Asynchronism and convergence rates in distributed optimization

Hamid Reza Feyzmahdavian, Arda Aytekin and Mikael Johansson
KTH - Royal Institute of Technology

Motivation

Optimization as iterative algorithms

Many optimization algorithms are iterations, e.g.

\[ x(t+1) = x(t) - \gamma \nabla f(x(t)) := Mx(t) \]

Optimizer \( x^\star \) is a fixed-point of \( M \).

Easy to analyze when \( M \) is a contraction mapping

\[ \|Mx - My\| \leq c \|x - y\| \quad \forall x, y \in \mathbb{R}^n \]

for some \( c \in [0, 1) \) and some norm \( \| \cdot \| \). Then \( \|x(t) - x^\star\| \leq c^t \|x(0) - x^\star\| \)

Ex. Gradient mapping when \( f \) is \( \mu \)-strongly convex with \( L \)-Lipschitz gradient

Distributed implementations and asynchrony

Emerging applications require distributed implementations

Communication delays, lack of synchronization ⇒ asynchronous iterations

The impact of asynchrony

Asynchrony can cause otherwise stable iterations to diverge, or slow down.

\[
\begin{align*}
x_1(t+1) &= x_1(t) - 0.75x_1(t) - 0.7x_2(t - \tau(t)) \\
x_2(t+1) &= x_2(t) - 0.75x_2(t) - 0.7x_1(t - \tau(t))
\end{align*}
\]

Need models and tools for asynchronous iterations!
A model for asynchronous iterations

A standard form for asynchronous iterations:

\[ x_i(t + 1) = \begin{cases} 
  \mathcal{M}_i(x_1(\tau^i_1(t)), \ldots, x_n(\tau^i_n(t))) & \text{if } t \in \mathcal{T}^i \\
  x_i(t) & \text{otherwise}
\end{cases} \]

Here, \( \mathcal{T}^i \) is the set of times when node \( i \) executes an update, and \( \tau^j_i(t) \) is the time when the most recent version of \( x_j \) available to node \( i \) at time \( t \) was computed.

Note: Can view \( t - \tau^j_i(t) \) as information delay from node \( j \) to \( i \) at time \( t \).

Chazan and Miranker (1969), Baudet (1978), Bertsekas and Tsitsiklis (1989), …

Partially asynchronous algorithms

The iteration

\[ x_i(t + 1) = \begin{cases} 
  \mathcal{M}_i(x_1(\tau^i_1(t)), \ldots, x_n(\tau^i_n(t))) & \text{if } t \in \mathcal{T}^i \\
  x_i(t) & \text{otherwise}
\end{cases} \]

is called partially asynchronous if there exists \( B > 0 \) such that

a) For every \( i, t \), at least one element of \( \{t, t + 1, \ldots, t + B - 1\} \) is in \( \mathcal{T}^i \).

b) For every \( i, j \) and all \( t \in \mathcal{T}^i \), we have \( 0 \leq t - \tau^j_i(t) \leq B - 1 \).

c) There holds \( \tau^i_i(t) = t \) for all \( i \) and all \( t \in \mathcal{T}^i \).

Bounded update intervals/information delays, direct access to “own” state

Totally asynchronous algorithms

The iteration

\[ x_i(t + 1) = \begin{cases} 
  \mathcal{M}_i(x_1(\tau^i_1(t)), \ldots, x_n(\tau^i_n(t))) & \text{if } t \in \mathcal{T}^i \\
  x_i(t) & \text{otherwise}
\end{cases} \]

is called totally asynchronous if

a) every set \( \mathcal{T}^i \) is an infinite subset of \( \mathbb{N}_0 \)

b) for every sequence \( \{t_k\} \) of elements of \( \mathcal{T}^i \) that tends to infinity, it holds that \( \lim_{k \to \infty} \tau^j_i(t_k) = \infty \) for all \( i, j \).

No node ceases to update, old information eventually purged out of system.

