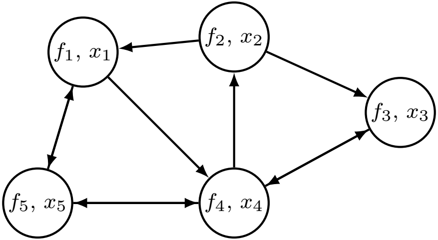
Systematic Analysis of Distributed Optimization Algorithms over Jointly-Connected Networks

Bryan Van Scoy
Miami University

Laurent Lessard
Northeastern University

IEEE Conference on Decision and Control December 16, 2020

$$
\begin{aligned}
\underset{x_1,\ldots,x_n}{\text{minimize}} \quad & \sum_{i=1}^{n} f_i(x_i) \\
\text{subject to} \quad & x_1 = x_2 = \ldots = x_n
\end{aligned}
$$



$$
W = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix}
$$

Want each agent to compute the global optimizer $x_\star$ by communicating with local neighbors and performing local computations.

DIGing

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

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

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

$$\|x_i(k) - x_\star\| \;=\; O\big(\rho^k\big)$$

iterations to converge $\propto\; -1/\log\rho$

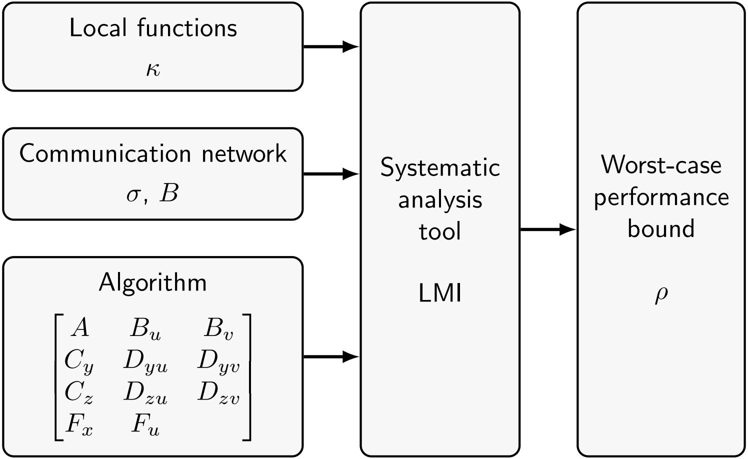
DIGing

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

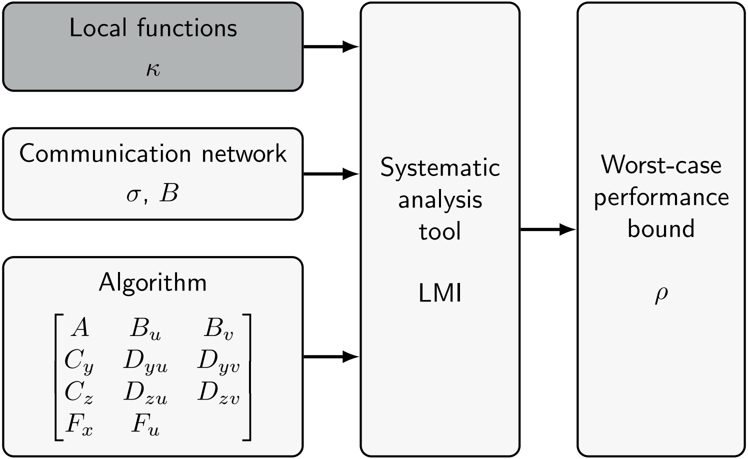
$$y_i(k+1) = \sum_{j=1}^{n} W_{ij}(k)\, y_j(k) + \nabla f_i\big(x_i(k+1)\big) - \nabla f_i\big(x_i(k)\big)$$

Issues:

- the bound grows with the number of agents
- the bound does not apply to large stepsizes
- the analysis is specific to DIGing



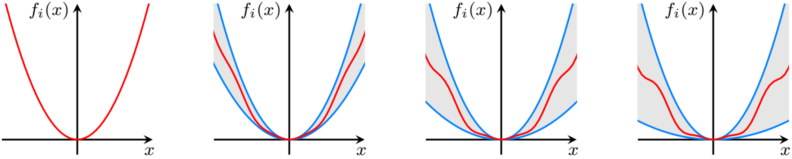
main contribution $=$ sharp and systematic analysis



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.

