Distributed Optimization with State-Dependent Communication in Multi-Agent Networks

Asu Ozdaglar

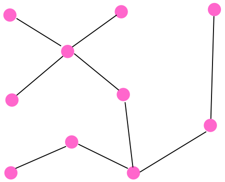
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology

ITMANET Review Meeting
January, 2011

Motivation

Recent interest in developing distributed algorithms for solving convex optimization problems in multi-agent networked systems.

Motivated by resource allocation problems and networked decision/data-processing problems among users with heterogeneous performance objectives in wireless networks.



- Key feature of such problems is the decentralized nature of information: Global objective is a combination of each agent's local objectives, feasibility is given by the intersection of agents' local constraints.

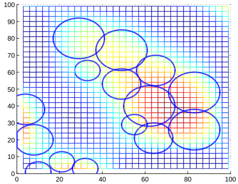
- Algorithms designed for such systems should:
  - Rely on local information and involve simple computations.
  - Robust to dynamic changes in network topology (due to link or node failures).

Literature

- Decomposition methods: Dual and primal-dual subgradient algorithms.
- Consensus-based schemes:
    - Parallel computation and optimization among different processors.
        - Tsitsiklis (84), Bertsekas and Tsitsiklis (95)
    - Consensus and cooperative control.
        - Jadbabaie *et al.* (03), Olfati-Saber and Murray (04), Boyd *et al.* (05), Olshevsky and Tsitsiklis (07), Tahbaz-Salehi, Jadbabaie (08), Fagnani, Zampieri (09)
    - Multi-agent optimization.
        - Deterministic communication models: Nedic, Ozdaglar (07), Nedic, Ozdaglar, Parrilo (08), Nedic, Olshevsky, Ozdaglar, Tsitsiklis (09),Zhu, Martinez (10)
        - Random communication models: Lobel, Ozdaglar (09), Baras and Matei (10), Agarwal, Duchi, Wainwright (10)
        - Incremental subgradient methods: Ram, Nedic, Veeravalli (09), Johansson, Rabi, Johansson (09)

This Talk

- “Consensus-based" schemes assume agents exchange information with their neighbors over an exogenous (fixed or time-varying) network topology
    - In many of the applications, network topology configured endogenously as a function of the agent states
    - Examples: Location optimization problems, rendezvous problems, sensor coverage.



- Distributed multi-agent optimization with state-dependent communication [Lobel, Ozdaglar, Feijer 10]
    - The communication network varies as the location of mobile agents changes in response to the objective they are trying to achieve.

Model

- We consider a set of nodes (or agents) $\mathcal{M} = \{1, \ldots, m\}$.
- The goal of the agents is to cooperatively solve the problem

$$\begin{aligned} \text{minimize} \quad & \sum_{i \in \mathcal{M}} f_i(x) \\ \text{subject to} \quad & x \in \bigcap_{i \in \mathcal{M}} X_i, \end{aligned}$$

  $f_i(x) : \mathbb{R}^n \to \mathbb{R}$ is a convex function and $X_i \subset \mathbb{R}^n$ is a convex set known only to agent i.
- State $x_i(k)$ represents agent i's estimate of the solution at time k.
- At each time k, agent i updates its estimate as

$$x_i(k+1) = P_{X_i}\left[\sum_{j \in \mathcal{M}} a_{ij}(k) x_j(k) - \alpha(k) d_i(k)\right],$$

  where $[a_{ij}(k)]_{j \in \mathcal{M}}$ is a weight vector, $\alpha(k) > 0$ is a stepsize, and $d_i(k)$ is a subgradient of $f_i(x)$ at $v_i(k) = \sum_{j=1}^{m} a_{ij}(k) x_j(k)$.
- We assume that the subgradients of each f_i are uniformly bounded, i.e., there exists some $L > 0$ such that for every $i \in \mathcal{M}$ and any $x \in \mathbb{R}^n$, we have

