Systematic Analysis of Distributed Optimization Algorithms over Jointly-Connected Networks

Bryan Van Scoy
Miami University

Laurent Lessard
Northeastern University
\[
\begin{aligned}
\text{minimize} & \quad \sum_{i=1}^{n} f_i(x_i) \\
\text{subject to} & \quad x_1 = x_2 = \ldots = x_n
\end{aligned}
\]

Want each agent to compute the global optimizer \( x_\star \) by communicating with local neighbors and performing local computations.
\[
    x_i(k+1) = \sum_{j=1}^{n} W_{ij}(k) x_j(k) - \alpha y_i(k)
\]

\[
    y_i(k+1) = \sum_{j=1}^{n} W_{ij}(k) y_j(k) + \nabla f_i(x_i(k+1)) - \nabla f_i(x_i(k))
\]

• analyzed in (Nedić, Olshevsky, and Shi, 2017) and (Qu, Li, 2018)
• linear convergence over time-varying networks

\[
    \|x_i(k) - x_*\| = O(\rho^k)
\]

iterations to converge \(\propto -1/\log \rho\)
DIGing

\[
x_i(k + 1) = \sum_{j=1}^{n} W_{ij}(k) x_j(k) - \alpha y_i(k)
\]

\[
y_i(k + 1) = \sum_{j=1}^{n} W_{ij}(k) y_j(k) + \nabla f_i(x_i(k + 1)) - \nabla f_i(x_i(k))
\]

Issues:

- the bound grows with the number of agents
- the bound does not apply to large stepsizes
- the analysis is specific to DIGing
Local functions
\[ \kappa \]

Communication network
\[ \sigma, B \]

Algorithm
\[
\begin{bmatrix}
A & B_u & B_v \\
C_y & D_{yu} & D_{yv} \\
C_z & D_{zu} & D_{zv} \\
F_x & F_u
\end{bmatrix}
\]

Systematic analysis tool
LMI

Worst-case performance bound
\[ \rho \]

main contribution = sharp and systematic analysis
main contribution = sharp and systematic analysis
Function assumptions

Each local objective function $f_i$ is $L$-smooth and $m$-strongly convex with respect to the global optimizer.

The condition ratio $\kappa = L/m$ characterizes how much the curvature varies.

For all $x$,

$$
\left( \nabla f_i(x) - \nabla f_i(x_*) - m(x - x_*) \right)^T \left( \nabla f_i(x) - \nabla f_i(x_*) - L(x - x_*) \right) \leq 0
$$
Local functions \( \kappa \)

Communication network \( \sigma, B \)

Algorithm

\[
\begin{bmatrix}
A & B_u & B_v \\
C_y & D_{yu} & D_{yv} \\
C_z & D_{zu} & D_{zv} \\
F_x & F_u & \\
\end{bmatrix}
\]

Systematic analysis tool LMI

Worst-case performance bound \( \rho \)
The spectral gap characterizes the connectivity of a graph.

Spectral gap

\[
\text{spectral gap} = \left\| \frac{1}{n} \mathbf{1} \mathbf{1}^T - W \right\|_2
\]
Network assumptions

**sparsity**  
\[ W_{ij}(k) = 0 \text{ if agent } i \text{ does not receive information from agent } j \text{ at time } k \]

**balanced**  
\[ W(k)1 = W(k)^T1 = 1 \]

**spectrum**  
The spectral gap of \( W(k) \) is less than or equal to one

**joint spectrum**  
for some fixed \( B \), the spectral gap of

\[
\left[ W(k) W(k+1) \cdots W(k+B-1) \right]^{1/B}
\]

is less than or equal to \( \sigma < 1 \)

The spectral gap \( \sigma \) characterizes the connectivity of the time-varying network averaged over \( B \) iterations.
Systematic analysis tool

