



university of
 groningen

faculty of mathematics
 and natural sciences

Convex Optimization in Multi-Agent Systems

Bachelor thesis in Mathematics and Applied Mathematics

July 2014

Student: I. E. Reijntjes

First Supervisor: Prof. dr. M. K. Camlibel

Second Supervisor: Prof. dr. A. C. D. van Enter

Abstract

In this thesis we are going to investigate convex optimization in the context of multi-agent systems. We represent a multi-agent system by a directed graph. If there is an edge from one vertex to another, then the second vertex receives certain information from the first vertex. We want to investigate an information exchange protocol that leads the agents to agree on a unique value that will, in turn, solve an underlying optimization problem.

Contents

1	Introduction	2
2	Preliminaries	4
2.1	Notation	4
2.2	Graph Theory	4
2.3	Convex Functions and the Subgradient	6
2.4	Kronecker Product	7
3	Main Results	8
3.1	System Dynamics	8
3.2	Reaching Consensus	10
3.3	Convergence Analysis	11
3.4	Example	19
4	Conclusion	23
A	Matlab Code	24

Chapter 1

Introduction

One of the first articles on distributed gradient optimization algorithms is dated in 1986 [5]. The types of problems that were studied are problems in which a certain number of agents try to optimize a given convex function. The problem is in discrete time and at every time step each agent performs its own computations and exchanges information with other agents. The communication between the agents is given by the system dynamics. The idea is that the states of all agents converge to a unique value that will optimize the function.

This information exchange is necessary since none of the agents has the required information to optimize the function by itself. Because the information is distributed among the agents, problems of this type are often called distributed optimization. This information exchange protocol is usually defined by a graph. If the graph is weighted, then a certain weight is assigned to the information received from the neighbors of each agent. One requirement is that the weight assigned to an agent's own state has to be significantly bigger compared to the weights assigned to the states of the neighbors.

The graph can be taken to be either time-independent or time-dependent. The time-dependent case is of course more complicated than the time-independent case because at every step the graph might change. Furthermore, the graph can be directed or undirected. If the graph is directed, then the communication between agents is not necessarily symmetric. When the graph is taken to be undirected, communication is always symmetric. As such, undirected graph topology simplifies the problem. The case where the graph is weighted, time-dependent, but undirected is considered in [4]. In this thesis we will focus on unweighted, time-independent and directed graphs.

For updating the state of each agent different methods have been studied in the past years. Our focus is on the subgradient method. With this method each agent computes a subgradient of its function. Together with the states of the neighboring agents, this subgradient is used to update the state. For this update, a stepsize is chosen for each of the agents at each iteration. For simplicity this stepsize is taken to be time-independent and the same for each agent.

The main interest in this field of study is about the convergence of the sub-gradient method. One wants to find out on what conditions the agents will reach consensus. These conditions may include for instance assumptions about the system dynamics and the sizes of the initial states and of the subgradients. Then one wants to know whether this state the agents agreed on will actually be the solution of the optimization problem. So far it has not been proven that this limit of the consensus state is the solution to the optimization problem, but one can show that the distance between this estimate and the actual optimal value is bounded.

In this research project we consider convex optimization in multi- agent systems. The situation is the following. We have a number of agents that try to find the unique vector x^* that minimizes the objective function $f(x)$. This objective function however, is not known to any of the agents. It is a convex combination of convex functions $f_i(x)$, each of which is known by agent i only.

In chapter 2 we will introduce the notation and some basic graph theory we will need. After these preliminaries we continue with describing the problem in chapter 3. We will also state the convergence results obtained in this section. The chapter ends with an example. Finally chapter 4 summarizes the conclusions.

Chapter 2

Preliminaries

In this section we will discuss some preliminaries needed in the sequel.

2.1 Notation

The notation used throughout will mainly be self-explanatory. However there are a few things we need to explain. A vector of ones i.e. $(1 \ 1 \ \dots \ 1)^\top$ will be denoted by $\mathbb{1}$.

Furthermore, a vector whose components are all non-negative and add up to 1 is called *stochastic*. If a matrix A is of the form $(a^1 \ a^2 \ \dots \ a^n)$, where each a^i is a stochastic vector, we say the matrix A is column-stochastic. Similarly A is called row-stochastic when all its rows are stochastic. Because we only use matrices that are row-stochastic, we simply refer to this property as stochastic.

If A is a matrix, then $[A]_i^j$ represents the entry in row i and column j of the matrix A .

2.2 Graph Theory

A graph is represented by $G = (V, E)$, where $V = \{1, 2, \dots, n\}$ is the set of vertices and $E \subseteq V \times V$ is the set of edges. In the case of a directed graph, the set of edges E contains ordered pairs of vertices. If the pair (i, j) is in the set E , then there is an edge from vertex i to vertex j . This does not imply that the pair (j, i) is also in the set E , since the graph is directed.

From every graph we can construct matrices which we can use in calculations and to verify certain properties of the corresponding graph. The two matrices that are of importance for us are the degree matrix and the adjacency matrix.

Definition 2.1 (Degree matrix). The degree matrix $D = d_{ij}$ is a diagonal $n \times n$ matrix of which for all $i \in \{1, 2, \dots, n\}$ the entry d_{ii} is equal to the incoming degree of vertex i .

The incoming degree of vertex i is defined as the number of edges $(j, i) \in E$, where $1 \leq j \leq n$. Another matrix we can define is the adjacency matrix A .

Definition 2.2 (Adjacency matrix). The adjacency matrix $A = a_{ij}$, is a matrix where $a_{ij} = \begin{cases} 1 & \text{if } (j, i) \in E \\ 0 & \text{otherwise.} \end{cases}$

From these matrices one can obtain the matrix M by $M = D^{-1}A$. We will use this matrix later on. The last definition we need to state is about strong connectedness.

Definition 2.3 (Strong connectedness). A directed graph is called strongly connected if for all $i \in V$, there exists a path to every vertex $j \in V$.

A path is a sequence of edges that connects a pair of vertices. For instance if we have $(i, j), (j, k) \in E$, then the sequence $\{(i, j), (j, k)\}$ is a path from vertex i to vertex j .