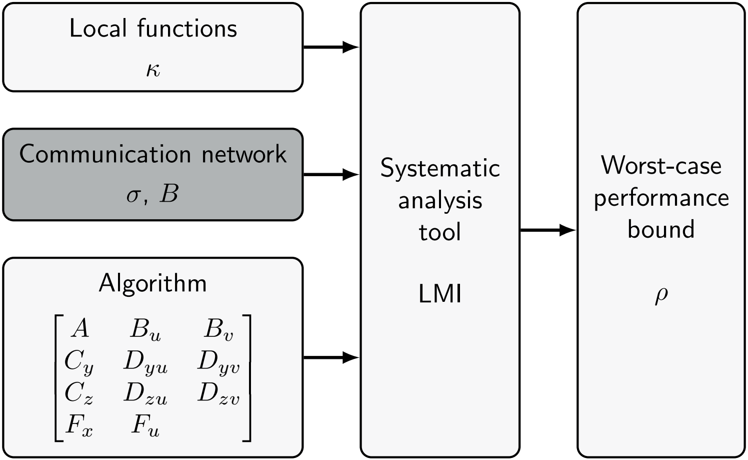


$\kappa = 1$ $\kappa = 2$ $\kappa = 4$ $\kappa = 8$

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

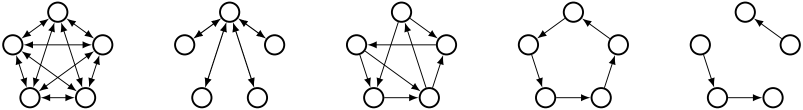
For all x,

$$\big(\nabla f_i(x) - \nabla f_i(x_\star) - m\,(x - x_\star)\big)^{\mathsf{T}}\big(\nabla f_i(x) - \nabla f_i(x_\star) - L\,(x - x_\star)\big) \le 0$$


Local functions
κ
Communication network
σ, 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
ρ

Spectral gap

The spectral gap characterizes the connectivity of a graph.



0.0 0.667 0.707 0.809 1.0

$$\text{spectral gap} \quad = \quad \left\| \tfrac{1}{n}\mathbf{1}\mathbf{1}^{\mathsf{T}} - W \right\|_2$$

Network assumptions

sparsity $W_{ij}(k) = 0$ if agent i does not receive information from agent j at time k

balanced $W(k)\mathbf{1} = W(k)^\mathsf{T}\mathbf{1} = \mathbf{1}$

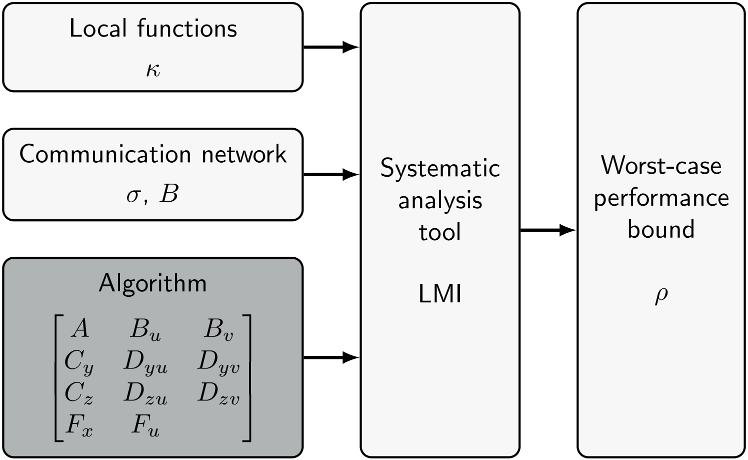
spectrum the spectral gap of $W(k)$ is less than or equal to one

joint spectrum for some fixed B, the spectral gap of

$$\Big[W(k)\,W(k+1)\cdots W(k+B-1)\Big]^{1/B}$$

is less than or equal to $\sigma < 1$

The spectral gap σ characterizes the connectivity of the time-varying network averaged over B iterations.