Challenge: quantify the impact of asynchronism

We address two key questions:

1. quantify how \( B \) impacts convergence of partially asynchronous iterations

2. establish convergence rates for classes of totally asynchronous iterations

We then use this insight to design delay-insensitive optimization algorithms.
Outline

1. Motivation
2. Problem formulation
3. Convergence rates of asynchronous iterations
4. Example: power control in wireless systems
5. A delayed incremental gradient method with linear convergence rate
6. Conclusions

Outline

1. Motivation
2. Problem formulation
3. Convergence rates of asynchronous iterations
4. Example: power control in wireless systems
5. A delayed incremental gradient method with linear convergence rate
6. Conclusions

Problem formulation

Consider iterations

\[ x(t + 1) = Mx(t) \]

where \( M \) is a pseudo-contraction

\[ \|Mx - x^*\| \leq c\|x - x^*\| \quad \forall x \in \mathbb{R}^n \]

with respect to a block-maximum norm

\[ \|x\|_b = \max_{1 \leq i \leq m} w_i \|x_i\| \]

(Here \( x = (x_1, \ldots, x_m) \in \mathbb{R}^n \), \( x_i \in \mathbb{R}^{n_i} \), and \( \| \cdot \|_i \) is any norm)

Challenge: Quantify the impact of asynchrony on the iterates.

Main result

Theorem 1. If

a) \( M \) is pseudo-contraction with modulus \( c \) w.r.t. block-maximum norm
b) There exist functions \( \beta^j : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \) and \( \Delta \in \mathbb{N}_0 \) such that, \( \forall t \geq \Delta \)

\[ t - t_k \leq \beta^j(t) \leq t \quad t \in (t_k, t_{k+1}] \]

for every two consecutive elements \( t_k \) and \( t_{k+1} \) in \( \mathcal{T}^i \).

c) There is a decreasing function \( \lambda : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \) with \( \lim_{t \rightarrow \infty} \lambda(t) = 0 \) and

\[ c \lim_{t \rightarrow \infty} \frac{\lambda(t_j) - \beta^j(t_j)}{\lambda(t)} < 1 \quad \forall i, j \]

Then, the sequence generated by (2) under total asynchronism satisfies

\[ \frac{1}{w_i} \|x_i(t) - x_i^*\| \leq M\lambda(t_i), \quad t \in (t_k, t_{k+1}] \]

for all \( i \) and all \( t \), where \( M \) is a positive constant.

Main result

Theorem 1. If

a) \( M \) is pseudo-contraction with modulus \( c \) w.r.t. block-maximum norm
b) There exist functions \( \beta^j : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \) and \( \Delta \in \mathbb{N}_0 \) such that, \( \forall t \geq \Delta \)

\[ t - t_k \leq \beta^j(t) \leq t \quad t \in (t_k, t_{k+1}] \]

for every two consecutive elements \( t_k \) and \( t_{k+1} \) in \( \mathcal{T}^i \).

c) There is a decreasing function \( \lambda : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \) with \( \lim_{t \rightarrow \infty} \lambda(t) = 0 \) and

\[ c \lim_{t \rightarrow \infty} \frac{\lambda(t_j) - \beta^j(t_j)}{\lambda(t)} < 1 \quad \forall i, j \]

Then, the sequence generated by (2) under total asynchronism satisfies

\[ \frac{1}{w_i} \|x_i(t) - x_i^*\| \leq M\lambda(t_i), \quad t \in (t_k, t_{k+1}] \]

for all \( i \) and all \( t \), where \( M \) is a positive constant.

Our approach

Use a continuous decreasing function \( \lambda : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \) satisfying

\[ \lim_{t \rightarrow \infty} \lambda(t) = 0 \]

and show that there is \( M > 0 \) such that

\[ \frac{1}{w_i} \|x_i(t) - x_i^*\| \leq M\lambda(t_k), \quad \forall t \in (t_k, t_{k+1}] \]

for all \( i \), all \( t \) and every pair of consecutive elements \( t_k \) and \( t_{k+1} \) in \( \mathcal{T}^i \).
Main results

**Main result (partially asynchronous iterations)**

**Theorem 2.** Let $\mathcal{M}$ be a pseudo-contraction in the block-maximum norm. Then, the iterates generated by (2) under partial asynchronism satisfy
\[
\frac{1}{w_i} \| x_i(t) - x_i^* \| \leq M \rho^{t_k} \quad t \in (t_k, t_{k+1})
\]
for every pair of consecutive elements $t_k$ and $t_{k+1}$ in $T^i$. Moreover,
\[
\rho = \frac{c}{2^{t-1}}
\]

**Note.** Convergence rate still linear. Slows down with increasing $B$.

Proof uses Theorem 1 with $\beta(t) = B$ and $\lambda(t) = \rho^t$.