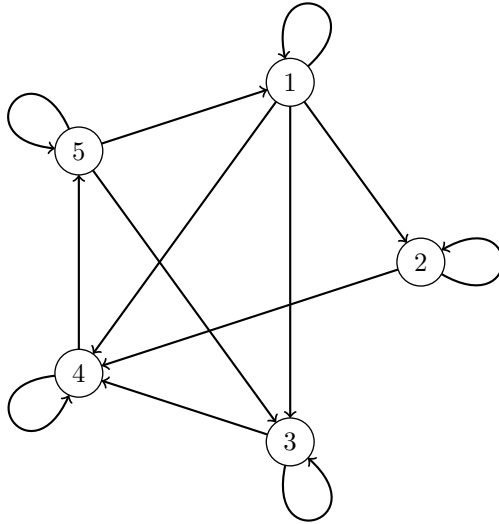


Figure 2.1: A strongly connected directed graph

Example 2.4. The graph in figure 2.1 is defined by $G = (V, E)$ where $V = \{1, 2, 3, 4, 5\}$ and $E = \{(1, 1), (1, 2), (1, 3), (1, 4), (2, 2), (2, 4), (3, 3), (3, 4), (4, 4), (4, 5), (5, 1), (5, 3), (5, 5)\}$. It can be checked that there exists a path from every vertex to every other vertex, so the graph is strongly connected. For this graph the degree matrix D and adjacency matrix A are:

$$D = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}, \quad A = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

Using these matrices we obtain M by $M = D^{-1}A$:

$$M = \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} & \frac{1}{2} \end{pmatrix}.$$

This property of a graph being strongly connected can be deduced from the matrix M defined above, in the following way [1]:

Theorem 2.5. *A directed graph G is strongly connected if and only if the corresponding matrix $M = D^{-1}A$ is irreducible, i.e. if for every $1 \leq i, j \leq n$, there exists a natural number q such that $[M^q]_i^j > 0$.*

2.3 Convex Functions and the Subgradient

In this section we will define both what a convex function is and what a subgradient is. These definitions can be found for instance in [2].

Definition 2.6 (Convex set). A set C is convex if for any $x, y \in C$ and any θ such that $0 \leq \theta \leq 1$, we have that $\theta x + (1 - \theta)y \in C$.

Hence, a set C is convex if the line segment between any two points in C is again in C .

Definition 2.7 (Convex function). A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is called a convex function if:

1. its domain is a convex set;
2. for all x and y in the domain of f and for all θ such that $0 \leq \theta \leq 1$, we have that

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y). \quad (2.1)$$

This means that the line segment between $(x, f(x))$ and $(y, f(y))$, lies on or above the graph of f for all x and y . Moreover, a function is *strictly convex* if there is a strict inequality in equation (2.1), whenever $x \neq y$ and $0 < \theta < 1$. From this definition, we can deduce that if a strictly convex function has a minimum, then this minimum is unique. Because of this property strictly convex functions are useful in optimization problems.

Definition 2.8 (Subgradient). The vector d is a subgradient of the function f at x if: $d^\top (y - x) \leq f(y) - f(x)$, for all y .

If the function $f(x)$ is differentiable at x we can take $d = \nabla f(x)$.

2.4 Kronecker Product

The Kronecker product is defined in the following way:

Definition 2.9 (Kronecker product). Let $A = a_{ij}$ be an $m \times n$ matrix and let $B = b_{ij}$ be a $p \times q$ matrix. Then for the Kronecker product we have:

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{pmatrix}.$$

Hence $A \otimes B$ is an $mp \times nq$ matrix.

There are two properties of the Kronecker product we will need later on. These are both stated below.

Property 2.10 (Associativity). The Kronecker product is associative, i.e. $(A \otimes B) \otimes C = A \otimes (B \otimes C)$.

Property 2.11 (Mixed-product property). If the matrices A, B, C and D have appropriate sizes such that AC and BD exist, then $(A \otimes B)(C \otimes D) = AC \otimes BD$. Particularly, if $C = A$ and $D = B$, we have $(A \otimes B)^2 = A^2 \otimes B^2$.

Chapter 3

Main Results

We have n agents communicating with each other via a certain protocol. Together they want to minimize a convex function but each agent only has information about one component of the function. Each component is a convex function itself from $\mathbb{R}^p \rightarrow \mathbb{R}$. So each agent minimizes its own component and exchanges information with its neighboring agents. We want to investigate if the agents will reach consensus and whether the solution found, is indeed the minimum. So the problem is:

$$\text{minimize } f(x), \tag{3.1}$$

subject to $x \in \mathbb{R}^p$, where $f(x)$ is a convex combination of $f_i(x)$, that will be determined by the dynamics of the multi-agent system. In turn each $f_i(x) : \mathbb{R}^p \rightarrow \mathbb{R}$ is a convex function, and is known to agent i only. We denote the optimal value by f^* , and assume that this is finite. Furthermore the optimal solution set is denoted by $X^* = \{x \in \mathbb{R}^p | f(x) = f^*\}$.

3.1 System Dynamics

We represent a multi-agent system by a directed graph $G = (V, E)$, where $V = \{1, 2, \dots, n\}$ is the set of vertices (also called the agents) and $E \subseteq V \times V$ is the set of edges. We require the graph G to be strongly connected because the state of each agent has to reach every other agent. For every vertex i we must have that $(i, i) \in E$ and if there is an edge from vertex j to vertex i , then $(j, i) \in E$. So if $(j, i) \in E$ then agent i receives information from agent j , and therefore we say agent j is a neighbor of agent i . The set of neighbors of vertex i is represented by:

$$N_i = \{j \in V | (j, i) \in E\}.$$

Next, we define the dynamics of the multi-agent system that we will work with. All agents compute a state for each time k . This is the estimate for the optimal value at that time. It is a vector in \mathbb{R}^p . Every agent i updates its state according to the following formula:

$$x^i(k+1) = \text{avg}\{x^j(k) | j \in N_i\} - \alpha d^i(k),$$

where $\text{avg}\{x^j(k)|j \in N_i\}$ represents taking the average of the states of all neighbors of vertex i . The constant α is the stepsize, which, for the sake of brevity, is taken to be equal for each agent i . Lastly the vector $d^i(k)$ is a subgradient of f_i at $x^i(k)$. By defining

$$x(k) = \begin{pmatrix} x^1(k) \\ x^2(k) \\ \vdots \\ x^n(k) \end{pmatrix} \text{ and } d(k) = \begin{pmatrix} d^1(k) \\ d^2(k) \\ \vdots \\ d^n(k) \end{pmatrix},$$

we arrive at the following compact form for the dynamics:

$$x(k+1) = (M \otimes I_p)x(k) - \alpha d(k). \quad (3.2)$$

Here the matrix M is a stochastic $n \times n$ matrix, representing the averages mentioned before. It can be obtained from the degree matrix and the adjacency matrix defined in section 2.2 in the way that $M = D^{-1}A$. The matrix I_p is the $p \times p$ identity matrix. So the Kronecker product $M \otimes I_p$ is an $np \times np$ matrix.