Local functions
κ
Communication network
σ, 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
ρ

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

$$u_i(k) = \nabla f_i\big(y_i(k)\big)$$

gradient computation

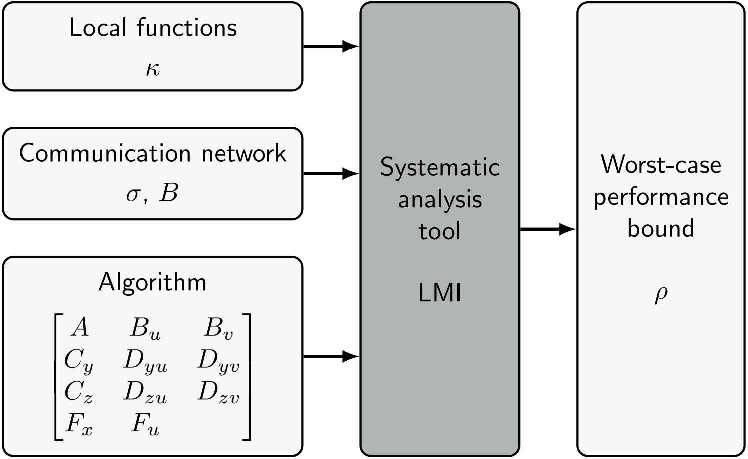
$$v_i(k) = \sum_{j=1}^{n} W_{ij}(k)\, z_j(k)$$

communication

$$0 = \sum_{j=1}^{n} \big(F_x\, x_j(k) + F_u\, u_j(k)\big)$$

invariant

		2 state variables		3 state variables
1 communicated variable	SVL	$\left[\begin{array}{cccc} 1-\gamma & \beta & -\alpha & -\gamma \\ -1 & 1 & 0 & -1 \\ \hline 1-\delta & 0 & 0 & \delta \\ \hline 1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & \end{array}\right]$	EXTRA	$\left[\begin{array}{ccccc} 1 & -\frac{1}{2} & \alpha & -\alpha & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 \\ \hline 1 & -\frac{1}{2} & 0 & 0 & 0 \\ \hline 1 & -1 & \alpha & 0 & \end{array}\right]$
	Exact Diffusion (ExDIFF)	$\left[\begin{array}{cccc} 1 & -1 & -\alpha & 1 \\ \frac{1}{2} & 0 & -\alpha & \frac{1}{2} \\ \hline 1 & 0 & -\frac{1}{2} & 0 \\ \hline 1 & 0 & 0 & 0 \\ \hline 1 & -1 & 0 & \end{array}\right]$	NIDS	$\left[\begin{array}{ccccc} 1 & -\frac{1}{2} & \frac{\alpha}{2} & -\frac{\alpha}{2} & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 \\ \hline 1 & -\frac{1}{2} & \frac{\alpha}{2} & -\frac{\alpha}{2} & 0 \\ \hline 1 & -1 & \alpha & 0 & \end{array}\right]$
2 communicated variables	Unified DIGing (uDIG)	$\left[\begin{array}{ccccc} 0 & -\alpha & -\alpha & 1 & 0 \\ \frac{L+m}{2} & 0 & -1 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 \\ -\frac{L+m}{2} & 1 & 1 & 0 & 0 \\ \hline 0 & 1 & 0 & & \end{array}\right]$	DIGing	$\left[\begin{array}{cccccc} 0 & -\alpha & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & -\alpha & 0 & 0 & 1 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 1 & -1 & 0 & & \end{array}\right]$
	Unified EXTRA (uEXTRA)	$\left[\begin{array}{ccccc} 0 & -\alpha & -\alpha & 1 & 0 \\ 0 & 0 & -1 & L & 1 \\ \hline 1 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & -L & 0 \\ \hline 0 & 1 & 0 & & \end{array}\right]$	AugDGM	$\left[\begin{array}{cccccc} 0 & 0 & 0 & 0 & 1 & -\alpha \\ 0 & 0 & -1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & -\alpha \\ \hline 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 1 & -1 & 0 & & \end{array}\right]$



