# Average-case Acceleration Through Spectral Density Estimation

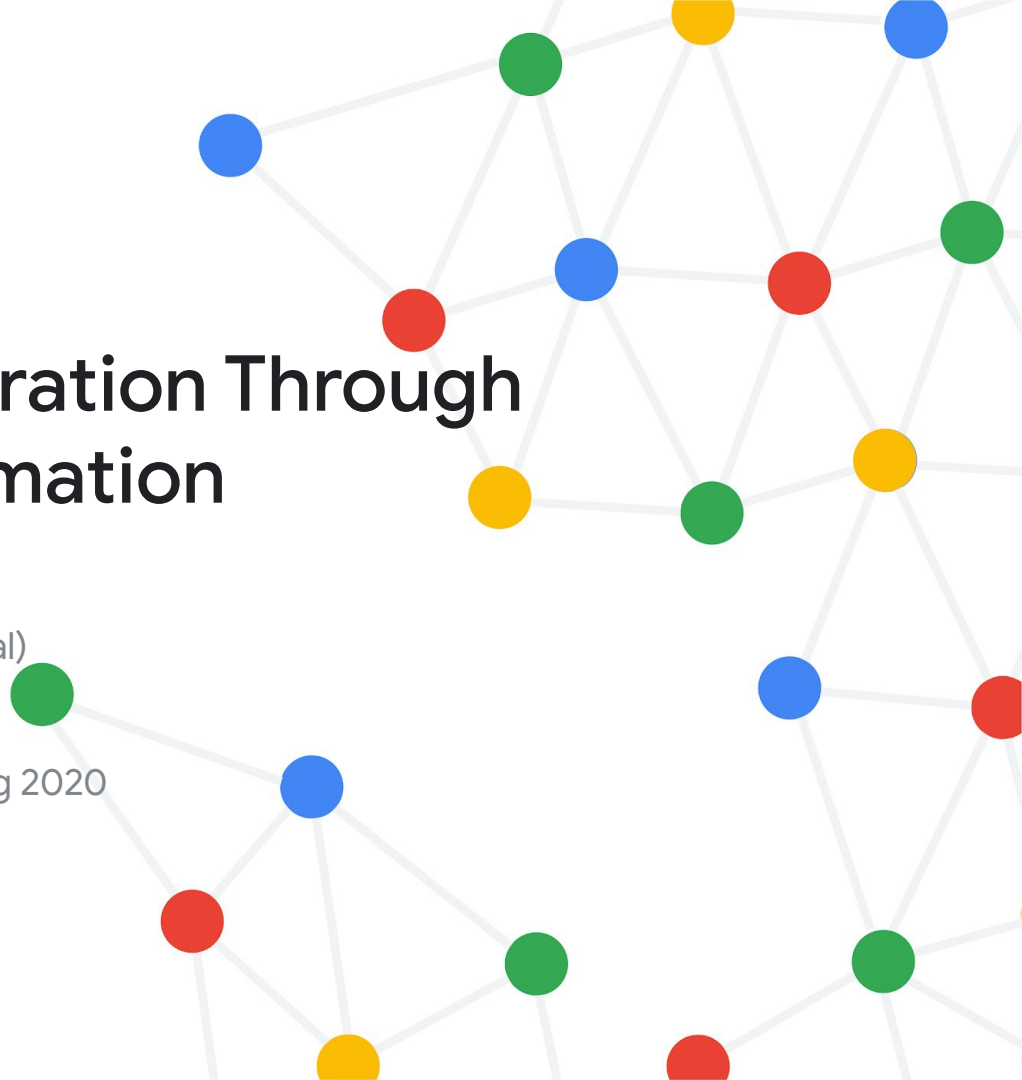
Fabian Pedregosa (Google Research)

Damien Scieur (Samsung SAIT AI Lab, Montréal)

International Conference on Machine Learning 2020

Google Research

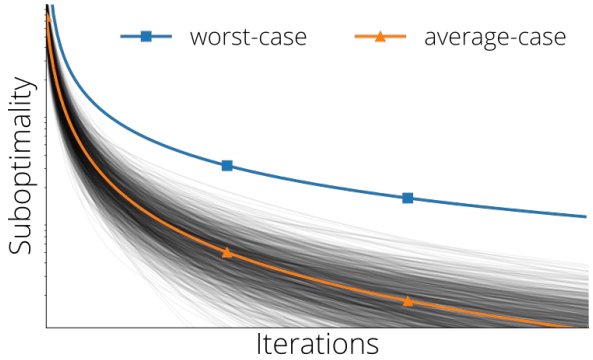
SAMSUNG
Advanced Institute
of Technology AI Lab
Montreal



# Complexity Analysis in Optimization

Worst-case analysis

- ✓ Bound on the complexity for *any* input.
- ✗ Potentially worse than observed runtime.




Simplex method (Dantzig, '98, Spielman & Teng '04)

- ✗ Exponential worst-case.
- ✓ Runtime typically polynomial.

Google Research

# Average-case Complexity

- ✓ Complexity *averaged* over all problem instances.
- ✓ Representative of the typical complexity.