By using equation (3.2) we can write $x(k)$ as a function of $x(0)$, α and $d(k)$. We have:

$$\begin{aligned} x(1) &= (M \otimes I_p)x(0) - \alpha d(0) \\ x(2) &= (M \otimes I_p)x(1) - \alpha d(1) \\ &= (M \otimes I_p)[(M \otimes I_p)x(0) - \alpha d(0)] - \alpha d(1) \\ &= (M \otimes I_p)^2 x(0) - \alpha(M \otimes I_p)d(0) - \alpha d(1) \\ &= (M \otimes I_p)^2 x(0) - \alpha[(M \otimes I_p)d(0) + d(1)] \\ x(3) &= (M \otimes I_p)x(2) - \alpha d(2) \\ &= (M \otimes I_p)[(M \otimes I_p)^2 x(0) - \alpha(M \otimes I_p)d(0) - \alpha d(1)] - \alpha d(2) \\ &= (M \otimes I_p)^3 x(0) - \alpha(M \otimes I_p)^2 d(0) - \alpha(M \otimes I_p)d(1) - \alpha d(2) \\ &= (M \otimes I_p)^3 x(0) - \alpha[(M \otimes I_p)^2 d(0) + (M \otimes I_p)d(1) + d(2)] \\ &\vdots \\ x(k) &= (M \otimes I_p)^k x(0) - \alpha(M \otimes I_p)^{k-1} d(0) - \alpha(M \otimes I_p)^{k-2} d(1) \\ &\quad - \dots - \alpha(M \otimes I_p)d(k-2) - \alpha d(k-1) \\ &= (M \otimes I_p)^k x(0) - \alpha[(M \otimes I_p)^{k-1} d(0) + (M \otimes I_p)^{k-2} d(1) \\ &\quad + \dots + (M \otimes I_p)d(k-2) + d(k-1)] \\ &= (M \otimes I_p)^k x(0) - \alpha \sum_{s=1}^k (M \otimes I_p)^{k-s} d(s-1). \end{aligned}$$

So for general k we get:

$$x(k) = (M \otimes I_p)^k x(0) - \alpha \sum_{s=1}^k (M \otimes I_p)^{k-s} d(s-1). \quad (3.3)$$

We can now use property 2.11 to simplify $(M \otimes I_p)^k$ by writing $(M \otimes I_p)^k = M^k \otimes I_p$. Hence equation (3.3) reduces to:

$$x(k) = (M^k \otimes I_p)x(0) - \alpha \sum_{s=1}^k (M^{k-s} \otimes I_p)d(s-1). \quad (3.4)$$

3.2 Reaching Consensus

We want the state of all agents to converge to the same limit. Therefore we are interested in the behaviour of the system (3.2) when $k \rightarrow \infty$. To investigate this, we need to know what $\lim_{k \rightarrow \infty} M^k$ is.

Lemma 3.1. Suppose that G is strongly connected. Then, there exists a vector $m \in \mathbb{R}^n$ such that $\lim_{k \rightarrow \infty} M^k = \mathbb{1}m^\top$.

Proof. We will first state two theorems we will be needing for the proof.

Theorem 3.2 (Gershgorin). *Let $M = (m_{ij})$ be a complex $n \times n$ matrix. For $1 \leq i \leq n$, let $R_i = \sum_{j \neq i} |m_{ij}|$. Let $D(m_{ii}, R_i)$ be the disc with radius R_i centred at m_{ii} . Then every eigenvalue of M is in one of the discs D .*

Theorem 3.3 (Perron-Frobenius). [3] *Let $M = (m_{ij})$ be an $n \times n$ positive matrix, i.e. $m_{ij} > 0$, for all $1 \leq i, j \leq n$.*

1. *There is a positive real number r , such that r is an eigenvalue of M and for any other eigenvalue λ we have $|\lambda| < r$. Thus the spectral radius is $\rho(M) = r$;*
2. *For the largest eigenvalue r , there exist a corresponding right eigenvector v and a corresponding left eigenvector w^\top , with the property $w^\top v = 1$, such that $\lim_{k \rightarrow \infty} \frac{M^k}{r^k} = vw^\top$;*

Moreover, if M is a non-negative matrix, i.e. $m_{ij} \geq 0$ for all $1 \leq i, j \leq n$, then the above statements still hold if and only if the matrix M is irreducible.

Since M is a stochastic matrix, it holds that $M\mathbb{1} = \mathbb{1}$. Therefore $\mu = 1$ is an eigenvalue of M with corresponding eigenvector $\mathbb{1}$. From theorem 3.2 and the fact that M is a stochastic matrix, it follows that for all eigenvalues λ of M we have $|\lambda| \leq 1$. In order to be able to use theorem 3.3 we need a positive matrix, but the matrix M is not necessarily a positive matrix. Since $m_{ij} \geq 0$, for all $1 \leq i, j \leq n$, M is merely a non-negative matrix. According to theorem 3.3 we require M to be irreducible. From theorem 2.5 we know that this is the case if and only if the corresponding graph G is strongly connected. Since we need the graph G to have this property, we are allowed to use theorem 3.3.1. It follows that each eigenvalue λ is strictly smaller than our largest eigenvalue $\mu = 1$. Now for the final step in the proof we use theorem 3.3.2. Because $r = 1$ the limit stated reduces to $\lim_{k \rightarrow \infty} M^k = vw^\top$. Since the eigenvectors v and w^\top have to satisfy $w^\top v = 1$, and we know that $v = \mathbb{1}$, w has to be a stochastic vector. As $w^\top = (w_1 \ w_2 \ \dots \ w_n)$, $w^\top v = w_1 + w_2 + \dots + w_n = 1$. Hence

indeed $\lim_{k \rightarrow \infty} M^k = \mathbb{1}m^\top$, where in fact m^\top is the normalized left eigenvector of M belonging to the eigenvalue $\mu = 1$. \square

The following lemma gives us the convergence rate for the limit stated in lemma 3.1. We will need this later on and will refer back when necessary.

Lemma 3.4. Let m be the vector of lemma 3.1, then for all $1 \leq i \leq n$ and for all $s \geq 0$ we have that:

$$\left| \left[(M^\top)^{k-s+1} \right]_i^j - m_i \right| \leq 2\gamma \delta^{\frac{k-s}{n-1}}, \quad (3.5)$$

for all $k \geq s$ and $1 \leq j \leq n$, where $\gamma = \frac{1 + (\frac{1}{n})^{-(n-1)}}{1 - (\frac{1}{n})^{n-1}}$, and $\delta = \left(1 - (\frac{1}{n})^{n-1}\right)$.

Of course this same bound holds for $\left| [M^{k-s+1}]_i^j - [m^\top]_i \right|$.

Proof. The proof of this lemma follows from lemma 4 of [4]. \square

3.3 Convergence Analysis

Suppose that after some time $t = \bar{k}$ we have that $d^i(k) = 0$, for all i , and for all k such that $k \geq \bar{k}$. We define $\bar{x}(k)$ to be the sequence of states generated by the agents. Then from equation (3.4) it follows:

$$\bar{x}(k) = (M^k \otimes I_p)x(0) - \alpha \sum_{s=1}^{\bar{k}} (M^{k-s} \otimes I_p)d(s-1),$$

since for all $k \geq \bar{k}$, $d(k) = 0$. We can now take the limit $k \rightarrow \infty$ on both sides to obtain:

$$\lim_{k \rightarrow \infty} \bar{x}(k) = \lim_{k \rightarrow \infty} \left[(M^k \otimes I_p)x(0) - \alpha \sum_{s=1}^{\bar{k}} (M^{k-s} \otimes I_p)d(s-1) \right];$$

$$\lim_{k \rightarrow \infty} \bar{x}(k) = \lim_{k \rightarrow \infty} (M^k \otimes I_p)x(0) - \alpha \lim_{k \rightarrow \infty} \sum_{s=1}^{\bar{k}} (M^{k-s} \otimes I_p)d(s-1).$$

Using lemma 3.1 this reduces to:

$$\lim_{k \rightarrow \infty} \bar{x}(k) = (\mathbb{1}m^\top \otimes I_p)x(0) - \alpha \sum_{s=1}^{\bar{k}} (\mathbb{1}m^\top \otimes I_p)d(s-1). \quad (3.6)$$

From equation (3.6), we can deduce that the state of all agents converges to the same value. This allows us to write the following relation that holds for all i :

$$\lim_{k \rightarrow \infty} \bar{x}^i(k) = (m^\top \otimes I_p)x(0) - \alpha \sum_{s=1}^{\bar{k}} (m^\top \otimes I_p)d(s-1).$$

In this equation both $x(0)$ and $d(s-1)$ are vectors in \mathbb{R}^{np} . So we now know that the agents will agree on a unique limit value. Hence, what remains to be shown, is that this limit value is indeed the solution of the underlying optimization problem.