**Example (“retarding divider”)**

Consider the iteration
\[
x(t + 1) = \begin{cases} 
\frac{1}{2} x(t), & t \in T \\
x(t), & t \not\in T 
\end{cases}
\]
where $x(t) \in \mathbb{R}$ and $T = \{2^k \mid k \in \mathbb{N}_0\}$.
Since $t_{k+1} - t_k = 2^k$, there is no uniform upper bound on inter-update times.

However, since
\[
t - t_k \leq \frac{t}{2} \leq t \quad \forall t \in (t_k, t_{k+1})
\]
$\beta(t) = t/2$ and $\lambda(t) = 1/t$ satisfy conditions of Theorem 1. It follows that
\[
|x(t)| \leq \frac{M}{t_k}, \quad t \in (t_k, t_{k+1})
\]

Thus,
\[
\frac{1}{w_i} \| x_i(t) - x_i^* \| \leq M' \rho^t,
\]
so error decays as $O(\rho^t)$

**Main result (linearly bounded delays)**

**Theorem 3.** If
a) $\mathcal{M}$ is a pseudo-contraction with modulus $c$ w.r.t. a block-maximum norm
b) For each $t \in T^i$, there exists $t' \in T^i$ such that $1 \leq t' - t \leq B$.
c) It holds that $0 \leq t - \tau_i(t) \leq \alpha t$ for all $i, j$ and all $t \geq t_\alpha$.

Then, the sequence generated by (2) under total asynchronism satisfies
\[
\frac{1}{w_i} \| x_i(t) - x_i^* \| \leq M \left( \frac{t}{B} + 1 \right)^{-\zeta} \quad t \in (t_k, t_{k+1})
\]
where $\zeta = \ln c / \ln(1 - \alpha)$.

**Note.** Bounded by polynomial function of time. Slower as delays increase.

**Discussion: iterate time vs. physical time**

Upper bound decreases only at iteration times, stays constant in between.

In physical time, convergence rate depends on how update times grow large.

For partially asynchronous iterations $t - B \leq t_k$ for $t \in (t_k, t_{k+1})$, so
\[
M \rho^{t_k} \leq M' \rho^{t-B} := M' \rho^t, \quad t \in (t_k, t_{k+1})
\]
Thus,
\[
\frac{1}{w_i} \| x_i(t) - x_i^* \| \leq M' \rho^t,
\]
so error decays as $O(\rho^t)$
Applications

Application: wireless power control

User $i$ transmits at power $p_i$, tries to maintain SINR target $\gamma_i$

$$\text{SINR}_i = \frac{g_{ii}p_i}{\sum_{j \neq i} g_{ij}p_j + v_i} \geq \gamma_i$$

Transmit powers that minimize total energy satisfy

$$\frac{g_{ii}p_i}{\sum_{j \neq i} g_{ij}p_j + v_i} = \gamma_i$$

or, equivalently

$$p_i = I_i(p)$$

where $I_i : \mathbb{R}_+^n \rightarrow \mathbb{R}_+$ is the interference function.

Corollary. Consider the asynchronous power control iteration, and assume

a) every mobile updates its power at least once every $B$ time units, and
b) no information is more than $D_{\text{max}}$ time units old.

If $I(p)$ is a $c$-contractive interference function, then

$$\frac{1}{\epsilon_i} |p_i(t) - p_i^*| \leq M \rho^k, \quad t \in (t_k^i, t_{k+1}^i)$$

where $M > 0$ and $t_k^i$ and $t_{k+1}^i$ are consecutive elements of $T_i$. Moreover,

$$\rho = c^{B+D_{\text{max}}}$$

Application: wireless power control

Transmit power control implements fixed-point iteration

$$p_i(t+1) = I_i(p(t))$$

Definition 1. $I : \mathbb{R}_+^n \rightarrow \mathbb{R}_+$ is a $c$-contractive interference function if

a) $I_i(p) \geq 0$

b) If $p \geq p'$ then $I_i(p) \geq I_i(p')$

c) There exists $c \in [0, 1)$ and a vector $\nu > 0$ such that for all $\epsilon > 0$

$$I_i(p + \epsilon \nu) \leq I_i(p) + c \epsilon \nu$$

Proposition. If $I : \mathbb{R}_+^n \rightarrow \mathbb{R}_+$ is a $c$-contractive interference function, then

it has a unique fixed-point $p^* \in \mathbb{R}_+^n$ and

$$\|I(p) - I(p')\|_\infty \leq c \|p - p'\|_\infty$$

Application: wireless power control

Simulations and bounds for two users in a four-user scenario

Linear interference functions, $B = D_{\text{max}} = 4$.