Local functions
κ
Communication network
σ, 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
ρ

LMI

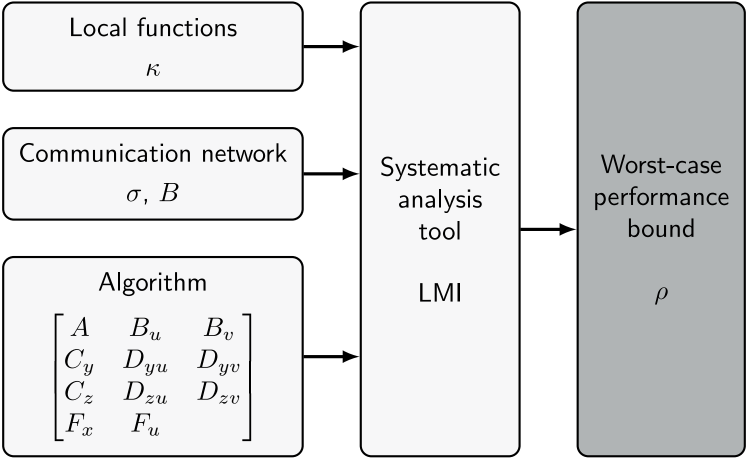
Find $P \succ 0$, $Q \succ 0$, $R \succeq 0$, $S \succeq 0$, and $\lambda \geq 0$ such that

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

and

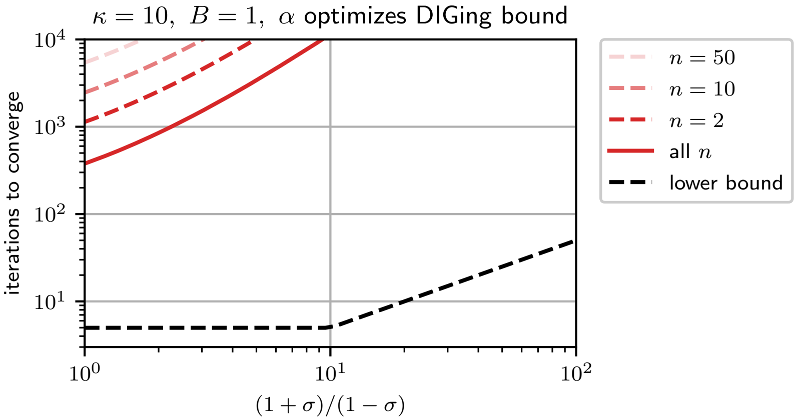
$$\begin{aligned} 0 \succeq \boldsymbol{\xi}_+^\mathsf{T} Q \, \boldsymbol{\xi}_+ - \rho^2 \, (\boldsymbol{\xi}^\mathsf{T} Q \, \boldsymbol{\xi}) &+ \sum_{\ell=0}^{B-1} \lambda(\ell) \begin{bmatrix} \boldsymbol{y}(\ell) \\ \boldsymbol{u}(\ell) \end{bmatrix}^\mathsf{T} \begin{bmatrix} -2 & \kappa+1 \\ \kappa+1 & -2\kappa \end{bmatrix} \begin{bmatrix} \boldsymbol{y}(\ell) \\ \boldsymbol{u}(\ell) \end{bmatrix} \\ &+ \begin{bmatrix} \boldsymbol{w}(0) \\ \boldsymbol{w}(B) \end{bmatrix}^\mathsf{T} \left(\begin{bmatrix} \sigma^{2B} & 0 \\ 0 & -1 \end{bmatrix} \otimes R \right) \begin{bmatrix} \boldsymbol{w}(0) \\ \boldsymbol{w}(B) \end{bmatrix} \\ &+ \sum_{\ell=0}^{B-1} \begin{bmatrix} \boldsymbol{z}(\ell) \\ \boldsymbol{w}(\ell) \\ \boldsymbol{v}(\ell) \\ \boldsymbol{w}(\ell+1) \end{bmatrix}^\mathsf{T} \left(\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \otimes S(\ell) \right) \begin{bmatrix} \boldsymbol{z}(\ell) \\ \boldsymbol{w}(\ell) \\ \boldsymbol{v}(\ell) \\ \boldsymbol{w}(\ell+1) \end{bmatrix} \end{aligned}$$

Solved in Julia using Convex.jl and Mosek.