We now define $y(\bar{k}) = \lim_{k \rightarrow \infty} \bar{x}^i(k)$:

$$y(\bar{k}) = (m^\top \otimes I_p)x(0) - \alpha \sum_{s=1}^{\bar{k}} (m^\top \otimes I_p)d(s-1).$$

This equation we obtained must hold for any \bar{k} and since this is the only dependent variable we can rename the index, and use k instead. This yields:

$$y(k) = (m^\top \otimes I_p)x(0) - \alpha \sum_{s=1}^k (m^\top \otimes I_p)d(s-1). \quad (3.7)$$

and since it holds for any k we can also state:

$$y(k+1) = (m^\top \otimes I_p)x(0) - \alpha \sum_{s=1}^{k+1} (m^\top \otimes I_p)d(s-1). \quad (3.8)$$

Now by subtracting equation (3.7) from equation (3.8) we obtain the following iterative relation:

$$y(k+1) - y(k) = -\alpha (m^\top \otimes I_p) d(k),$$

or equivalently

$$y(k+1) = y(k) - \alpha (m^\top \otimes I_p) d(k). \quad (3.9)$$

This formula above can be written without the Kronecker product in the following way:

$$y(k+1) = y(k) - \alpha \sum_{r=1}^n m_r d^r(k). \quad (3.10)$$

Now we want to investigate whether the sequence $\{y(k)\}$ actually converges to the optimal value in X^* . Hence, we are interested in the minimal distance between $y(k)$ and any $x \in X^*$. In order to set a bound for this distance, we will first look at the distance from $y(k)$ to any $x \in \mathbb{R}^p$ and later, by letting $x \in X^*$ and taking the infimum we can set a bound for the distance between $y(k)$ and the optimal solution set X^* . From here on we will use the notation

$$f(x) = \sum_{r=1}^n m_r f_r(x). \quad (3.11)$$

This function $f(x)$ is the objective function from (3.1) we want to minimize.

The following lemma sets a bound for the distance between the estimate vector $y(k)$ and the optimal vector $x \in X^*$.

Lemma 3.5. Let the sequence $\{y(k)\}$ be generated by iteration (3.10), and the sequence $\{x(k)\}$ by iteration (3.4). Let $\{g^r(k)\}$ be a sequence of subgradients of f_r at $y(k)$. Then:

1. For any $x \in \mathbb{R}^p$ and for all $k \geq 0$, we have:

$$\begin{aligned} \|y(k+1) - x\|^2 &\leq \|y(k) - x\|^2 \\ &\quad + 2\alpha \sum_{r=1}^n m_r (\|d^r(k)\| + \|g^r(k)\|) \|y(k) - x^r(k)\| \\ &\quad - 2\alpha [f(y(k)) - f(x)] + \alpha^2 \|d(k)\|^2. \end{aligned} \quad (3.12)$$

2. When the optimal solution set X^* is non-empty, then for all $k \geq 0$ we have:

$$\begin{aligned} \text{dist}^2(y(k+1), X^*) &\leq \text{dist}^2(y(k), X^*) \\ &\quad + 2\alpha \sum_{r=1}^n m_r (\|d^r(k)\| + \|g^r(k)\|) \|y(k) - x^r(k)\| \\ &\quad - 2\alpha [f(y(k)) - f^*] + \alpha^2 \|d(k)\|^2. \end{aligned} \quad (3.13)$$

Here $\text{dist}^2(y, X)$ is the least distance from the vector y to the set X .

Proof. Using relation (3.9), we obtain:

$$\|y(k+1) - x\|^2 = \left\| y(k) - x - \alpha \sum_{r=1}^n m_r d^r(k) \right\|^2,$$

implying that

$$\begin{aligned} \|y(k+1) - x\|^2 &\leq \|y(k) - x\|^2 - 2\alpha \sum_{r=1}^n m_r [d^r(k)]^\top [y(k) - x] \\ &\quad + \alpha^2 \left\| \sum_{r=1}^n m_r d^r(k) \right\|^2. \end{aligned} \quad (3.14)$$

We will now consider each of the terms on the right hand side individually. Let us first investigate the second term. In particular we focus on $[d^r(k)]^\top [y(k) - x]$. For any $1 \leq r \leq n$, we have:

$$\begin{aligned} [d^r(k)]^\top [y(k) - x] &= [d^r(k)]^\top [y(k) - x^r(k)] + [d^r(k)]^\top [x^r(k) - x] \\ &\geq -\|d^r(k)\| \|y(k) - x^r(k)\| + [d^r(k)]^\top [x^r(k) - x], \end{aligned} \quad (3.15)$$

where the inequality follows from Cauchy-Schwarz inequality. Since $d^r(k)$ is a subgradient of f_r at $x^r(k)$ we are allowed to write:

$$[d^r(k)]^\top [x^r(k) - x] \geq f_r(x^r(k)) - f_r(x),$$

for all $1 \leq r \leq n$ and for all $x \in \mathbb{R}^n$. Next we will use a subgradient $g^r(k)$ of f_r at $y(k)$, which yields:

$$\begin{aligned} f_r(x^r(k)) - f_r(x) &= f_r(x^r(k)) - f_r(y(k)) + f_r(y(k)) - f_r(x) \\ &\geq [g^r(k)]^\top (x^r(k) - y(k)) + f_r(y(k)) - f_r(x) \\ &\geq -\|g^r(k)\| \|x^r(k) - y(k)\| + f_r(y(k)) - f_r(x). \end{aligned} \quad (3.16)$$

Combining inequality (3.15) and inequality (3.16) yields:

$$\begin{aligned} [d^r(k)]^\top [y(k) - x] &\geq -(\|d^r(k)\| + \|g^r(k)\|) \|y(k) - x^r(k)\| \\ &\quad + f_r(y(k)) - f_r(x). \end{aligned}$$

Multiplying both sides by m_r and summing over all r gives us:

$$\begin{aligned} \sum_{r=1}^n m_r [d^r(k)]^\top [y(k) - x] &\geq -\sum_{r=1}^n m_r (\|d^r(k)\| + \|g^r(k)\|) \|y(k) - x^r(k)\| \\ &\quad + [f(y(k)) - f(x)], \end{aligned} \quad (3.17)$$

where the last term follows from equation (3.11). Now for the third term in inequality (3.14), we have:

$$\left\| \sum_{r=1}^n m_r d^r(k) \right\|^2 = \|m^\top d(k)\|^2 \leq \|m\|^2 \|d(k)\|^2 \leq \|d(k)\|^2, \quad (3.18)$$

where the last inequality follows from the fact that the vector m is stochastic.

