# Curvature-Exploiting Acceleration of Elastic Net Computation

**Vien V. Mai and Mikael Johansson**
**KTH - Royal Institute of Technology**

## The elastic net problem

Workhorse in ML and modern statistics

$$\underset{x \in \mathbb{R}^d}{\text{minimize}} \left\{ \frac{1}{2n} \|Ax - b\|_2^2 + \frac{\gamma_2}{2} \|x\|_2^2 + \gamma_1 \|x\|_1 \right\}$$

Special instances: $\gamma_1 = 0 \Rightarrow$ Ridge regression; $\gamma_2 = 0 \Rightarrow$ Lasso

In many real-world data sets, Hessian of the smooth part

$$\nabla^2 f(x) = \frac{1}{n} A^\top A + \gamma_2 I = C + \gamma_2 I$$

has **rapidly decaying spectrum**.

## Related work

**Deterministic first-order methods:**

- PGD: $O\left(dn\kappa \log \frac{1}{\epsilon}\right)$
- `FISTA`: $O\left(dn\sqrt{\kappa} \log \frac{1}{\epsilon}\right)$

$$\kappa = \frac{\lambda_1(\boldsymbol{C}+\gamma_2 \mathbf{I})}{\lambda_d(\boldsymbol{C}+\gamma_2 \mathbf{I})}$$

**Stochastic first-order methods:**

- `ProxSVRG`: $O\left(d\left(n+\tilde{\kappa}\right) \log \frac{1}{\epsilon}\right)$
- Katyusha: $O\left(d\left(n+\sqrt{n\tilde{\kappa}}\right) \log \frac{1}{\epsilon}\right)$

$$\tilde{\kappa} = \frac{\mathrm{tr}(\boldsymbol{C}+\gamma_2 \mathbf{I})}{\lambda_d(\boldsymbol{C}+\gamma_2 \mathbf{I})}$$

**Challenge:** exploit second-order information despite non-smoothness.

## Main contribution

Novel 2nd-order optimization algorithm computes $\varepsilon$-optimal solution in time

$$\mathcal{O}(d(n + c\,\tilde{\kappa}) \log \frac{1}{\varepsilon})$$

Stochastic first-order methods have $c = 1$, our method has

$$c = \frac{r\lambda_r + \sum_{i>r} \lambda_i}{\sum_{i=1}^{r} \lambda_i + \sum_{i>r} \lambda_i} \ll 1$$

Dramatic improvement when $C$ has rapidly decaying spectrum

## Proposed algorithm

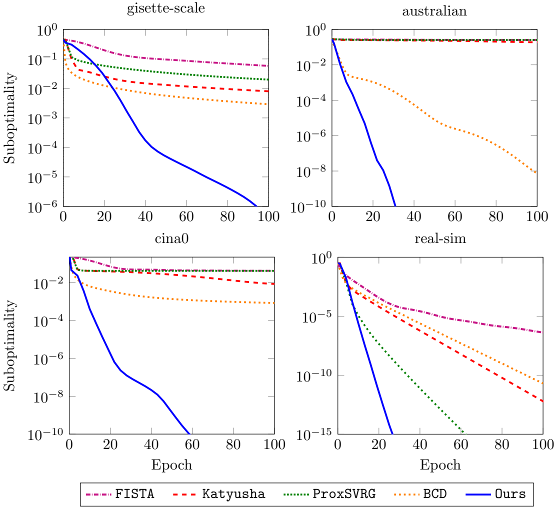
Two building blocks:

**1.** Approximation of smooth Hessian using randomized block Lanczos

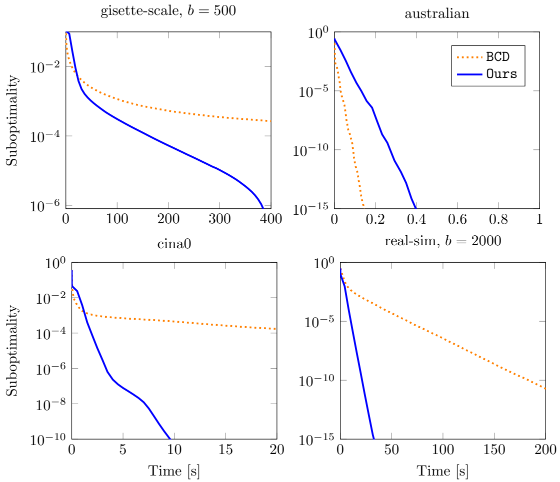
**2.** Proximal Newton method with stochastic gradients

- Exploits finite-sum structure
- Uses momentum acceleration to increase mini-batch size
- Makes clever use of error control and warm start

# Experimental results: suboptimality vs. iteration counts



# Experimental results: suboptimality vs. run-times



## Thank you!

Please come visit our poster at:

**Room Pacific Ballroom #196**

Code: `https://github.com/vienmai/elasticnet`