High-Dimensional Private ERM

by Greedy Coordinate Descent

Paul Mangold[1], Aurélien Bellet[1], Joseph Salmon[2], Marc Tommasi[1]

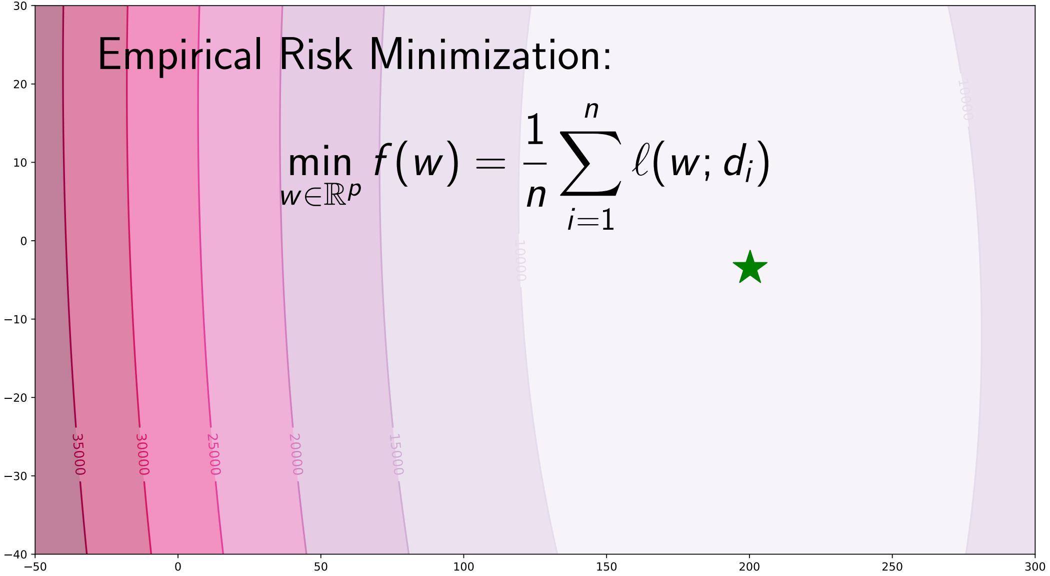
[1]Inria Lille [2]Univ. Montepellier

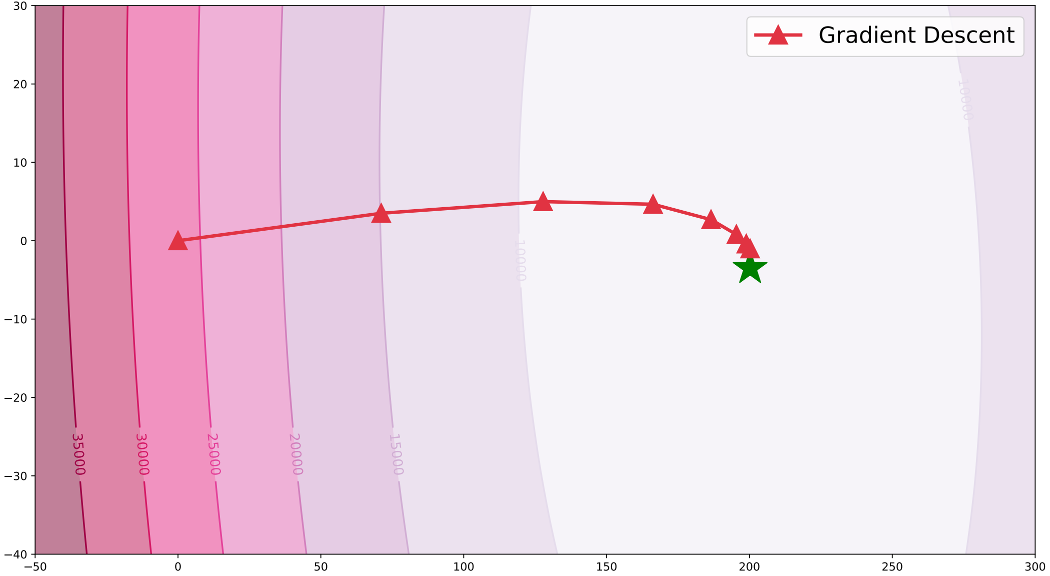
3PML Workshop @Meta

November 9, 2022

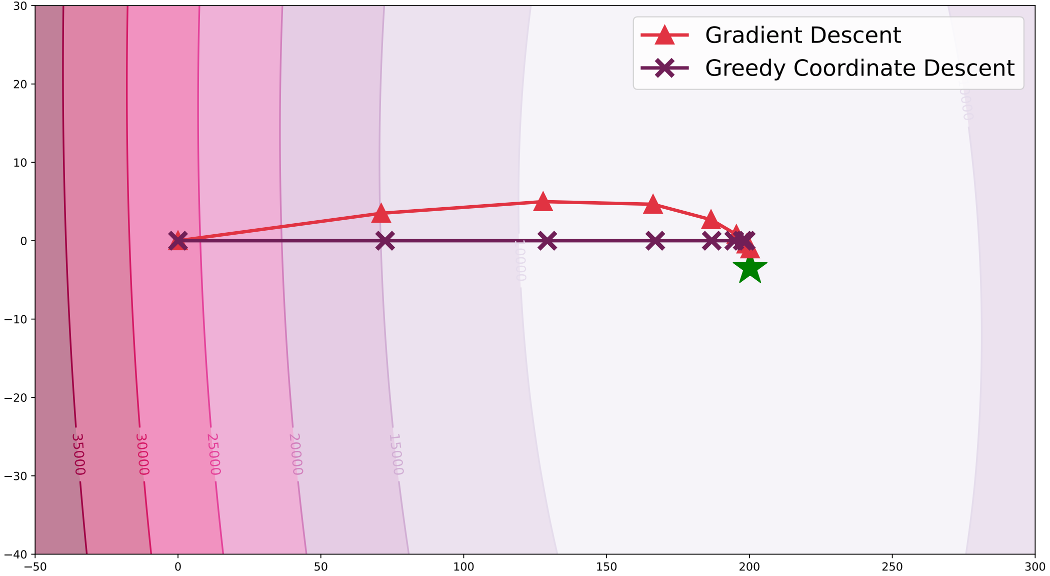
Empirical Risk Minimization:

$$\min_{w \in \mathbb{R}^p} f(w) = \frac{1}{n} \sum_{i=1}^{n} \ell(w; d_i)$$





Gradient Descent
35000
30000
25000
20000
15000
10000
10000
30
20
10
0
−10
−20
−30
−40
−50
0
50
100
150
200
250
300



Gradient Descent
Greedy Coordinate Descent
35000
30000
25000
20000
15000
10000
30
20
10
0
−10
−20
−30
−40
−50
0
50
100
150
200
250
300

Differentially Private ERM

$$w^{\text{priv}} \approx \arg\min_{w \in \mathbb{R}^p} f(w) = \frac{1}{n} \sum_{i=1}^{n} \ell(w; d_i)$$

such that w^{priv} is (ϵ, δ)-DP

Differential Privacy

$\mathcal{A} : D \mapsto w^{\text{priv}}$ is (ϵ, δ)-*Differentially Private*

$$\Pr\left[\mathcal{A}(D) \in \mathcal{S}\right] \leq e^{\epsilon} \Pr\left[\mathcal{A}(D') \in \mathcal{S}\right] + \delta$$

