Learning to Relax:

Setting solver parameters across a sequence of linear system instances

Misha Khodak, Edmond Chow, Nina Balcan, Ameet Talwalkar

khodak@cmu.edu

Online learning to solve linear systems

Goal: solve T **s.p.d. linear systems**

- for each matrix-vector pair $(\mathbf{A}_1, \mathbf{b}_1), \dots, (\mathbf{A}_T, \mathbf{b}_T)$ find $\mathbf{x}$ s.t. $\|\mathbf{A}_t\mathbf{x} - \mathbf{b}_t\|_2 \le \varepsilon\|\mathbf{b}_t\|_2$ for $\varepsilon \ll 1$
- solve each using a solver **parameterized** by $\omega \in \Omega$ that takes $I(\mathbf{A}, \mathbf{b}, \omega)$ iterations.

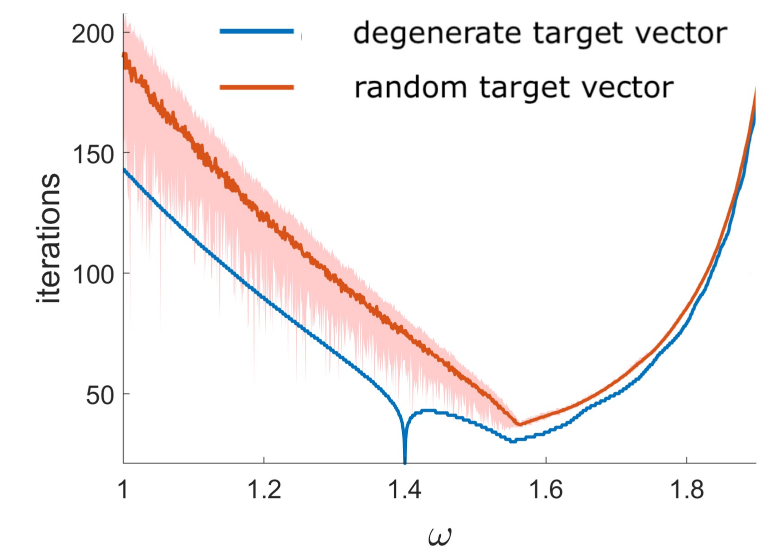
1. Can we find the best $\omega^* \in \Omega$ using only the iteration cost as feedback? In other words, can we **minimize regret**

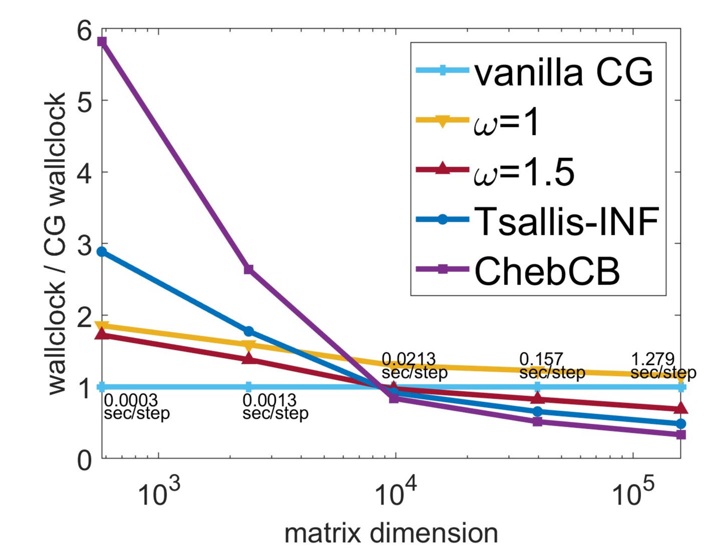
$$Reg(\omega^*) = \sum_{t=1}^{T} I(\mathbf{A}_t, \mathbf{b}_t, \omega_t) - \sum_{t=1}^{T} I(\mathbf{A}_t, \mathbf{b}_t, \omega^*)$$