$$\|d\| \leq L \qquad \text{for all } d \in \partial f_i(x).$$

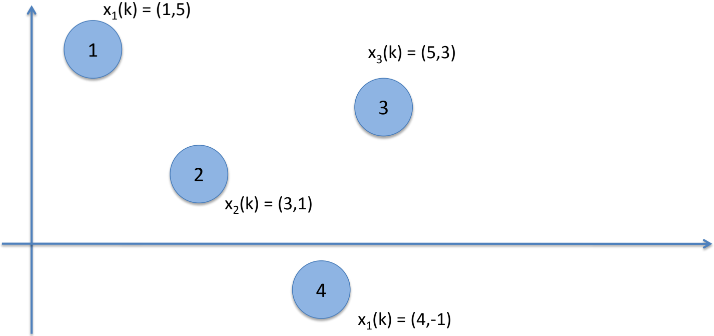
State-Dependent Communication

- Consider a location optimization problem for mobile agents.
  - Agent i's estimate $x_i(k)$ at time k is also its physical location.

- The network topology is a (random) function of the agent states.

State-Dependent Communication

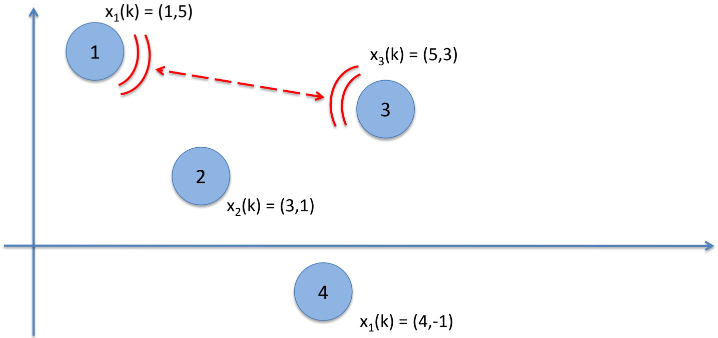
- Consider a location optimization problem for mobile agents.
  - Agent i's estimate $x_i(k)$ at time k is also its physical location.



- The network topology is a (random) function of the agent states.

State-Dependent Communication

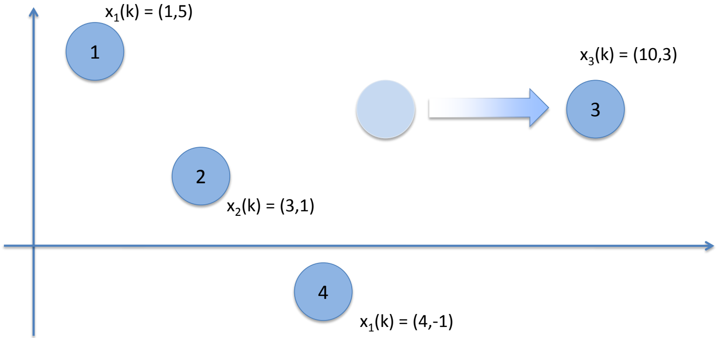
- Consider a location optimization problem for mobile agents.
  - Agent i's estimate $x_i(k)$ at time k is also its physical location.



- The network topology is a (random) function of the agent states.

State-Dependent Communication

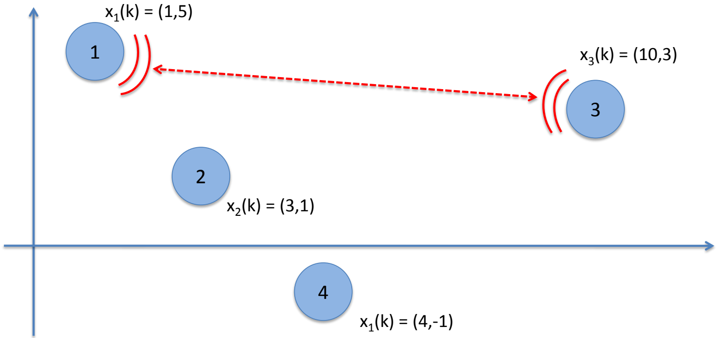
- Consider a location optimization problem for mobile agents.
  - Agent i's estimate $x_i(k)$ at time k is also its physical location.