Now we can use both equation (3.17) and equation (3.18) to obtain the bound stated in equation (3.12):

$$\begin{aligned} \|y(k+1) - x\|^2 &\leq \|y(k) - x\|^2 + 2\alpha \sum_{r=1}^n m_r (\|d^r(k)\| + \|g^r(k)\|) \|y(k) - x^r(k)\| \\ &\quad - 2\alpha [f(y(k)) - f(x)] + \alpha^2 \|d(k)\|^2. \end{aligned}$$

As stated before, the proof of lemma 3.5.2 follows by letting $x \in X^*$ and taking the infimum. Hence:

$$\begin{aligned} \text{dist}^2(y(k+1), X^*) &\leq \text{dist}^2(y(k), X^*) \\ &\quad + 2\alpha \sum_{r=1}^n m_r (\|d^r(k)\| + \|g^r(k)\|) \|y(k) - x^r(k)\| \\ &\quad - 2\alpha [f(y(k)) - f^*] + \alpha^2 \|d(k)\|^2. \end{aligned} \quad (3.19)$$

□

We see that the distance from $y(k+1)$ to the optimal solution set X^* , is bounded by the distance of $y(k)$ to set X^* and some other terms, proportional to α . Therefore by choosing the stepsize α smaller, the bound will become stricter.

By rewriting this inequality (3.19) we can obtain a certain bound for $f(y(k)) - f^*$. We have:

$$\begin{aligned} f(y(k)) - f^* &\leq \frac{1}{2\alpha} (\text{dist}^2(y(k), X^*) - \text{dist}^2(y(k+1), X^*)) \\ &\quad + \sum_{r=1}^n m_r (\|d^r(k)\| + \|g^r(k)\|) \|y(k) - x^r(k)\| \\ &\quad + \frac{\alpha}{2} \|d(k)\|^2. \end{aligned}$$

Since this inequality holds for all k we can say:

$$\begin{aligned} f(y(0)) - f^* &\leq \frac{1}{2\alpha} (\text{dist}^2(y(0), X^*) - \text{dist}^2(y(1), X^*)) \\ &\quad + \sum_{r=1}^n m_r (\|d^r(0)\| + \|g^r(0)\|) \|y(0) - x^r(0)\| \\ &\quad + \frac{\alpha}{2} \|d(0)\|^2 \\ f(y(1)) - f^* &\leq \frac{1}{2\alpha} (\text{dist}^2(y(1), X^*) - \text{dist}^2(y(2), X^*)) \\ &\quad + \sum_{r=1}^n m_r (\|d^r(1)\| + \|g^r(1)\|) \|y(1) - x^r(1)\| \\ &\quad + \frac{\alpha}{2} \|d(1)\|^2 \\ &\quad \vdots \\ f(y(k-1)) - f^* &\leq \frac{1}{2\alpha} (\text{dist}^2(y(k-1), X^*) - \text{dist}^2(y(k), X^*)) \\ &\quad + \sum_{r=1}^n m_r (\|d^r(k-1)\| + \|g^r(k-1)\|) \|y(k-1) - x^r(k-1)\| \\ &\quad + \frac{\alpha}{2} \|d(k-1)\|^2. \end{aligned}$$

Summing all these up and omitting the non-positive terms yields:

$$\begin{aligned} \sum_{h=0}^{k-1} f(y(h)) - kf^* &\leq \frac{1}{2\alpha} \text{dist}^2(y(0), X^*) \\ &\quad + \sum_{h=0}^{k-1} \left(\sum_{r=1}^n m_r (\|d^r(h)\| + \|g^r(h)\|) \|y(h) - x^r(h)\| \right) \\ &\quad + \frac{\alpha}{2} \sum_{h=0}^{k-1} \|d(h)\|^2. \end{aligned} \tag{3.20}$$

If the bound on the right hand side becomes very small, then we know that our estimate $y(k)$ is very close to the solution of the minimization problem, since $f(y(k))$ is close to f^* .

We need to set bounds for each of the terms on the right hand side. The first term is already bounded, so that leaves us with the second and third term. The next two propositions provide us with bounds for these terms.

Proposition 3.6. Assume that for each $1 \leq r \leq n$, the sequences of subgradients $\{d^r(k)\}$ and $\{g^r(k)\}$ are bounded. I.e. there exists a scalar L , such that $\max\{\|d^r(k)\|, \|g^r(k)\|\} \leq L$, for all $k \geq 0$. Then $\|d(h)\|^2 \leq nL^2$ and $\|g(h)\|^2 \leq nL^2$.

Proof. This is easily verified in the following way:

$$\|d(h)\|^2 = \|d^1(h)\|^2 + \|d^2(h)\|^2 + \dots + \|d^n(h)\|^2 \leq L^2 + L^2 + \dots + L^2 = nL^2.$$

The same holds for $\|g(h)\|^2$. \square

Next our goal is to find a uniform bound for $\|y(k) - x^r(h)\|$ that will hold for all $1 \leq r \leq n$. Since $\sum_{r=1}^n m_r \|y(k) - x^r(k)\| \leq \sum_{r=1}^n \|y(k) - x^r(k)\| \leq \sum_{r=1}^n \|y(k) - x^r(k)\|^2 = \|\bar{y}(k) - x(k)\|$, where $\bar{y}(k) = (y(k) \ y(k) \ \dots \ y(k))^\top = \mathbb{1} \otimes y(k)$, it is sufficient to have a bound for $\|\bar{y}(k) - x(k)\|$. In the next proposition, we will show that this term is indeed bounded:

Proposition 3.7. For all $k \geq 0$, a uniform upper bound for $\|\bar{y}(k) - x(k)\|$ is given by:

$$\|\bar{y}(k) - x(k)\| \leq C, \quad (3.21)$$

$$\text{where } C = \alpha p \sqrt{n} L \left(2n^2 \gamma + \frac{2n^2 \gamma}{1 - \delta^{\frac{1}{n-1}}} \right).$$

Proof. Substituting the equation for $y(k)$ obtained in (3.7), in $\bar{y}(k) = \mathbb{1} \otimes y(k)$, yields

$$\bar{y}(k) = \mathbb{1} \otimes \left[(m^\top \otimes I_p) x(0) - \alpha \sum_{s=1}^k (m^\top \otimes I_p) d(s-1) \right].$$

By using the associativity of the Kronecker product, and the fact that:

$$\mathbb{1} \otimes m^\top = \begin{pmatrix} m^\top \\ m^\top \\ \vdots \\ m^\top \end{pmatrix} = \mathbb{1} m^\top,$$

we can say:

$$\mathbb{1} \otimes (m^\top \otimes I_p) = (\mathbb{1} \otimes m^\top) \otimes I_p = \mathbb{1} m^\top \otimes I_p.$$

Then for $\bar{y}(k)$ we obtain:

$$\bar{y}(k) = (\mathbb{1} m^\top \otimes I_p) x(0) - \alpha \sum_{s=1}^k (\mathbb{1} m^\top \otimes I_p) d(s-1). \quad (3.22)$$