2. If we have structural info, e.g. $\mathbf{A}_t = \mathbf{A} + c_t\mathbf{I}$, can we learn the **instance-optimal policy** $\pi^*: \mathbb{R} \mapsto \Omega$

$$Reg(\pi^*) = \sum_{t=1}^{T} I(\mathbf{A}_t, \mathbf{b}_t, \omega_t) - \sum_{t=1}^{T} I(\mathbf{A}_t, \mathbf{b}_t, \pi^*(c_t))$$

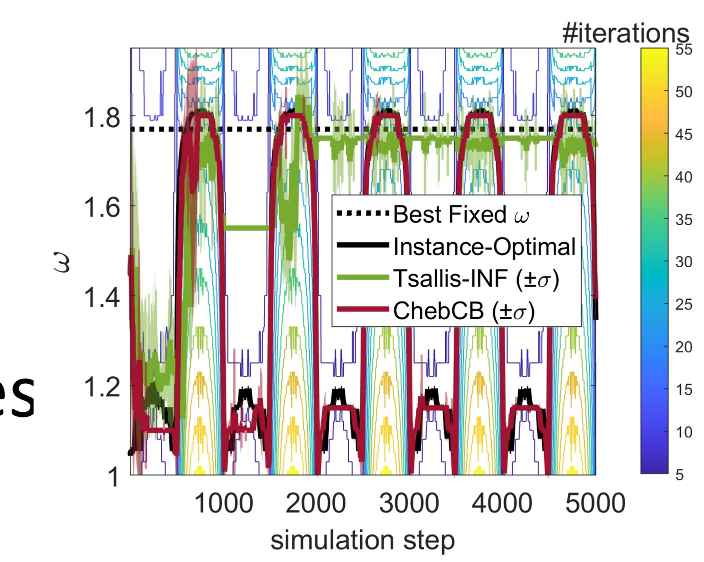
References:
[1] Abernethy, Lee, Tewari. NeurIPS 2015.
[2] Auer, Cesa-Bianchi, Freund, Schapire. SIAM J. Computing.
[3] Foster, Rakhlin. ICML 2020.





Applications:

- numerical simulation
- solving nonlinear systems
- many graphics subroutines
- …



For example:
solving the heat equation with conductivity κ_t involves matrices $\mathbf{A}_t = \frac{c}{\kappa_t}\mathbf{I} - \nabla^2$, where c is a constant and ∇ is a discrete Laplacian

ChebCB algorithm:
set uniform discretization $\mathbf{g} \in \Omega^d$
for instance $t = 1, \dots, T$

- fit truncated Chebyshev series to costs $I(\mathbf{A}_s, b_s, \mathbf{g}_{[i_s]})$ at context, parameters pairs $(c_s, \mathbf{g}_{[i_s]})$
- sample i_t via inverse gap-weighting with weights determined by applying the fitted regressors to $(c_t, \mathbf{g}_{[i_t]})$
- solve $\mathbf{A}_t\mathbf{x} = \mathbf{b}_t$ using parameter $\mathbf{g}_{[i_t]}$

Learning the relaxation parameter ω of SOR

We study **successive over-relaxation** (SOR), a linear solver parameterized by $\omega \in [1,2)$ often used as a preconditioner for conjugate gradient (CG) or as a multigrid smoother. To prevent degeneracy, we assume $\mathbf{b}_t$ are drawn i.i.d. from a truncated Gaussian.

1. Standard bandit algos (e.g. Exp3 [2] / Tsallis-INF [1]) attain sublinear regret w.r.t. a fixed relaxation parameter:

$$Reg(\omega^*) = O\left(T^{2/3}\log\tfrac{1}{\varepsilon}\right)$$

2. We introduce **ChebCB**, which incorporates Chebyshev regression into a standard contextual bandit algo [3]. It attains sublinear regret w.r.t. the instance optimal parameters:

$$Reg\ (\pi^*) = O\left(\sqrt{n}T^{9/11}\log\tfrac{1}{\varepsilon}\right)$$

See paper for non-stochastic analysis, SOR-preconditioned CG results, and new proof techniques for data-driven algos.
Future: multi-parameter preconditioners, non-s.p.d. solvers