- The network topology is a (random) function of the agent states.

State-Dependent Communication

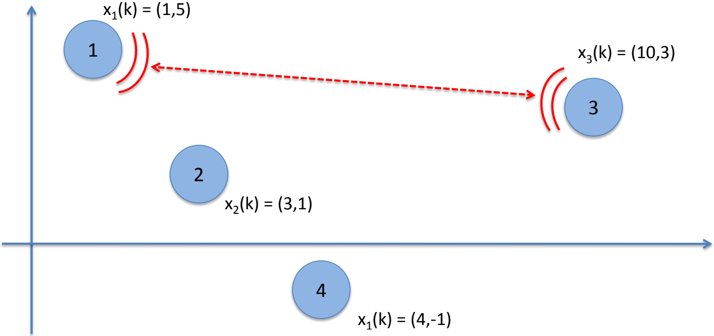
- Consider a location optimization problem for mobile agents.
  - Agent i's estimate $x_i(k)$ at time k is also its physical location.



- The network topology is a (random) function of the agent states.

State-Dependent Communication

- Consider a location optimization problem for mobile agents.
  - Agent i's estimate $x_i(k)$ at time k is also its physical location.



- The network topology is a (random) function of the agent states.

Network Communication Model

- We define the communication matrix for the network at time k as $A(k) = [a_{ij}(k)]_{i,j \in \mathcal{M}}$.
- We assume that $A(k)$ satisfies the following:
  - The communication matrix $A(k)$ is doubly stochastic for all $k \geq 0$ with probability one.
  - There exists some $\gamma > 0$ such that $a_{i,i}(k) \geq \gamma$ for all $i \in \mathcal{M}$ and all $k \geq 0$ with probability one.
- We further assume that the probability of communication between two agents is potentially small if their estimates are far apart: There exists some $C > 0$, $K > 0$ and $\delta \in (0, 1)$ such that for all (i, j), all $k \geq 0$ and all $\bar{x} \in \mathbb{R}^{m \times n}$,

$$P(a_{ij}(k) \geq \gamma | x(k) = \bar{x}) \geq \min\left\{\delta, \frac{K}{\|\bar{x}_i - \bar{x}_j\|^C}\right\}.$$

Counterexample

- Two agents solve a one dimensional minimization problem over $X_1 = X_2 = [0, \infty)$.
- Agent objective functions given by $f_1(x) = 2x$ and $f_2(x) = -x$.
- Global objective function is $f(x) = x$ and global optimum is $x^* = 0$.
- Let $C > 1$, $\alpha(k) = \alpha > 0$ and, for simplicity, assume $x_1(0) = 0$.
- For any $k \geq 0$,

$$P(\text{agents don't communicate after } k) = \prod_{j=0}^{\infty} \left(1 - \min\left\{\delta, (\alpha j)^{-C}\right\}\right) \geq \epsilon > 0.$$

- Consider all periods k immediately after the agents communicated.
- By the Borel-Cantelli Lemma, $\lim_{k\to\infty} x_2(k) = \infty$ with probability 1

Counterexample

- Two agents solve a one dimensional minimization problem over $X_1 = X_2 = [0, \infty)$.
- Agent objective functions given by $f_1(x) = 2x$ and $f_2(x) = -x$.
- Global objective function is $f(x) = x$ and global optimum is $x^* = 0$.
- Let $C > 1$, $\alpha(k) = \alpha > 0$ and, for simplicity, assume $x_1(0) = 0$.
- For any $k \geq 0$,

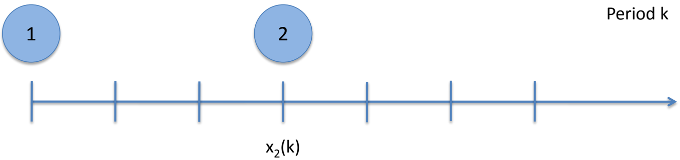
$$P(\text{agents don't communicate after } k) = \prod_{j=0}^{\infty} \left(1 - \min\left\{\delta, (\alpha j)^{-C}\right\}\right) \geq \epsilon > 0.$$