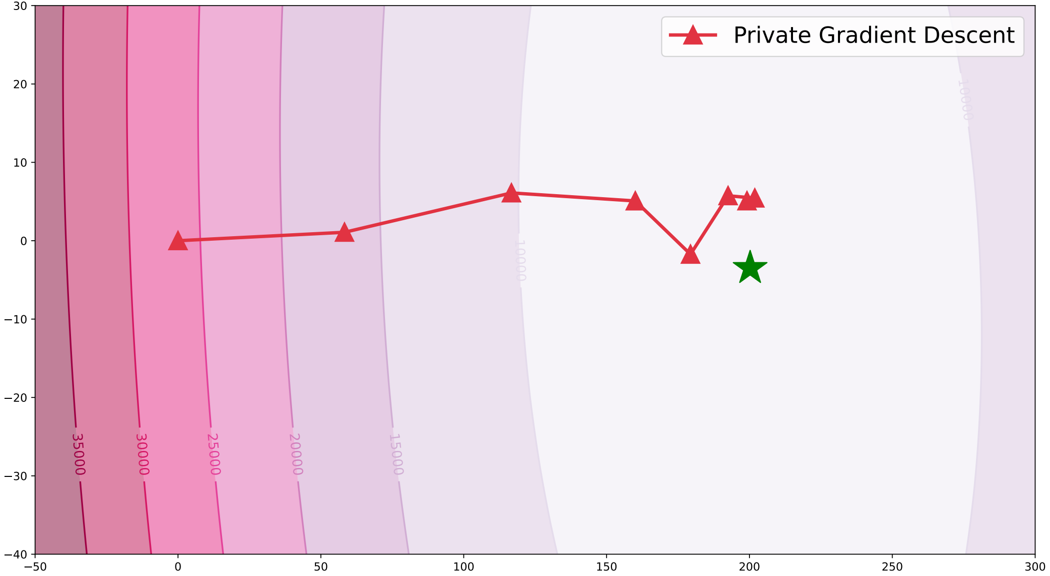
(where D and D' differ on one element)

Private Gradient Descent

For T iterations:

$$w^{t+1} = w^t - \eta \left(\nabla f(w^t) + \mathcal{N}(\sigma^2 \mathbb{1}_p) \right)$$

Noise scale: $\sigma \propto \frac{\sqrt{Tp}}{n\epsilon}$



Private Gradient Descent
35000
30000
25000
20000
15000
10000

Utility: $\mathbb{E}[f(w) - f^*] = \ ?$

assuming f and ∇f are Lipschitz

Utility: $\mathbb{E}[f(w) - f^*] = \ ?$

assuming f and ∇f are Lipschitz

- Convex: $\widetilde{O}\left(\frac{\sqrt{p}}{n\epsilon}\right)$
- Strongly-Convex: $\widetilde{O}\left(\frac{p}{n^2\epsilon^2}\right)$

Can we choose updates
“more wisely”?

Private **Greedy** CD

For T iterations:

$$w_j^{t+1} = w_j^t - \eta_j \left(\nabla_j f(w^t) + \mathbf{Lap}(\lambda_j) \right)$$

where $j = \arg\max_{j' \in [p]} \left| \nabla_{j'} f(w^t) + \mathbf{Lap}(\lambda_{j'}) \right|$