When we use equations (3.22) and (3.4), for the distance between $\bar{y}(k)$ and $x(k)$ we get:

$$\|\bar{y}(k) - x(k)\| = \left\| \left[(\mathbb{1} m^\top \otimes I_p) x(0) - \alpha \sum_{s=1}^k (\mathbb{1} m^\top \otimes I_p) d(s-1) \right] - \left[(M^k \otimes I_p) x(0) - \alpha \sum_{s=1}^k (M^{k-s} \otimes I_p) d(s-1) \right] \right\|$$

$$= \left\| \left[(\mathbb{1}m^\top - M^k) \otimes I_p \right] x(0) - \alpha \sum_{s=1}^k \left[(\mathbb{1}m^\top - M^{k-s}) \otimes I_p \right] d(s-1) \right\|. \quad (3.23)$$

By using the triangle inequality on equation (3.23) we obtain:

$$\begin{aligned} \|\bar{y}(k) - x(k)\| &\leq \left\| \left[(\mathbb{1}m^\top - M^k) \otimes I_p \right] x(0) \right\| \\ &\quad + \alpha \left\| \sum_{s=1}^k \left[(\mathbb{1}m^\top - M^{k-s}) \otimes I_p \right] d(s-1) \right\|. \end{aligned}$$

Now we consider both terms on the right hand side individually. We have:

$$\begin{aligned} \left\| \left[(\mathbb{1}m^\top - M^k) \otimes I_p \right] x(0) \right\| &\leq \left\| (\mathbb{1}m^\top - M^k) \otimes I_p \right\| \|x(0)\| \\ &\leq p \left\| \mathbb{1}m^\top - M^k \right\| \|x(0)\|. \end{aligned} \quad (3.24)$$

And similarly for the second term:

$$\begin{aligned} &\alpha \left\| \sum_{s=1}^k \left[(\mathbb{1}m^\top - M^{k-s}) \otimes I_p \right] d(s-1) \right\| \\ &\leq \alpha \sum_{s=1}^k \left\| (\mathbb{1}m^\top - M^{k-s}) \otimes I_p \right\| \|d(s-1)\| \\ &\leq \alpha p \sum_{s=1}^k \left\| \mathbb{1}m^\top - M^{k-s} \right\| \|d(s-1)\|. \end{aligned} \quad (3.25)$$

Combining inequalities (3.24) and (3.25) we obtain:

$$\|\bar{y}(k) - x(k)\| \leq p \left\| \mathbb{1}m^\top - M^k \right\| \|x(0)\| + \alpha p \sum_{s=1}^k \left\| \mathbb{1}m^\top - M^{k-s} \right\| \|d(s-1)\|. \quad (3.26)$$

We need make an assumption about the size of the vector $x(0)$, in order to set a bound. Therefore we assume that for all $1 \leq i \leq n$ we have that $\|x^i(0)\| \leq \alpha L$, where L is the bound of the subgradients. Then we have:

$$\begin{aligned} \|x(0)\| &= \sqrt{\|x^1(0)\|^2 + \|x^2(0)\|^2 + \dots + \|x^n(0)\|^2} \\ &\leq \sqrt{(\alpha L)^2 + (\alpha L)^2 + \dots + (\alpha L)^2} = \sqrt{n} \alpha L. \end{aligned}$$

With this information and by using the boundedness of $\|d(k)\|$ equation (3.26) becomes:

$$\begin{aligned} \|\bar{y}(k) - x(k)\| &\leq \alpha p \sqrt{n} L \left\| \mathbb{1}m^\top - M^k \right\| + \alpha p \sqrt{n} L \sum_{s=1}^k \left\| (\mathbb{1}m^\top - M^{k-s}) \right\| \\ &= \alpha p \sqrt{n} L \left(\left\| \mathbb{1}m^\top - M^k \right\| + \sum_{s=1}^k \left\| (\mathbb{1}m^\top - M^{k-s}) \right\| \right). \end{aligned} \quad (3.27)$$

To continue from here, we use lemma 3.4. For the first term of the right hand side of equation (3.27) we have:

$$\|\mathbb{1}m^\top - M^k\| \leq n^2 \left| [M^k]_i^j - [m^\top]_i \right| \leq 2n^2\gamma\delta^{\frac{k-1}{n-1}},$$

which, since $0 < \delta^{\frac{k-1}{n-1}} < 1$, reduces to:

$$\|\mathbb{1}m^\top - M^k\| < 2n^2\gamma. \quad (3.28)$$

For the second term we obtain:

$$\sum_{s=1}^k \|(\mathbb{1}m^\top - M^{k-s})\| \leq n^2 \sum_{s=1}^k \left| [M^{k-s}]_i^j - [m^\top]_i \right| \leq 2n^2\gamma \sum_{s=1}^k \delta^{\frac{k-s}{n-1}}. \quad (3.29)$$

For the sum in the equation above we have:

$$\sum_{s=1}^k \delta^{\frac{k-s}{n-1}} = \sum_{s=1}^k \left(\delta^{\frac{1}{n-1}}\right)^{k-s} = \sum_{s=0}^{k-1} \left(\delta^{\frac{1}{n-1}}\right)^s = \frac{\left(\delta^{\frac{1}{n-1}}\right)^k - 1}{\delta^{\frac{1}{n-1}} - 1}.$$

So we also have:

$$\sum_{s=1}^k \delta^{\frac{k-s}{n-1}} \leq \sum_{s=1}^{\infty} \delta^{\frac{k-s}{n-1}} = \frac{1}{1 - \delta^{\frac{1}{n-1}}}. \quad (3.30)$$

Hence, combining inequality (3.29) with (3.30) yields:

$$\sum_{s=1}^k \|(\mathbb{1}m^\top - M^{k-s})\| \leq \frac{2n^2\gamma}{1 - \delta^{\frac{1}{n-1}}}. \quad (3.31)$$

By using inequalities (3.28) and (3.31), equation (3.27) reduces to:

$$\|\bar{y}(k) - x(k)\| \leq C = \alpha p \sqrt{n} L \zeta, \text{ where } \zeta = \left(2n^2\gamma + \frac{2n^2\gamma}{1 - \delta^{\frac{1}{n-1}}}\right),$$

which is what we wanted to show. \square

Proposition 3.8. Let $\hat{y}(k)$ be the averaged vector, i.e. $\hat{y}(k) = \frac{1}{k} \sum_{h=0}^{k-1} y(h)$. Then we have:

$$f(\hat{y}(k)) \leq f^* + \frac{1}{2\alpha k} \text{dist}^2(y(0), X^*) + 2LC + \frac{1}{2}\alpha n L^2. \quad (3.32)$$

Proof. By using propositions 3.6 and 3.7, equation (3.20) reduces to:

$$\sum_{h=0}^{k-1} f(y(h)) - k f^* \leq \frac{1}{2\alpha} \text{dist}^2(y(0), X^*) + 2kLC + \frac{\alpha n k L^2}{2},$$

or equivalently, if we divide both sides by k :

$$\frac{1}{k} \sum_{h=0}^{k-1} f(y(h)) - f^* \leq \frac{1}{2\alpha k} \text{dist}^2(y(0), X^*) + 2LC + \frac{1}{2}\alpha n L^2. \quad (3.33)$$

