Lasso Screening Rules via Dual Polytope Projection

Jie Wang, Jiayu Zhou, Peter Wonka, Jieping Ye
Compute Science and Engineering | Arizona State University

Poster
Sat81

Standard Lasso

$$\beta^*(\lambda) = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda\|\beta\|_1.$$

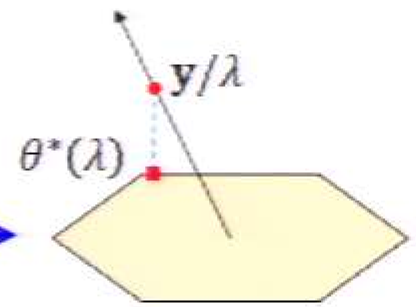
KKT 1

$$\mathbf{y} = -\mathbf{X}\beta^*(\lambda) + \lambda\theta^*(\lambda)$$

Dual Problem of Lasso

$$\theta^*(\lambda) = \sup_{\theta \in \mathbb{R}^n} \frac{1}{2}\|\mathbf{y}\|_2^2 - \frac{\lambda^2}{2}\left\|\theta - \frac{\mathbf{y}}{\lambda}\right\|_2^2, |x_i^T\theta| \le 1, i = 1, 2, \cdots, p.$$

$$\theta^*(\lambda) = P_{\mathcal{F}}\left(\frac{\mathbf{y}}{\lambda}\right)$$



Key Ingredients of DPP Rules

- KKT 2

$$(\theta^*(\lambda))^T \mathbf{x}_i \in \begin{cases} \text{sign}([\beta^*(\lambda)]_i), & \text{if } [\beta^*(\lambda)]_i \neq 0 \\ [-1, 1], & \text{if } [\beta^*(\lambda)]_i = 0 \end{cases}$$

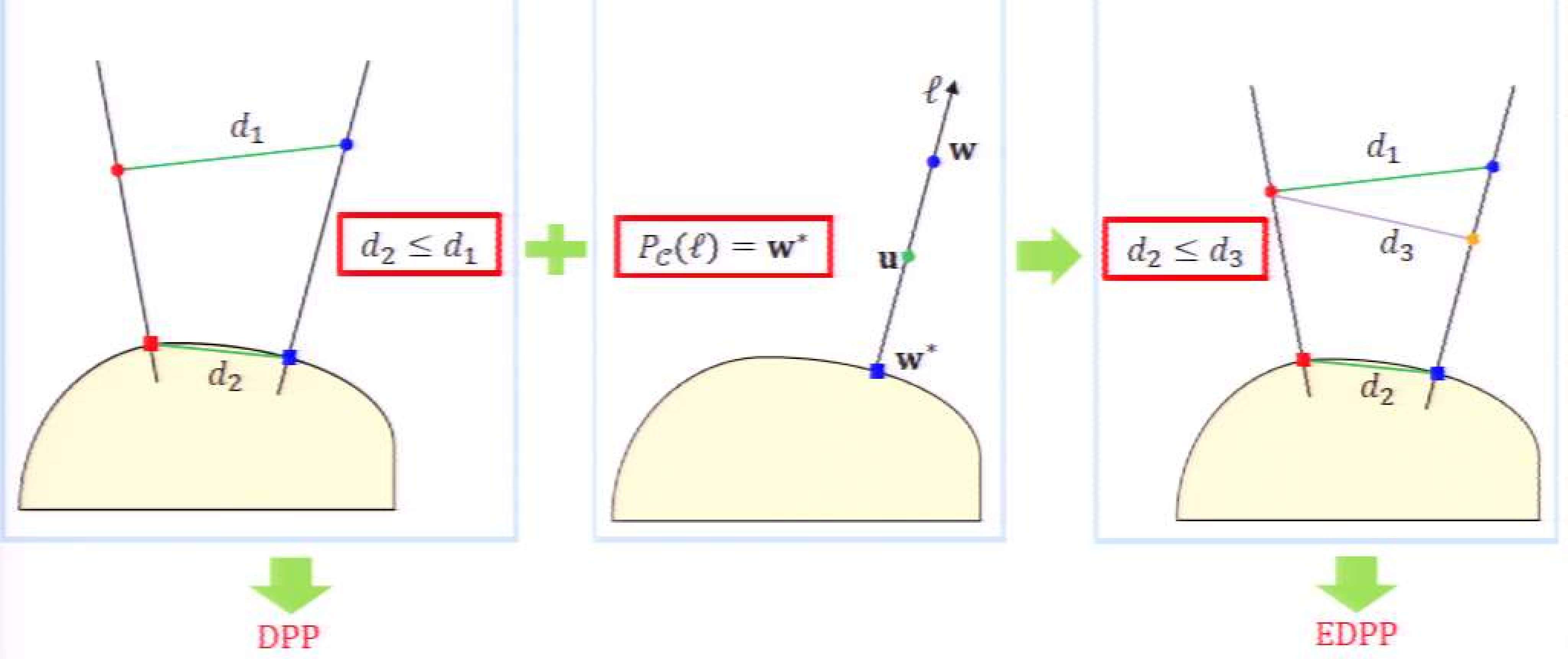
$|(\theta^*(\lambda))^T \mathbf{x}_i| < 1 \Rightarrow [\beta^*(\lambda)]_i = 0$ ← NOT applicable ($\theta^*(\lambda)$: unknown)