Better bounds, sometimes better algorithms

→ Quicksort (Hoare ’62): Fast average-case sorting

*Rarely used in optimization*

Google Research

## Main contributions

Average-case analysis for optimization on quadratics.

Optimal methods under this analysis.

# Problem Distribution: Random Quadratics

$$\min_{x \in \mathbb{R}^d} \left\{ f(x) \overset{\text{def}}{=} \frac{1}{2}(x - x^\star)^\top H (x - x^\star) \right\},$$

where $H$, $x^\star$ are random matrix, vector.

- ✓ exact runtime known depends on eigenvalues(**H**).
- ✓ shares (some) dynamics of real problems, e.g., Neural Tangent Kernel (Jacot et al., 2018).

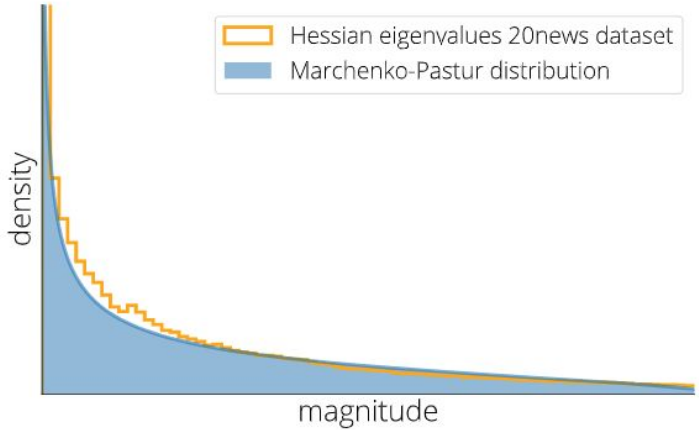
Google Research

# Example: Random Least Squares

$$\min_{\boldsymbol{x}\in\mathbb{R}^d} \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|^2, \text{ with } \boldsymbol{b}=\boldsymbol{A}\boldsymbol{x}^\star$$

When elements of **A** are iid, standardized:

Spectrum of **H** will be close to Marchenko-Pastur.




Google Research

# Expected Error For Gradient-Based Methods

$$\text{expected error} = \mathbb{E}\|\boldsymbol{x}_t - \boldsymbol{x}^\star\|^2 = \overbrace{R^2}^{\text{initialization}} \int_{\mathbb{R}} \overbrace{P_t^2}^{\text{algorithm}} \overbrace{\mathrm{d}\mu}^{\text{problem}}$$

$R^2$ is the distance to optimum at initialization $\mathbb{E}\|\boldsymbol{x}_0 - \boldsymbol{x}^\star\|^2$.

Fixed

Problem difficulty represented by expected density Hessian eigenvalue $\mathrm{d}\mu$

$P_t$ is a polynomial of degree t determined from the optimization algorithm.

Flexible: algorithm design

Google Research

# Average-case Optimal Method

**Goal**: Find method with minimal expected error = $R^2 \int_{\mathbb{R}} P_t^2 \, d\mu$ (initialization: $R^2$; algorithm: $P_t^2$; problem: $d\mu$)

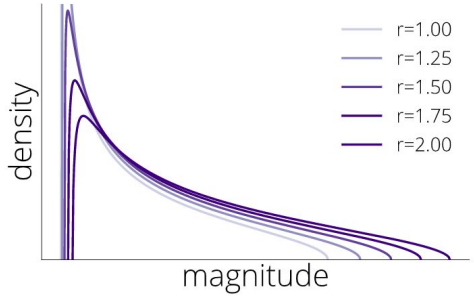
**Algorithms ↔ Polynomials**

Find **polynomial** $P_t$ of degree t that minimizes expected error (with proper normalization).

**Solution:** Polynomial of degree t, orthogonal wrt to $\lambda d\mu(\lambda)$ .

Google Research

# Marchenko-Pastur Acceleration




**Model for** $d\mu$ = Marchenko-Pastur(r, $\sigma$).

r and $\sigma$ estimated from:

- Largest eigenvalue
- Trace of H

No need to know strong convexity constant.

**Algorithm**

$$\boldsymbol{x}_t = \boldsymbol{x}_{t-1} + \underbrace{a_t(r,\sigma)}_{\text{momentum}}(\boldsymbol{x}_{t-2} - \boldsymbol{x}_{t-1}) + \underbrace{b_t(r,\sigma)}_{\text{step-size}}\nabla f(\boldsymbol{x}_{t-1})$$

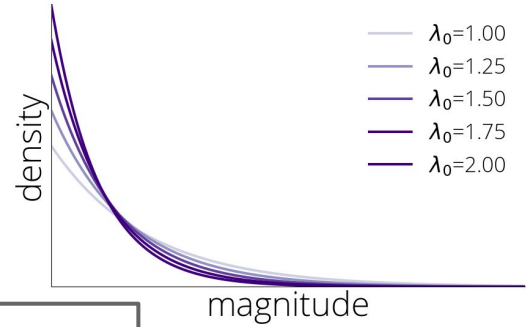
Simple momentum-like method, low memory requirements.