Since the function $f(x)$ is a convex function, from (2.1) it follows that:

$$\frac{1}{k} \sum_{h=0}^{k-1} f(y(h)) \geq f\left(\frac{1}{k} \sum_{h=0}^{k-1} y(h)\right) = f(\hat{y}(k)).$$

Hence equation (3.33) becomes:

$$f(\hat{y}(k)) - f^* \leq \frac{1}{2\alpha k} \text{dist}^2(y(0), X^*) + 2LC + \frac{1}{2} \alpha n L^2.$$

Therefore we have shown what we wanted to prove:

$$f(\hat{y}(k)) \leq f^* + \frac{1}{2\alpha k} \text{dist}^2(y(0), X^*) + 2LC + \frac{1}{2} \alpha n L^2. \quad (3.34)$$

□

This proposition tells us that the value of the function f at the average of the limit of each $x^i(k)$, is less than the minimal value of f , plus some terms depending on various constants.

We notice that in inequality (3.34), the right hand side consists of three terms. The first term is inversely proportional to k . This implies that when k increases to infinity, this term will decrease to zero and therefore have no influence on the size of the bound. The last two terms are both proportional to the stepsize α , which can be chosen. If we take a small value for α , then the error will become smaller. So one can guarantee a smaller error, by choosing the value for α accordingly.

3.4 Example

In this section we will give an example, for which we will show that the agents reach consensus and that this value will indeed solve the optimization problem. We consider five agents, for which the communication is given by the graph G in figure 2.1. As seen in section 2.2 the corresponding matrix M is given by:

$$M = \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Using lemma 3.1, we can state that $\lim_{k \rightarrow \infty} M^k = \mathbb{1} m^\top$, where m^\top is a left eigenvector corresponding to the eigenvalue $\lambda = 1$, such that $m^\top \mathbb{1} = 1$. It can be verified that $\lim_{k \rightarrow \infty} M^k = \mathbb{1} \begin{pmatrix} \frac{10}{37} & \frac{4}{37} & \frac{3}{37} & \frac{8}{37} & \frac{12}{37} \end{pmatrix}$. For each of the agent i we define a convex function $f_i(x) : \mathbb{R} \rightarrow \mathbb{R}$ that has to be optimized. Hence in this example we take $p = 1$.

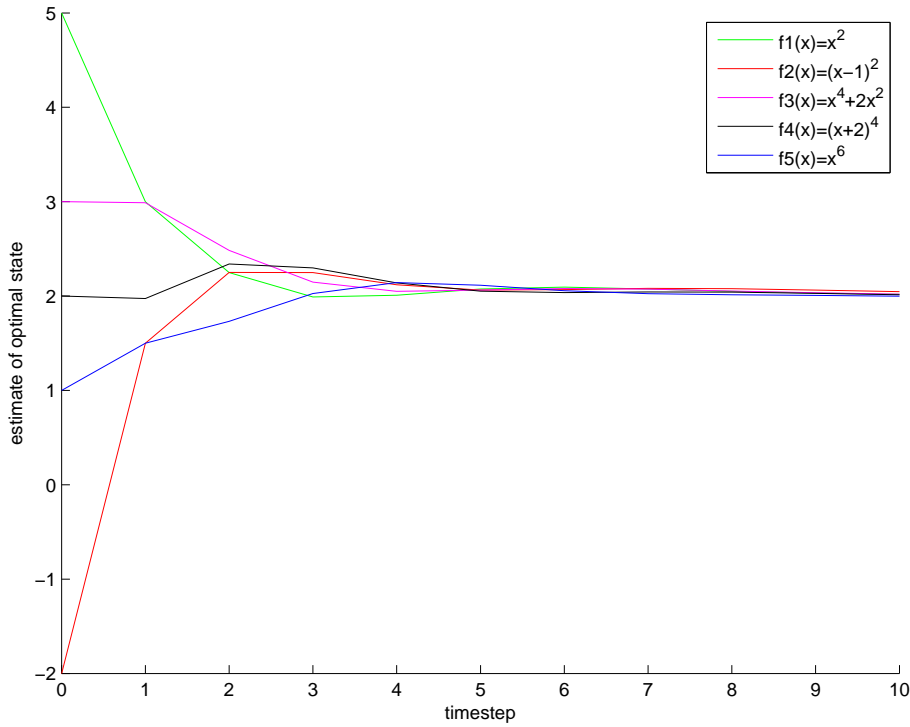


Figure 3.1: Agents agreeing on consensus state

The five functions we take are:

$$\begin{aligned}
 f_1(x) &= x^2 \\
 f_2(x) &= (x - 1)^2 \\
 f_3(x) &= x^4 + 2x^2 \\
 f_4(x) &= (x + 2)^4 \\
 f_5(x) &= x^6
 \end{aligned}$$

Since the functions f_i are continuous and differentiable, the unique subgradient is defined by the derivative of the function.

By using Matlab we generated the plots in figure 3.1 and 3.2. The code used to obtain these plots, can be found in appendix A. We chose the stepsize $\alpha = 0.0001$.

What we see in figure 3.1 is that the agents agree on a limit vector very fast. This makes sense since α is taken quite small, which results in only small influence of the subgradient compared to the averaging with the neighboring states. At first sight it may seem that the state will converge to a value close to 2, but this is not the case. If we set k to 10000 we see that the agents converge to a value

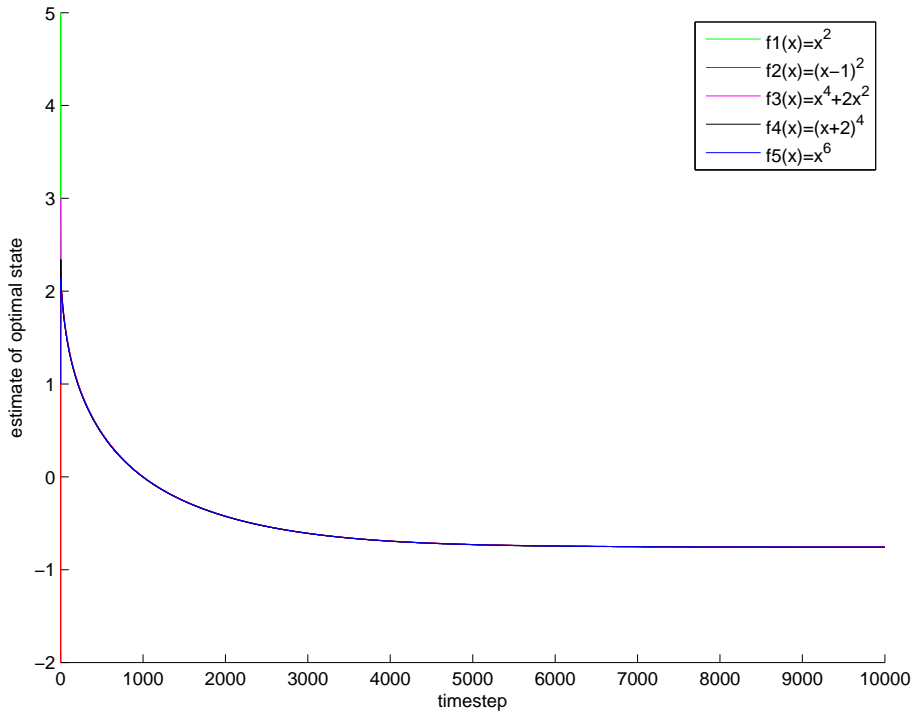


Figure 3.2: Consensus state converging to optimal solution

between -1 and -0.5 . For $k = 10000$ the vector computed is:

$$x = \begin{pmatrix} -0.7566 \\ -0.7559 \\ -0.7560 \\ -0.7572 \\ -0.7569 \end{pmatrix}. \quad (3.35)$$

This convergence to the optimal value is as slow as it is, because of the small influence of the subgradient. If the value for α is taken bigger, then it will take longer for the agents to reach consensus, but the convergence to the optimal value will be faster.