↓ relaxed

$$\max_{\theta \in \boldsymbol{\Theta}} |(\theta)^T \mathbf{x}_i| < 1 \Rightarrow [\beta^*(\lambda)]_i = 0, \quad \text{s.t.} \quad \theta^*(\lambda) \in \boldsymbol{\Theta}.$$

← General testing rule

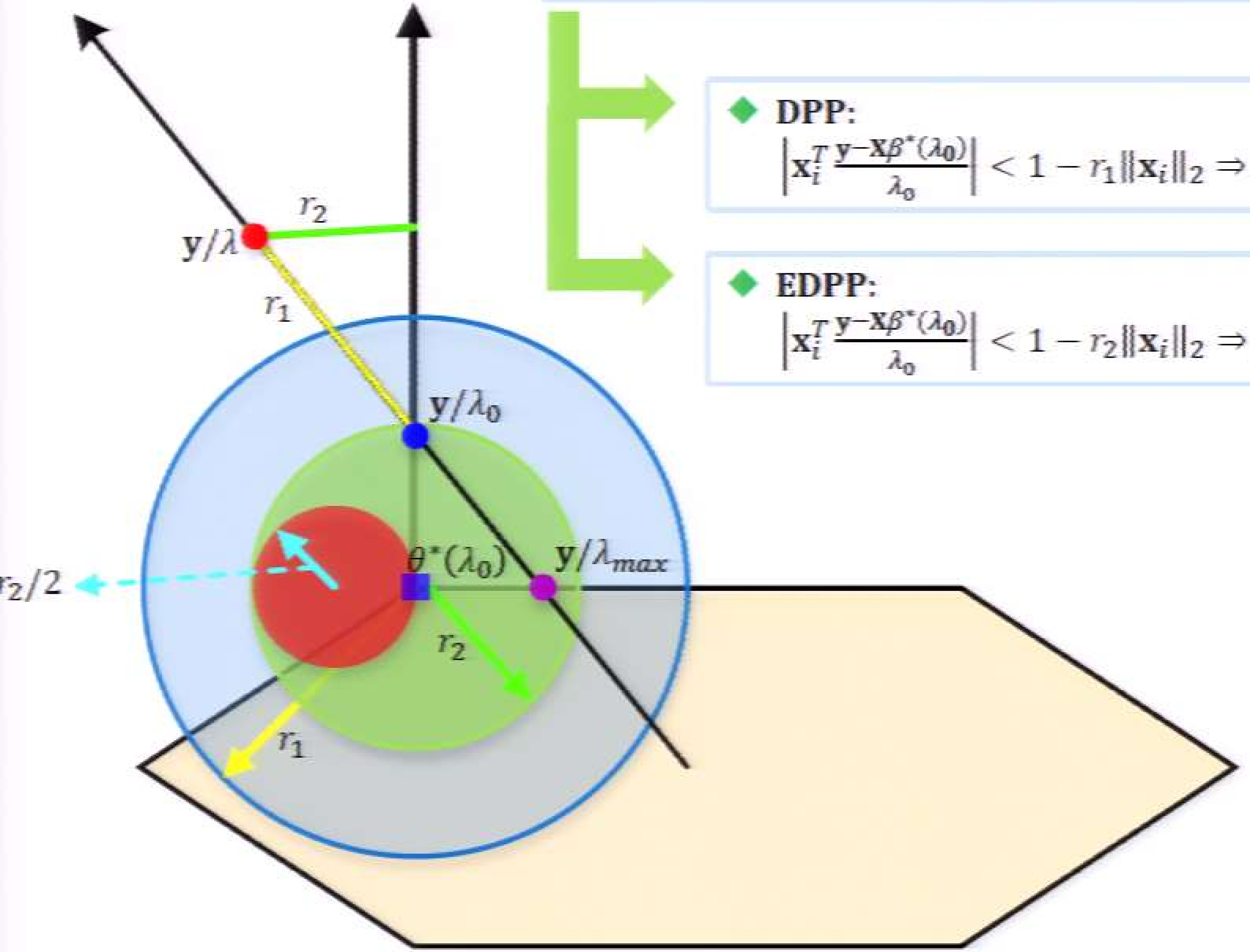
- Nonexpansiveness
- Projections of Ray



Geometric Intuitions of DPP Rules

- Where is $\theta^*(\lambda) = P_{\mathcal{F}}(\mathbf{y}/\lambda)$ given $\theta^*(\lambda_0)$?