Google Research

# Decaying Exponential Acceleration

**Model for** $d\mu$ = decaying exponential($\lambda_0$).

**Unbounded** largest eigenvalue. Only access to Tr(**H**).



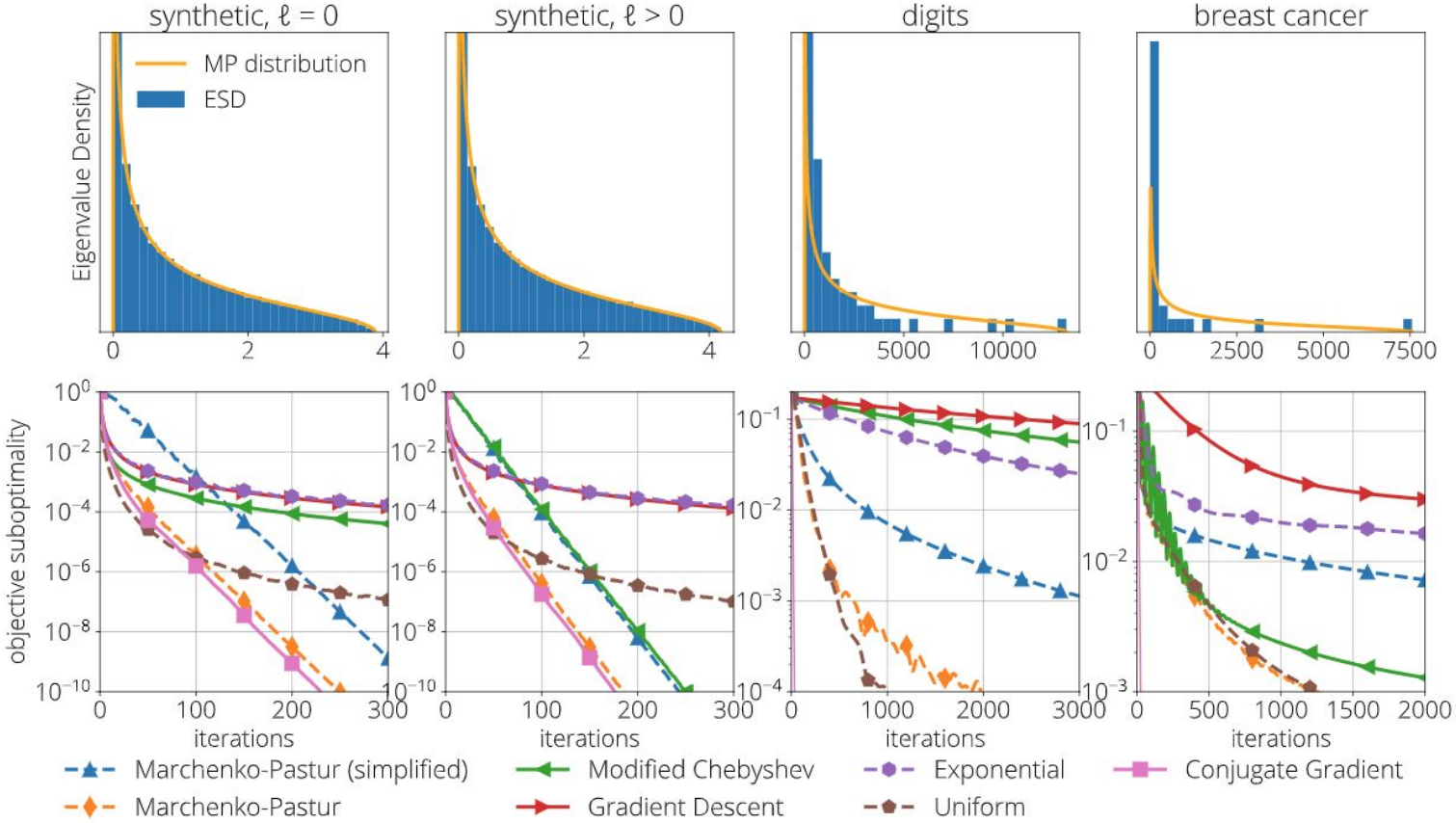

**Algorithm**

$$\boldsymbol{x}_t = \boldsymbol{x}_{t-1} + \frac{t-1}{t+1}(\boldsymbol{x}_{t-1} - \boldsymbol{x}_{t-2}) - \frac{\lambda_0}{t+1}\nabla f(\boldsymbol{x}_{t-1})$$

- Decaying step-size
- Similar to Polyak averaging

Google Research

# Benchmarks: Least Squares



# Conclusions

Average-case analysis based on random quadratics.

Optimal methods under different eigenvalue distribution.

✓ Acceleration without knowledge of strong convexity.

In **paper**

\+ More methods, convergence rates, empirical extension to non-quadratic objectives.

**Follow-up work** on asymptotic analysis

(Scieur and P., *"Universal Average-Case Optimality of Polyak Momentum"*)

Google Research