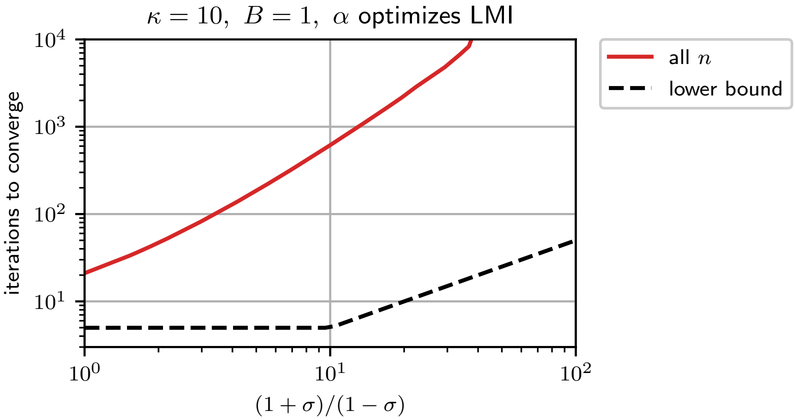
Local functions
κ
Communication network
σ, B
Algorithm
$\begin{bmatrix} A & B_u & B_v \\ C_y & D_{yu} & D_{yv} \\ C_z & D_{zu} & D_{zv} \\ F_x & F_u & \end{bmatrix}$
Systematic analysis tool
LMI
Worst-case performance bound
ρ

DIGing



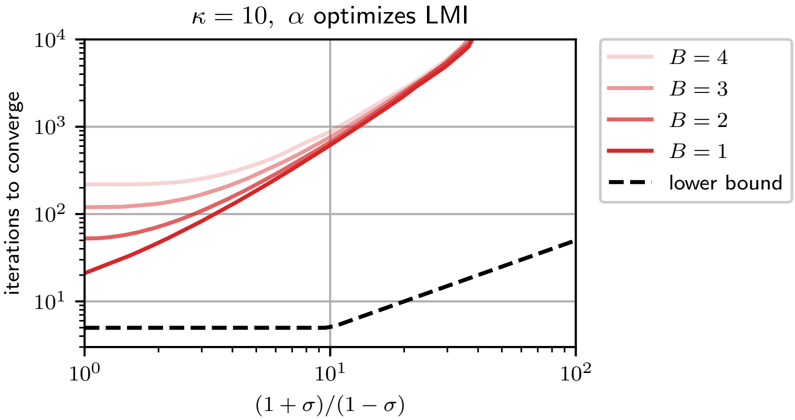
- dashed lines are from (Nedić, Olshevsky, and Shi, 2017)
- solid line is from our LMI
- lower bound corresponds to only optimization and only consensus

DIGing

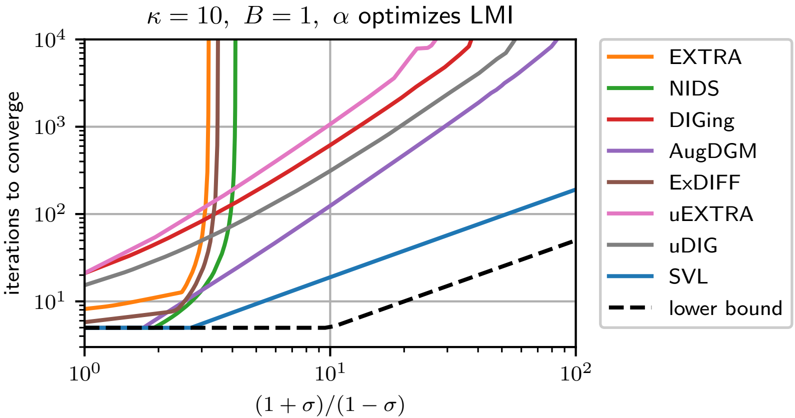


- the bound in (Nedić, Olshevsky, and Shi, 2017) is vacuous

DIGing


$\kappa = 10,\ \alpha$ optimizes LMI
iterations to converge
10^4
10^3
10^2
10^1
10^0
10^1
10^2
$(1+\sigma)/(1-\sigma)$
$B = 4$
$B = 3$
$B = 2$
$B = 1$
lower bound