Bounds valid, but not tight (for these users)
Application: wireless power control

Assume that information delay for user 1 grows increasingly large

\[ t - \tau^1_j(t) = t - \tau^1(t) = [0.1t] \]

while other delays, execution times remain unchanged.

Simulations and bounds from Theorem 3.

Proof sketch

**Theorem 1 (recollection and interpretation)** If

a) \( \mathcal{M} \) is pseudo-contraction with modulus \( c \) w.r.t. block-maximum norm

b) There exist functions \( \beta^i : \mathbb{R}_+ \to \mathbb{R}_+ \) and \( \Delta \in \mathbb{N}_0 \) such that, \( \forall t \geq \Delta \)

\[ t - t^*_k \leq \beta^i(t) \leq t \quad t \in (t^*_k, t^*_{k+1}] \]

for every two consecutive elements \( t^*_k \) and \( t^*_{k+1} \) in \( T^i \).

c) There is a decreasing function \( \lambda : \mathbb{R}_+ \to \mathbb{R}_+ \) with \( \lim_{t \to \infty} \lambda(t) = 0 \) and

\[ c \lim_{t \to \infty} \frac{\lambda(t^*_j(t) - \beta^j(t^*_j(t)))}{\lambda(t)} < 1 \quad \forall i, j \]

Then, the sequence generated by (2) under total asynchronism satisfies

\[ \frac{1}{w_i} \| x_i(t) - x^*_i \|_i \leq M \lambda(t^*_k), \quad t \in (t^*_k, t^*_{k+1}] \]

for all \( i \) and all \( t \), where \( M \) is a positive constant.

Proof sketch

Step 1. Find initial time \( \hat{T} \) such that hypotheses satisfied for \( t = 0, \ldots, \hat{T} \):

Let \( t^*_0 \) be smallest element of \( T^i \). By total asynchronism, there is \( \hat{T} \) such that

\[ \tau^j_i(t) = \max \{ \Delta, \max_{1 \leq m \leq k} t^*_m + 1 \} \quad \forall t \geq \hat{T} \]

By condition c), we can find \( \hat{T} \) such that

\[ c \lambda \left( \tau^j_i(t) - \beta^j(\tau^j_i(t)) \right) \leq \lambda(t) \quad \forall t \geq \hat{T} \]

Let \( \hat{T} = \max \{ \hat{T}, \hat{T} \} \) and define \( M = \| x(0) - x^* \|_{w} / \lambda(\hat{T}) \).

Since \( \{ x \mid \| x(t) - x^* \|_{w} \leq \| x(0) - x^* \|_{w} \} \) is invariant and \( \lambda(t) \) decreasing

\[ \frac{1}{w_i} \| x_i(t) - x^*_i \|_i \leq M \lambda(t^*_k), \quad t \in (t^*_k, t^*_{k+1}] \]

for all \( t = 0, \ldots, \hat{T} \).

Step 2. Induction: assume true until \( t' \), show that it holds for \( t' + 1 \).

First consider \( t' \in T^i \), and define \( k' : t' \in (t^*_{k'}, t^*_{k'+1}] \). Then, by a)

\[ \frac{1}{w_i} \| x_i(t' + 1) - x^*_i \|_i \leq c \max_{1 \leq j \leq m} \left\{ \frac{1}{w_j} \| x_j(\tau^j(t')) - x^*_j \|_j \right\} \]

Noting that \( \tau^j(t') \leq t' \), we apply the induction hypothesis and find

\[ \frac{1}{w_j} \| x_j(\tau^j(t')) - x^*_j \|_j \leq M \lambda(t^*_{k'}) \leq M \lambda(t^*_{j'}) - \beta^j(\tau^j(t'))) \leq \frac{M}{c} \lambda(t') \]

It thus holds

\[ \frac{1}{w_i} \| x_i(t' + 1) - x^*_i \|_i \leq M \lambda(t') = M \lambda(t^*_{k'+1}) \]

Since \( t' + 1 \in (t^*_{k'}, t^*_{k'+2}] \), the assertion holds for \( t' + 1 \). (\( t' \not\in T^i \) trivial)
Outline

1. Motivation
2. Problem formulation
3. Convergence rates of asynchronous iterations
4. Example: power control in wireless systems
5. A delayed incremental gradient method with linear convergence rate
6. Conclusions

So far...