- Consider all periods k immediately after the agents communicated.
- By the Borel-Cantelli Lemma, $\lim_{k\to\infty} x_2(k) = \infty$ with probability 1

Counterexample

- Two agents solve a one dimensional minimization problem over $X_1 = X_2 = [0, \infty)$.
- Agent objective functions given by $f_1(x) = 2x$ and $f_2(x) = -x$.
- Global objective function is $f(x) = x$ and global optimum is $x^* = 0$.
- Let $C > 1$, $\alpha(k) = \alpha > 0$ and, for simplicity, assume $x_1(0) = 0$.
- For any $k \geq 0$,

$$P(\text{agents don't communicate after } k) = \prod_{j=0}^{\infty} \left(1 - \min\left\{\delta, (\alpha j)^{-C}\right\}\right) \geq \epsilon > 0.$$

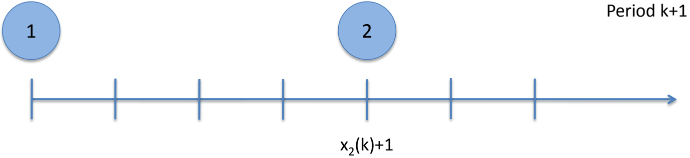


- Consider all periods k immediately after the agents communicated.
- By the Borel-Cantelli Lemma, $\lim_{k \to \infty} x_2(k) = \infty$ with probability 1

Counterexample

- Two agents solve a one dimensional minimization problem over $X_1 = X_2 = [0, \infty)$.
- Agent objective functions given by $f_1(x) = 2x$ and $f_2(x) = -x$.
- Global objective function is $f(x) = x$ and global optimum is $x^* = 0$.
- Let $C > 1$, $\alpha(k) = \alpha > 0$ and, for simplicity, assume $x_1(0) = 0$.
- For any $k \geq 0$,

$$P(\text{agents don't communicate after } k) = \prod_{j=0}^{\infty} \left(1 - \min\left\{\delta, (\alpha j)^{-C}\right\}\right) \geq \epsilon > 0.$$

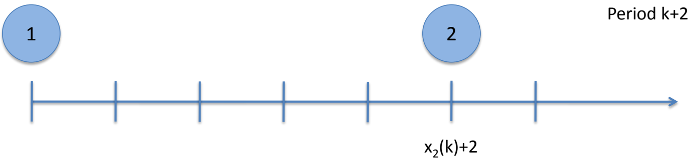


- Consider all periods k immediately after the agents communicated.
- By the Borel-Cantelli Lemma, $\lim_{k\to\infty} x_2(k) = \infty$ with probability 1

Counterexample

- Two agents solve a one dimensional minimization problem over $X_1 = X_2 = [0, \infty)$.
- Agent objective functions given by $f_1(x) = 2x$ and $f_2(x) = -x$.
- Global objective function is $f(x) = x$ and global optimum is $x^* = 0$.
- Let $C > 1$, $\alpha(k) = \alpha > 0$ and, for simplicity, assume $x_1(0) = 0$.
- For any $k \geq 0$,

$$P(\text{agents don't communicate after } k) = \prod_{j=0}^{\infty} \left(1 - \min\left\{\delta, (\alpha j)^{-C}\right\}\right) \geq \epsilon > 0.$$

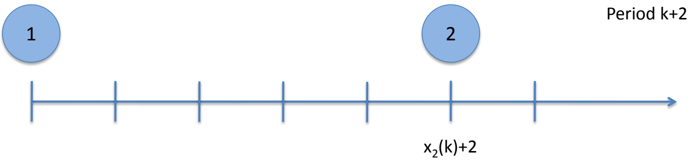


- Consider all periods k immediately after the agents communicated.
- By the Borel-Cantelli Lemma, $\lim_{k\to\infty} x_2(k) = \infty$ with probability 1