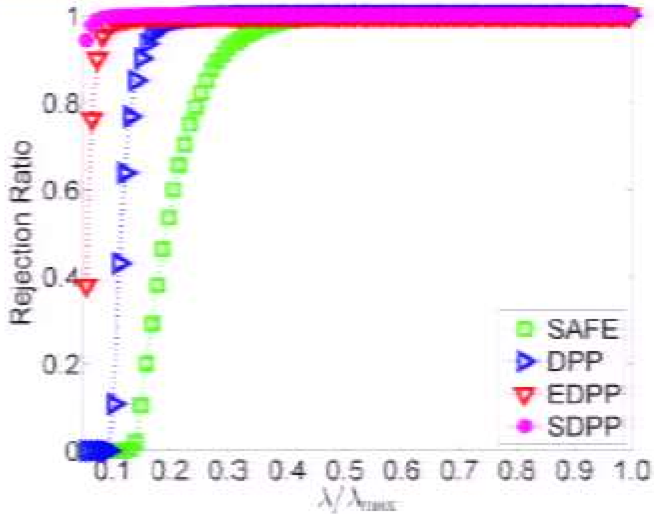
- Let $\lambda_{max} = \max_i |\mathbf{x}_i^T \mathbf{y}|$. If $\lambda > \lambda_{max}$, $\beta^*(\lambda) = 0, \theta^*(\lambda) = \mathbf{y}/\lambda$. Otherwise, assume $\beta^*(\lambda_0)$ is known.



- **DPP:**
$$\left|\mathbf{x}_i^T \frac{\mathbf{y} - \mathbf{X}\beta^*(\lambda_0)}{\lambda_0}\right| < 1 - r_1 \|\mathbf{x}_i\|_2 \Rightarrow [\beta^*(\lambda)]_i = 0.$$