To check whether the agents indeed converge to the optimal value, we plotted the function that needs to be optimized. By using equation (3.11), we find that this function is:

$$f(x) = \frac{10}{37}x^2 + \frac{4}{37}(x-1)^2 + \frac{3}{37}(x^4+2x^2) + \frac{8}{37}(x+2)^4 + \frac{12}{37}x^6. \quad (3.36)$$

The plot of this equation can be found in figure 3.3. We can indeed confirm that the optimal value is close to -0.75 .

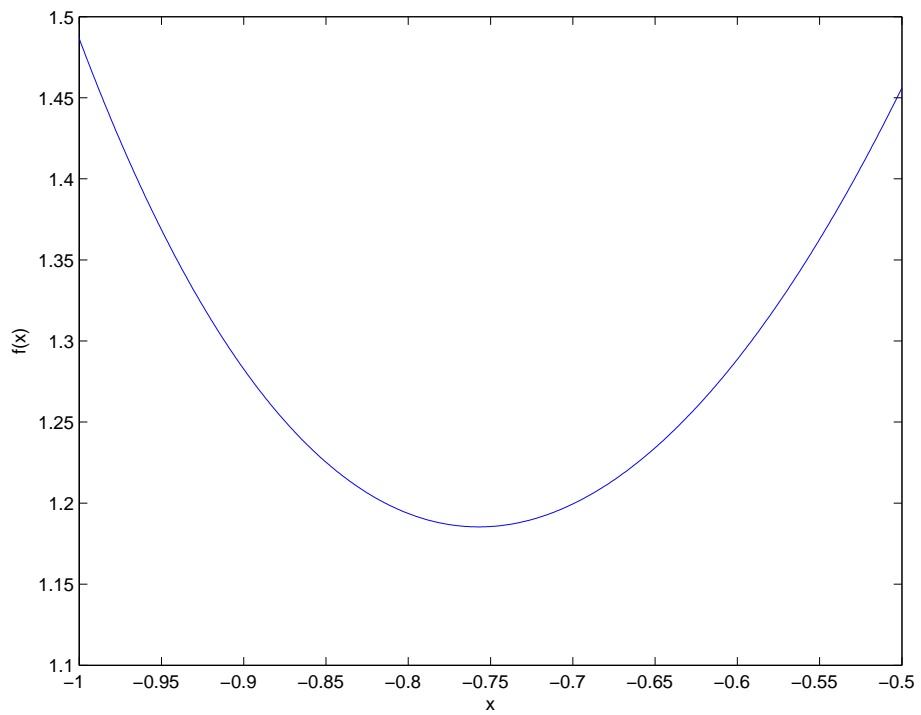


Figure 3.3: The function that is being optimized

Chapter 4

Conclusion

In this thesis we investigated convex optimization in multi-agent systems. The multi-agent system is described by a directed graph, from we obtain a matrix M . To each agent one convex function is known and our goal is to optimize a convex combination of these functions. This convex combination is determined by the system dynamics. At every time step, each agent computes a subgradient at its current estimate of the optimal value. Then it updates this estimate by combining estimates of its incoming neighbors and using the subgradient.

We showed that each row of M^k converges to a stochastic vector, in particular the left eigenvalue corresponding to the eigenvalue 1. We also showed that the rate of this convergence is geometric. This result has been used in the convergence analysis of the subgradient method.

Next we proved that at each iterate and for each agent, the distance between the current estimate and the optimal value is bounded by the distance from the estimate one step before, and some constant terms proportional to the stepsize. Finally we showed that the difference between the function value at the averaged vector of estimates and the optimal function value is bounded. Again the constant terms in the bound are proportional to the stepsize. This implies that by choosing the stepsize smaller, the bound will become stricter.

From these findings we can conclude that by making a good choice for the stepsize, we can get the function values of the estimates arbitrarily close to the optimal value. Hence we did not prove converge to the optimal value, but we can get as to close to this value as we like by choosing the stepsize accordingly.

Appendix A

Matlab Code

```
function convexoptimization
alpha=.0001;
k=10;
x=[];
x(1,1)=5;
x(2,1)=-2;
x(3,1)=3;
x(4,1)=2;
x(5,1)=1;
%f1=x^2;
%d1=2*x;
%f2=(x-1)^2;
%d2=2*(x-1);
%f3=x^4+2*x^2;
%d3=4*x^3+4*x;
%f4=(x+2)^4;
%d4=4*(x+2)^3;
%f5=x^6;
%d5=6*x^5;
M=[1/2 0 0 0 1/2; 1/2 1/2 0 0 0; 1/3 0 1/3 0 1/3; 1/4 1/4 1/4 1/4
0; 0 0 0 1/2 1/2];

for i=2:k+1
    x(1,i)=M(1,1:5)*x(1:5,i-1)-alpha*2*x(1,i-1);
    x(2,i)=M(2,1:5)*x(1:5,i-1)-alpha*2*(x(2,i-1)-1);
    x(3,i)=M(3,1:5)*x(1:5,i-1)-alpha*(4*x(3,i-1)^3+4*x(3,i-1));
    x(4,i)=M(4,1:5)*x(1:5,i-1)-alpha*(4*(x(4,i-1)+2)^3);
    x(5,i)=M(5,1:5)*x(1:5,i-1)-alpha*(6*x(5,i-1)^5);
end
t=[0:k];
hold on
plot(t,x(1,1:k+1),'g');
plot(t,x(2,1:k+1),'r');
plot(t,x(3,1:k+1),'m');
plot(t,x(4,1:k+1),'k');
plot(t,x(5,1:k+1));
xlabel('timestep');
ylabel('estimate of optimal state');
legend('f1(x)=x^2','f2(x)=(x-1)^2','f3(x)=x^4+2x^2','f4(x)=(x+2)^4',
,'f5(x)=x^6');
end
```

Bibliography

- [1] A. Berman and R. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. Academic Press, 1979.
- [2] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [3] C.D. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, 2000.
- [4] A. Nedić and A. Ozdaglar. Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control*, 2009.
- [5] J.N. Tsitsiklis, D.P. Bertsekas, and M. Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Transactions on Automatic Control*, 1986.