Counterexample

- Two agents solve a one dimensional minimization problem over $X_1 = X_2 = [0, \infty)$.
- Agent objective functions given by $f_1(x) = 2x$ and $f_2(x) = -x$.
- Global objective function is $f(x) = x$ and global optimum is $x^* = 0$.
- Let $C > 1$, $\alpha(k) = \alpha > 0$ and, for simplicity, assume $x_1(0) = 0$.
- For any $k \geq 0$,

$$P(\text{agents don't communicate after } k) = \prod_{j=0}^{\infty} \left(1 - \min\left\{\delta, (\alpha j)^{-C}\right\}\right) \geq \epsilon > 0.$$



- Consider all periods k immediately after the agents communicated.
- By the Borel-Cantelli Lemma, $\lim_{k\to\infty} x_2(k) = \infty$ with probability 1

Analysis of Updates

- We rewrite the estimate update as

$$
\begin{aligned}
x_i(k+1) &= v_i(k) - \alpha(k)d_i(k) + e_i(k),\\
e_i(k) &= P_{X_i}[v_i(k) - \alpha(k)d_i(k)] - \Big(v_i(k) - \alpha(k)d_i(k)\Big),
\end{aligned}
$$

where $v_i(k) = \sum_{j=1}^{m} a_{ij}(k)x_j(k)$ and $e_i(k)$ is the projection error at time k.

- For any $s \geq 0$ and any $k \geq s$, we define the transition matrices

$$\Phi(k,s) = A(s)A(s+1)\cdots A(k-1)A(k) \qquad \text{for all } s \text{ and } k \text{ with } k \geq s,$$

and relate the estimates at k and s with $k > s$,

$$
\begin{aligned}
x_i(k+1) = \sum_{j=1}^{m}[\Phi(k,s)]_{ij}x_j(s) \quad &- \quad \sum_{r=s+1}^{k}\sum_{j=1}^{m}[\Phi(k,r)]_{ij}\alpha(r-1)d_j(r-1) - \alpha(k)d_i(k)\\
&+ \quad \sum_{r=s+1}^{k}\sum_{j=1}^{m}[\Phi(k,r)]_{ij}e_j(r-1) + e_i(k).
\end{aligned}
$$

- We introduce the disagreement metric ρ,

$$\rho(k,s) = \max_{i,j\in\mathcal{M}}\left|[\Phi(k,s)]_{ij} - \frac{1}{m}\right| \qquad \text{for all } k \geq s \geq 0.$$

Analysis of Updates

Lemma

For all $k \geq 0$,

$$\max_{i,h\in\mathcal{M}} \|x_i(k) - x_h(k)\| \leq \Delta + 2mL\sum_{r=0}^{k-1}\alpha(r) + 2\sum_{r=0}^{k-1}\sum_{j=1}^{m}\|e_j(r)\|,$$

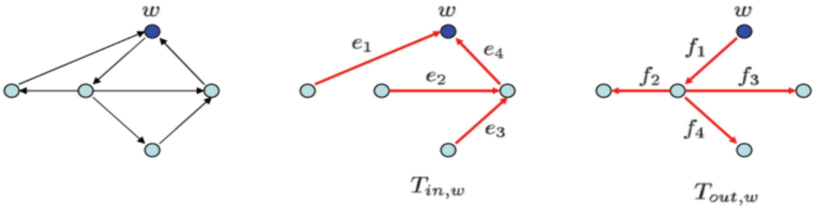
where $\Delta = 2m\max_{j\in\mathcal{M}}\|x_j(0)\|$.

- The lemma establishes a bound on the distance between the agents' estimates.
- If there exists some $M > 0$ such that $\|e_i(k)\| \leq M\alpha(k)$ for all i and $k \geq 0$, then with probability 1, $\max_{i,h\in\mathcal{M}}\|x_i(k) - x_h(k)\| \leq \Delta + 2m(L+M)\sum_{r=0}^{k-1}\alpha(r)$.
- For all k, we define the reachable set

$$R_M(k) = \left\{ x \in \mathbb{R}^{m\times n} \mid \max_{i,h\in\mathcal{M}}\|x_i - x_h\| \leq \Delta + 2m(L+M)\sum_{r=0}^{k-1}\alpha(r) \right\}.$$

Propagation of Information

- We next construct an even in which edges among the agents are activated sequentially over time, so that information propagates from every agent to every other agent.
- To define this event, we fix a node $w \in \mathcal{M}$ and consider two directed spanning trees rooted at w with a specific order.



