leqslant \frac{37\bar{R} m}{15 A_N}.
$$

And due to the weak duality, we obtain

$$
\sum_{k=1}^{n} f_k(\mathtt{x}_{\bullet k}^N) - \sum_{k=1}^{n} f_k(\mathtt{x}_{\bullet k}^*) \leqslant \frac{37\bar{R}^2 m}{9A_N}.
$$

And using the estimation (48) for the A_t, $t \geqslant 0$ we obtain the statement on the Theorem.

$\square$