Established rather general convergence estimates for asynchronous iterations.
Psuedo-contraction in block-maximum norm essential to analysis.

When the gradient iteration

\[ x(t+1) = x(t) - \gamma \nabla f(x(t)) \]

is a contraction mapping, this is typically w.r.t. the Euclidean norm.

Can we use our insight to design delay-insensitive optimization algorithms?

A delayed incremental gradient method

Delayed incremental gradient methods

Common set-up in machine-learning applications:

\[
\text{minimize } \frac{1}{M} \sum_{m=1}^{M} f_m(x) 
\]

Centralized coordinator, workers that compute delayed (partial) gradients

State-of-the-art


\[
i(t) = \mathcal{U}[1,M] \\
x(t+1) = x(t) - \gamma \nabla f_i(t)(x(t - \tau(t)))
\]

Converges linearly to ball around origin.

Limitations:
- Analysis assumes strong convexity and bounded gradients (!)
- Convergence proof valid for one particular value of \( \gamma \).
- Step-size depends on \( M \), max-delay and gradient norms at optimum

Note. Iterations mixing delayed and current states often hard to analyze.
A delayed incremental gradient method

Delayed gradient iterations

Instead of updating based on delayed gradient
\[ x(t + 1) = x(t) - \gamma \nabla f(x(t - \tau(t))) \]
we consider updating based on delayed gradient mapping,
\[ x(t + 1) = x(t - \tau(t)) - \gamma \nabla f(x(t - \tau(t))) \]  \hspace{1cm} (1)

Proposition 1. Let \( f \) be \( \mu \)-strongly convex and have \( L \)-Lipschitz continuous gradient. If \( 0 \leq \tau(t) \leq \tau_{\text{max}} \) for all \( t \), then \( \{x(t)\} \) generated by (1) satisfies
\[ \|x(t) - x^*\| \leq \left( \frac{\kappa - 1}{\kappa + 1} \right) \frac{t}{\tau_{\text{max}} + 1} \]
where \( \kappa = L/\mu \).

M. Johansson (KTH)  Embrp't14 - Lucca, Italy - September 8-9, 2014  29 / 40

A delayed incremental gradient method

Our algorithm

To minimize
\[ f(x) = \frac{1}{M} \sum_{m=1}^{M} f_m(x) \]
we propose the following algorithm
\[ i(t) = U[1, M] \]
\[ s(t) = x(t - \tau(t)) - \gamma \nabla f_{i(t)}(x(t - \tau(t))) \]
\[ x(t + 1) = (1 - \theta)x(t) + \theta s(t) \]

M. Johansson (KTH)  Embrp't14 - Lucca, Italy - September 8-9, 2014  30 / 40

A delayed incremental gradient method

Delayed gradient iterations: quadratic objective functions

Consider minimization of the quadratic function
\[ f(x) = \frac{1}{2} (Lx_1^2 + \mu x_2^2) \]
with \( \tau(t) = 1 \) for all \( t \).

Then, delayed gradient iteration has convergence factor
\[ c_g = \frac{\kappa}{\kappa + 1} \]
while the delayed prox iteration has convergence factor
\[ c_p = \sqrt{\kappa^2 - 1} \frac{1}{\kappa + 1} < c_g \]

Potentially faster and easier to analyze

M. Johansson (KTH)  Embrp't14 - Lucca, Italy - September 8-9, 2014  31 / 40

Main result

Theorem 4. Assume that
a) each \( f_m \) is convex and has \( L_m \)-Lipschitz gradient on \( \mathbb{R}^n \)
b) the overall objective \( f \) is \( \mu \)-strongly convex
Then, if \( \gamma \in (0, \mu/\max_m L_m^2) \) the iterates generated by our method satisfy
\[ E_t[f(x(t))] - f^* \leq c^t(f(x(0)) - f^*) + e \]
with
\[ c = \left( 1 - 2\gamma \mu \theta \left( 1 - \gamma \frac{\max_m L_m^2}{\mu} \right) \right)^{1/(\tau_{\text{max}} + 1)} \]
and
\[ e = \frac{\gamma \max_m L_m}{2M(\mu - \gamma \max_m L_m^2)} \sum_{m=1}^{M} \|\nabla f_m(x^*)\| \]

Note. Linear convergence to ball around optimum. Error/speed trade-off.