- Let G(k) denote the event in which each edge in the spanning trees $T_{in,w}$ and $T_{out,w}$ are activated sequentially following time k in this order.
- If the event $G(k)$ occurs, then information from every agent reaches every other agent by period $k + 2(m - 1)$.
- Implies a contraction in the disagreement metric $\rho(k, s)$.

Propagation of Information

Lemma

Assume that there exists some $M > 0$ such that $\|e_i(k)\| \leq M\alpha(k)$ for all i and $k \geq 0$. For all $k \geq s$ and any $\bar{x} \in R_M(s)$,

$$P(G(k)|x(s) = \bar{x}) \geq \min\left\{\delta, \frac{1}{(\Delta + 2m(L+M)\sum_{r=1}^{k+2m-3}\alpha(r))^C}\right\}^{2(m-1)}$$

Lemma

For any $k > s > 0$, let t be a positive integer such that $s < s_1 < s_2 < \cdots < s_t < k$ and $s_{i+1} - s_i \geq 2(m-1)$ for all $i = 1, \ldots, t-1$. Suppose that events $G(s_i)$ occur for each $i = 1, \ldots, t$. Then,

$$\rho(k,s) \leq 2\left(1 + \frac{1}{\gamma^{2(m-1)}}\right)\left(1 - \gamma^{2(m-1)}\right)^t.$$

Contraction Bound

Assumption

The stepsize sequence $\{\alpha(k)\}_{k\in\mathbb{N}}$ *satisfies*

$$\lim_{k\to\infty} k\log^{p}(k)\alpha(k)=0 \qquad \textit{for all } p<1.$$

- Example: $\alpha(k)=\frac{1}{k\log(k)}$ (satisfies: $\sum_{k=0}^{\infty}\alpha(k)=\infty, \quad \sum_{k=0}^{\infty}\alpha^{2}(k)<\infty$) .

Proposition

Assume that there exists some $M>0$ *such that* $\|e_i(k)\|\leq M\alpha(k)$ *for all* $i\in\mathcal{M}$ *and* $k\in\mathbb{N}$. *Then, there exists a scalar* $\mu>0$, *an increasing function* $\beta(s):\mathbb{N}\to\mathbb{R}_+$ *and a function* $S(q):\mathbb{N}\to\mathbb{N}$ *such that*

$$\beta(s)\leq s^{q} \qquad \textit{for all } q>0 \textit{ and all } s\geq S(q) \tag{1}$$

$$\textit{and} \qquad E[\rho(k,s)|x(s)=\bar{x}]\leq\beta(s)e^{-\mu\sqrt{k-s}} \quad \textit{for all } k\geq s\geq 0,\ \bar{x}\in R_M(s).$$

- A time-nonhomogeneous contraction bound.

Agent Disagreements

- To measure the agent disagreements $\|x_i(k) - x_j(k)\|$, we consider their average and measure disagreement with respect to this average:

$$y(k) = \frac{1}{m}\sum_{j=1}^{m} x_j(k) \qquad \text{for all } k.$$

- We have

$$y(k+1) = y(k) - \frac{\alpha(k)}{m}\sum_{i=1}^{m} d_i(k) + \frac{1}{m}\sum_{i=1}^{m} e_i(k).$$

Lemma

For all i and $k \geq 0$, we have

$$\begin{aligned}\|x_i(k) - y(k)\| \quad &\leq \quad m\rho(k-1,0)\sum_{j=1}^{m}\|x_j(0)\| + mL\sum_{r=0}^{k-2}\rho(k-1,r+1)\alpha(r) + 2\alpha(k-1)L \\ &\quad + \sum_{r=0}^{k-2}\rho(k-1,r+1)\sum_{j=1}^{m}\|e_j(r)\| + \|e_i(k-1)\| + \frac{1}{m}\sum_{j=1}^{m}\|e_j(k-1)\|.\end{aligned}$$

Analysis of the Distributed Subgradient Method

Assumption: Constraint set $X_i = X$ for all i.