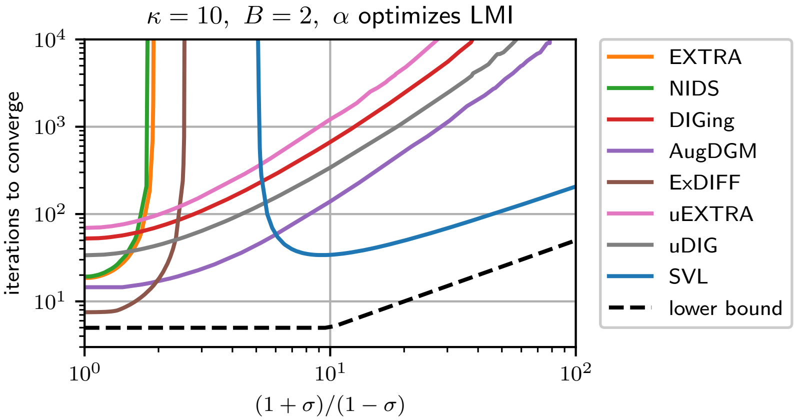
Algorithm comparison



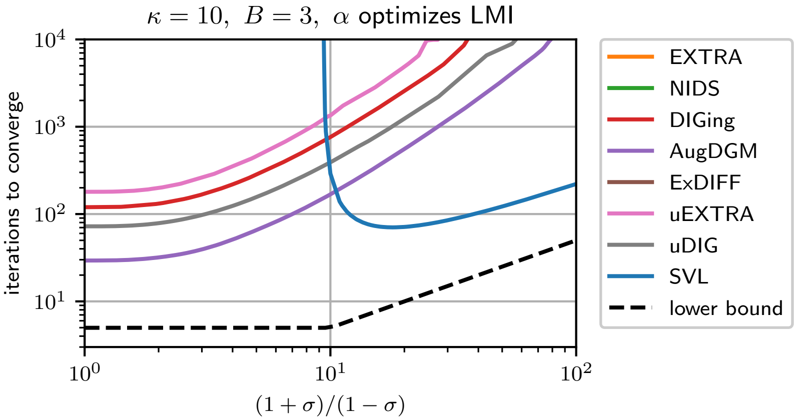
SVL is designed to optimize the worst-case performance when $B = 1$.

Sundararajan, Van Scoy, and Lessard (2020)

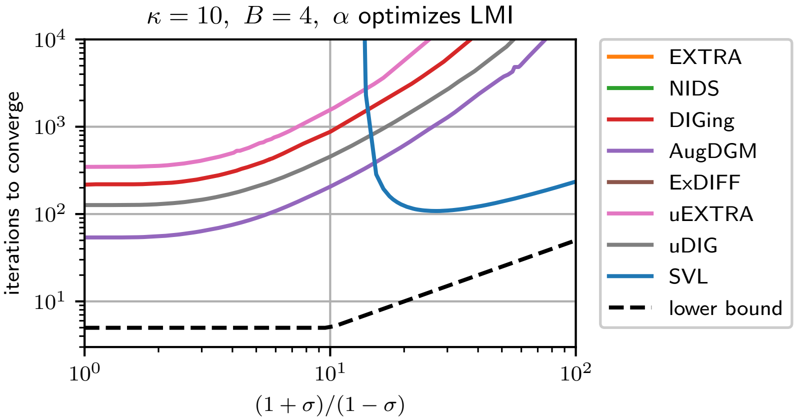
Algorithm comparison



Algorithm comparison



Algorithm comparison



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`