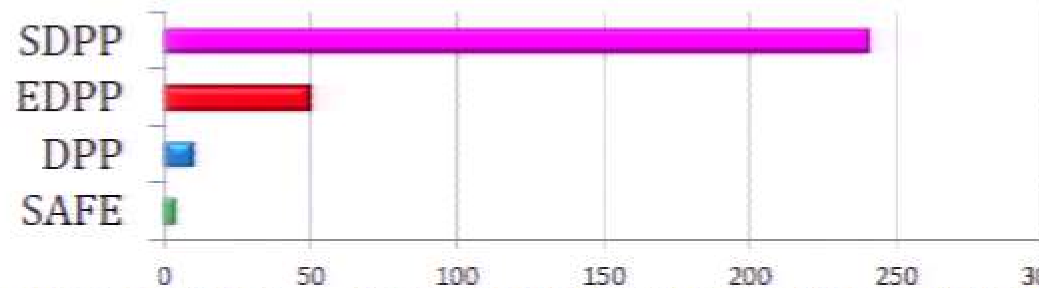
- **EDPP:**
$$\left|\mathbf{x}_i^T \frac{\mathbf{y} - \mathbf{X}\beta^*(\lambda_0)}{\lambda_0}\right| < 1 - r_2 \|\mathbf{x}_i\|_2 \Rightarrow [\beta^*(\lambda)]_i = 0.$$

Results

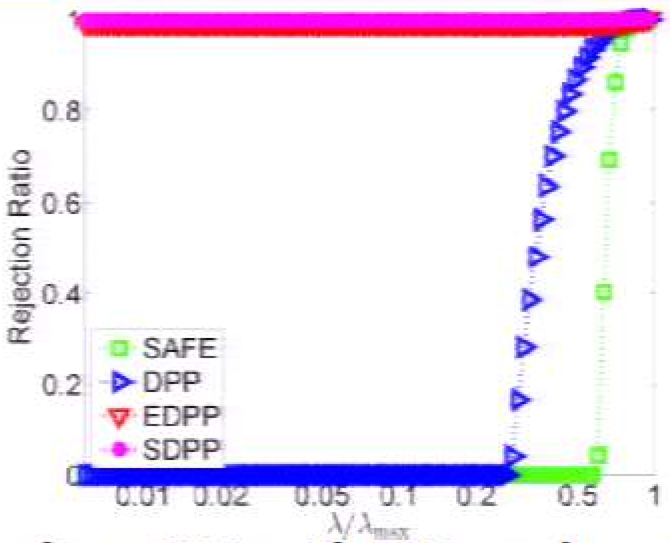


	solver	SAFE	DPP	EDPP	SDPP
time (s)	2245.26	685.12	233.85	45.56	9.34

Speedup

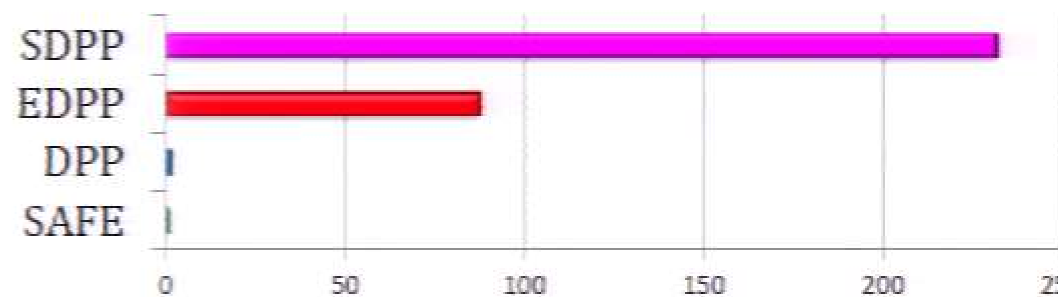


Results on MNIST along a sequence of 100 parameters (λ/λ_{max}) evenly distributed in (0.05, 1). Data matrix is of size 784×50,000.



	solver	SAFE	DPP	EDPP	SDPP
time (s)	38258.72	37882.41	31214.41	436.66	164.99

Speedup



Results on SNPs with 747 samples and 504,095 features. The parameter sequence contains 100 parameters (λ/λ_{max}) evenly distributed in (0.005, 1) along the log scale.

Our method can be extended to Group Lasso, Fused Lasso, Mixed-norm Regularization, SVM, LAD etc.