Lemma

For all i and k,

$$\|e_i(k)\| \leq 2L\alpha(k)$$

- Relies on nonexpansiveness of projection.
- The state $x(k)$ is in the reachable set $R_{2L}(k)$ with probability one.
- Simplifies the preceding bound:

$$\begin{aligned} \|x_i(k) - y(k)\| \quad \leq \quad & m\rho(k-1,0)\sum_{j\in\mathcal{N}}\|x_j(0)\| + \\ & 3mL\sum_{r=0}^{k-2}\rho(k-1,r+1)\alpha(r) + 6\alpha(k-1)L \end{aligned}$$

Analysis of the Distributed Subgradient Method

Assumption: Constraint set $X_i = X$ for all i.

Lemma

For all i and k,

$$\|e_i(k)\| \leq 2L\alpha(k)$$

- Relies on nonexpansiveness of projection.
- The state $x(k)$ is in the reachable set $R_{2L}(k)$ with probability one.
- Simplifies the preceding bound:

$$\begin{aligned}\|x_i(k) - y(k)\| \quad &\leq \quad m\rho(k-1,0)\sum_{j\in\mathcal{N}}\|x_j(0)\| + \\ &\qquad 3mL\sum_{r=0}^{k-2}\rho(k-1,r+1)\alpha(r) + 6\alpha(k-1)L\end{aligned}$$

Taking Expectations

- Using the contraction bound, we have

$$E\left[\sum_{r=0}^{k-2} \rho(k-1, r+1)\alpha(r)\right] \leq \sum_{r=0}^{k-2} \beta(r+1)e^{\mu\sqrt{k-r-2}}\alpha(r)$$

- Using the Stepsize Assumption,

$$\lim_{k\to\infty} \sum_{r=0}^{k-2} \beta(r+1)e^{\mu\sqrt{k-r-2}}\alpha(r) = 0$$

Proposition

For all i, we have

$$\lim_{k\to\infty} E[\|x_i(k) - y(k)\|] = 0, \quad \textit{and}$$

$$\liminf_{k\to\infty} \|x_i(k) - y(k)\| = 0 \qquad \textit{with probability one.}$$

Convergence

Using supermartingale convergence results, we obtain the following:

Proposition

For all i, we have:

(a) $\sum_{k=2}^{\infty} \alpha(k)\|x_i(k) - y(k)\| < \infty$ *with probability one.*

(b) $\lim_{k\to\infty} \|x_i(k) - y(k)\| = 0$ *with probability one.*

Theorem

There exists an optimal solution $x^* \in X^*$ *such that for all i*

$$\lim_{k\to\infty} x_i(k) = x^* \qquad \textit{with probability one}.$$

When the Constraint Sets X_i are Different

Assumption: The constraint sets X_i are compact.

Lemma

Assume that the stepsize sequence satisfies $\alpha(k) \to 0$ as k goes to infinity. The projection errors $e_i(k)$ converge to zero as $k \to \infty$, i.e.,

$$\lim_{k\to\infty} \|e_i(k)\| = 0 \qquad \textit{for all } i.$$

Theorem

Assume that the stepsize sequence satisfies $\sum_k \alpha(k) = \infty$ and $\sum_k \alpha^2(k) < \infty$. Then, there exists an optimal solution $x^* \in X^*$ such that for all i

$$\lim_{k\to\infty} x_i(k) = x^* \qquad \textit{with probability one}.$$

Conclusions

- We presented distributed algorithms for solving convex optimization problems over networks.
- We studied a projected multi-agent subgradient algorithm under state-dependent communication among networked agents.
    - We showed the iterates with constant stepsize may diverge.
    - We proposed a method based on fast diminishing stepsizes and showed convergence using a time non-homogeneous contraction bound.
- Future Directions:
    - Network effects.
    - Second-order methods for general multi-agent optimization problems.
    - Primal-dual methods for global constraints.