LMI

Worst-case performance bound $\rho$

Local functions $\kappa$

Communication network $\sigma, B$

Algorithm

$$
\begin{bmatrix}
A & B_u & B_v \\
C_y & D_{yu} & D_{yv} \\
C_z & D_{zu} & D_{zv} \\
F_x & F_u
\end{bmatrix}
$$
Algorithm

\[
\begin{bmatrix}
  x_i(k+1) \\
y_i(k) \\
z_i(k)
\end{bmatrix} =
\begin{bmatrix}
  A & B_u & B_v \\
  C_y & D_{yu} & D_{yv} \\
  C_z & D_{zu} & D_{zv}
\end{bmatrix}
\begin{bmatrix}
  x_i(k) \\
u_i(k) \\
v_i(k)
\end{bmatrix}
\]

state update

\[u_i(k) = \nabla f_i(y_i(k))\]

gradient computation

\[v_i(k) = \sum_{j=1}^{n} W_{ij}(k) z_j(k)\]

communication

\[0 = \sum_{j=1}^{n} (F_x x_j(k) + F_u u_j(k))\]

invariant
<table>
<thead>
<tr>
<th>2 state variables</th>
<th>3 state variables</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SVL</strong></td>
<td><strong>EXTRA</strong></td>
</tr>
</tbody>
</table>
| \[
\begin{pmatrix}
1 & -\gamma & \beta & -\alpha & -\gamma \\
-1 & 1 & 0 & 1 & -1 \\
1 & -\delta & 0 & 0 & \delta \\
1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0
\end{pmatrix}
\] | \[
\begin{pmatrix}
1 & -\frac{1}{2} & \alpha & -\alpha & 1 \\
1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 \\
1 & -\frac{1}{2} & 0 & 0 & 0
\end{pmatrix}
\] |
| 1 communicated variable | |

<table>
<thead>
<tr>
<th><strong>Exact Diffusion (ExDIFF)</strong></th>
<th><strong>NIDS</strong></th>
</tr>
</thead>
</table>
| \[
\begin{pmatrix}
1 & 1 & -\alpha & 1 \\
\frac{1}{2} & 0 & 0 & -\alpha & \frac{1}{2} \\
1 & 0 & -\frac{1}{2} & 0 \\
1 & 0 & 0 & 0 \\
1 & -1 & 0
\end{pmatrix}
\] | \[
\begin{pmatrix}
1 & -\frac{1}{2} & \alpha & -\alpha & \frac{1}{2} & 1 \\
1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 \\
1 & -\frac{1}{2} & \alpha & -\alpha & 0
\end{pmatrix}
\] |
| 2 communicated variables | |

<table>
<thead>
<tr>
<th><strong>Unified DIGing (uDIG)</strong></th>
<th><strong>DIGing</strong></th>
</tr>
</thead>
</table>
| \[
\begin{pmatrix}
0 & -\alpha & -\alpha & 1 & 0 \\
0 & 0 & -1 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 \\
-\frac{L+m}{2} & 1 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0
\end{pmatrix}
\] | \[
\begin{pmatrix}
0 & -\alpha & 0 & 0 & 1 & 0 \\
0 & 0 & -1 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & -\alpha & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 0
\end{pmatrix}
\] |
| 2 communicated variables | |

<table>
<thead>
<tr>
<th><strong>Unified EXTRA (uEXTRA)</strong></th>
<th><strong>AugDGM</strong></th>
</tr>
</thead>
</table>
| \[
\begin{pmatrix}
0 & -\alpha & -\alpha & 1 & 0 \\
0 & 0 & -1 & L & 1 \\
1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & -L & 0 \\
0 & 1 & 0 & 0 & 0
\end{pmatrix}
\] | \[
\begin{pmatrix}
0 & 0 & 0 & 0 & 1 & -\alpha \\
0 & 0 & -1 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & -\alpha \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 0
\end{pmatrix}
\] |
| 2 communicated variables | |
Systematic analysis tool

LMI

Worst-case performance bound \( \rho \)

Algorithm

\[
\begin{bmatrix}
A & B_u & B_v \\
C_y & D_{yu} & D_{yv} \\
C_z & D_{zu} & D_{zv} \\
F_x & F_u & \end{bmatrix}
\]

Local functions \( \kappa \)