M. Johansson (KTH)  Embrp't14 - Lucca, Italy - September 8-9, 2014  32 / 40
A delayed incremental gradient method

Numerical results

Representative convergence behaviour

Comparison with Hogwild!

Our algorithm converges faster with theoretically justified stepsizes.

Proof sketch

**Lemma 5.** Let \( \{V(t)\} \) be a sequence of real numbers satisfying

\[
V(t + 1) \leq pV(t) + q \max_{t - \tau(t) \leq s \leq t} V(s) + r
\]

for some non-negative numbers \( p, q \) and \( r \). If \( p + q < 1 \), and

\[
0 \leq \tau(t) \leq \tau_{\text{max}}
\]

Then,

\[
V(t) \leq c^t V(0) + e
\]

where \( c = (p + q)^{1/(1+\tau_{\text{max}})} \) and \( e = r/(1 - p - q) \).

Proof of Lemma 5. First note that since \( p + q < 1 \),

\[
1 \leq (p + q)^{-\tau_{\text{max}}/(1+\tau_{\text{max}})}
\]

so, since \( c = (p + q)^{1/(1+\tau_{\text{max}})} \),

\[
p + qc^{-\tau_{\text{max}}} = p + q(p + q)^{-\tau_{\text{max}}} \leq (p + q)(p + q)^{-\tau_{\text{max}}} = c
\]

Assertion holds for \( t = 0 \). Assume that it holds for \( t = 0, \ldots, \bar{t} \). Then

\[
V(\bar{t}) \leq c^\bar{t} V(0) + e, \quad V(s) \leq c^s V(0) + e \quad s = \bar{t} - \tau_{\text{max}}, \ldots, \bar{t}
\]

We then have

\[
V(\bar{t} + 1) \leq p c^\bar{t} V(0) + pe + q \max_{\bar{t} - \tau(\bar{t}) \leq s \leq \bar{t}} c^s V(0) + qe + r
\]

\[
\leq pc^\bar{t} V(0) + pe + qc^{\bar{t} - \tau_{\text{max}}} V(0) + qe + r = c^{\bar{t}+1} V(0) + e.
\]
A delayed incremental gradient method

Proof sketch

Proof of Theorem 4. Consider

\[ V(t + 1) = \mathbb{E}_t [f(x(t + 1))] - f^* = \mathbb{E}_{t-1} \left[ \mathbb{E}_{t-1} [f(x(t + 1))] \right] - f^* \]

Since \( f \) is convex and \( \theta \in [0, 1] \),

\[ f(x(t + 1)) - f^* = f((1 - \theta)x(t) + \theta s(t)) - f^* \leq (1 - \theta)(f(x(t)) - f^*) + \theta(f(s(t)) - f^*) \]

We establish the following bound on \( f(s(t)) - f^* \):

\[ \mathbb{E}_{t-1} [f(s(t))] - f^* \leq \left( 1 - 2\mu \gamma \left( 1 - \frac{\alpha \max_m L^2_m}{\mu} \right) \right) (f(x(t - \tau(t))) - f^*) + \frac{\gamma^2 \max_m L_m}{M} \sum_{m=1}^{M} \| \nabla f_m(x^*) \|^2 \]

Allows to express \( V(t + 1) \) in terms of \( V(t), \ldots V(t - \tau_{\text{max}}) \) plus error term.

Conclusions

Conclusions

- Convergence analysis of asynchronous iterations
- A general theorem covering both totally and partially asynchronism
- Asynchronism affects rates, not only factors
- A delayed incremental gradient method
- Running averages of delayed incremental gradient mappings
- Converges faster, and under less restrictive assumptions, than alternatives
- Not everything is in “the book” - many open problems!

References

References

Complete statements and proofs can be found in