Communication network \( \sigma, B \)
Find $P \succ 0$, $Q \succ 0$, $R \succeq 0$, $S \succeq 0$, and $\lambda \geq 0$ such that

$$0 \succeq \Psi^T \left( \xi^T P \xi - \rho^2 (\xi^T P \xi) + \sum_{\ell=0}^{B-1} \lambda(\ell) \begin{bmatrix} y(\ell) \\ u(\ell) \end{bmatrix}^T \begin{bmatrix} -2 & \kappa + 1 \\ \kappa + 1 & -2\kappa \end{bmatrix} \begin{bmatrix} y(\ell) \\ u(\ell) \end{bmatrix} \right) \Psi$$

and

$$0 \succeq \xi^T Q \xi - \rho^2 (\xi^T Q \xi) + \sum_{\ell=0}^{B-1} \lambda(\ell) \begin{bmatrix} y(\ell) \\ u(\ell) \end{bmatrix}^T \begin{bmatrix} -2 & \kappa + 1 \\ \kappa + 1 & -2\kappa \end{bmatrix} \begin{bmatrix} y(\ell) \\ u(\ell) \end{bmatrix}$$

$$+ \begin{bmatrix} w(0) \\ w(B) \end{bmatrix}^T \left( \begin{bmatrix} \sigma^2 B & 0 \\ 0 & -1 \end{bmatrix} \otimes R \right) \begin{bmatrix} w(0) \\ w(B) \end{bmatrix}$$

$$+ \sum_{\ell=0}^{B-1} \begin{bmatrix} z(\ell) \\ w(\ell) \\ v(\ell) \\ w(\ell + 1) \end{bmatrix}^T \left( \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \otimes S(\ell) \right) \begin{bmatrix} z(\ell) \\ w(\ell) \\ v(\ell) \\ w(\ell + 1) \end{bmatrix}$$

Solved in Julia using Convex.jl and Mosek.
Local functions $\kappa$

Communication network $\sigma, B$

Algorithm

$$
\begin{bmatrix}
A & B_u & B_v \\
C_y & D_{yu} & D_{yv} \\
C_z & D_{zu} & D_{zv} \\
F_x & F_u & F_v
\end{bmatrix}
$$

Systematic analysis tool

LMI

Worst-case performance bound $\rho$
DIGing

\[ \kappa = 10, \; B = 1, \; \alpha \text{ optimizes DIGing bound} \]

- dashed lines are from (Nedić, Olshevsky, and Shi, 2017)
- solid line is from our LMI
- lower bound corresponds to only optimization and only consensus
\( \kappa = 10, \ B = 1, \ \alpha \) optimizes LMI

\[ \frac{(1 + \sigma)}{(1 - \sigma)} \]

- the bound in (Nedić, Olshevsky, and Shi, 2017) is vacuous
$\kappa = 10$, $\alpha$ optimizes LMI

![Graph showing iterations to converge with different $B$ values and a lower bound.](image-url)

- $B = 4$
- $B = 3$
- $B = 2$
- $B = 1$
- lower bound
SVL is designed to optimize the worst-case performance when $B = 1$. 

Sundararajan, Van Scoy, and Lessard (2020)
Algorithm comparison

\( \kappa = 10,\ B = 2,\ \alpha \) optimizes LMI

\[
\frac{1 + \sigma}{1 - \sigma}
\]

iterations to converge

\( \kappa = 10,\ B = 2,\ \alpha \) optimizes LMI

EXTRA
NIDS
DIGing
AugDGM
ExDIFF
uEXTRA
uDIG
SVL
lower bound
Algorithm comparison

\[ \kappa = 10, \ B = 3, \ \alpha \text{ optimizes LMI} \]

\[ \frac{1 + \sigma}{1 - \sigma} \]

\begin{align*}
10^0 & \quad 10^1 & \quad 10^2 \\
\text{iterations to converge} & \quad \text{lower bound}
\end{align*}

\begin{align*}
(1 + \sigma)/(1 - \sigma) & \quad 10^0 & \quad 10^1 & \quad 10^2 \\
\text{EXTRA} & \quad \text{NIDS} & \quad \text{DIGing} & \quad \text{AugDGM} & \quad \text{ExDIFF} & \quad \text{uEXTRA} & \quad \text{uDIG} & \quad \text{SVL}
\end{align*}
Algorithm comparison

\( \kappa = 10, \ B = 4, \ \alpha \) optimizes LMI

\[
\frac{1 + \sigma}{1 - \sigma}
\]
**Summary**

```
main contribution  =  sharp and systematic analysis
```

**open questions:**

- is SVL asymptotically optimal in the limit as $\sigma \to 1$?
- what is the optimal algorithm for $B > 1$?

```
https://vanscoy.github.io
```