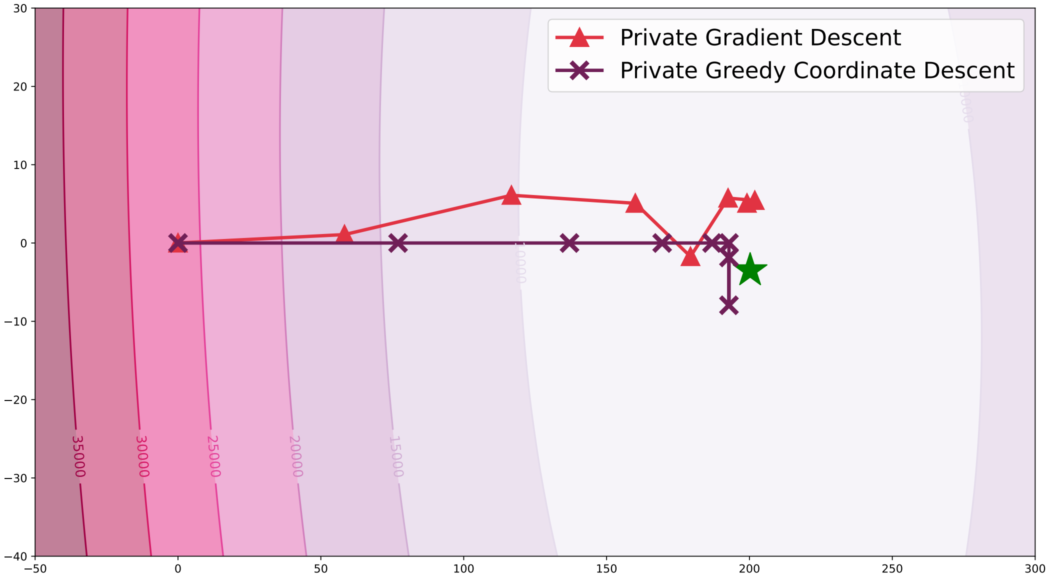
Private **Greedy** CD

For T iterations:

$$w_j^{t+1} = w_j^t - \eta_j \left(\nabla_j f(w^t) + \mathbf{Lap}(\lambda_j) \right)$$

where $j = \underset{j' \in [p]}{\arg\max} \left| \nabla_{j'} f(w^t) + \mathbf{Lap}(\lambda_{j'}) \right|$

Noise scale: $\lambda_j \propto \frac{\sqrt{T}}{n\epsilon}$, independent on the dimension!!


Private Gradient Descent
Private Greedy Coordinate Descent
35000
30000
25000
20000
15000
10000
30
20
10
0
−10
−20
−30
−40
−50
0
50
100
150
200
250
300

Utility: $\mathbb{E}[f(w) - f^*] = \ ?$

assuming f and ∇f are Lipschitz

Utility: $\mathbb{E}[f(w) - f^*] = \ ?$

assuming f and ∇f are Lipschitz

For imbalanced objective/problems with sparse solutions:

Utility: $\mathbb{E}[f(w) - f^*] = \ ?$

assuming f and ∇f are Lipschitz

For imbalanced objective/problems with sparse solutions:

- Convex: $\widetilde{O}\left(\frac{\log p}{n\epsilon}\right)$
- Strongly-Convex: $\widetilde{O}\left(\frac{\log p}{n^2\epsilon^2}\right)$

When is the dependence logarithmic?

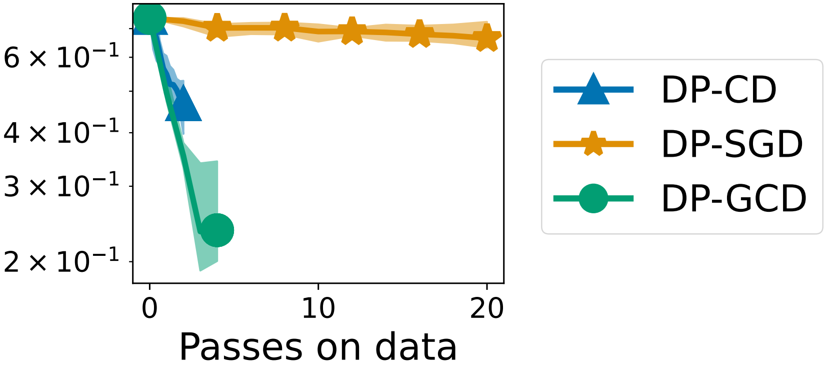
- Imbalanced problems:
  - $\|w^0 - w^*\|_{L,1} = \sum_{j=1}^{p} L_j^{1/2}|w_j^0 - w_j^*|$ is small
  - strong-convexity constant w.r.t. ℓ_1-norm is large

When is the dependence logarithmic?

- Imbalanced problems:
  - $\|w^0 - w^*\|_{L,1} = \sum_{j=1}^{p} L_j^{1/2} |w_j^0 - w_j^*|$ is small
  - strong-convexity constant w.r.t. ℓ_1-norm is large
- Sparse solutions (strongly-convex loss):
  - w^* has few non-zero coordinates
  - few total number of iterations/iterates remain sparse

Logistic Regression ($n = 1000, p = 100$)

$w^* \sim \text{lognormal}(\sigma = 2)^p$



Wrap up

- Private Greedy CD provably works!
- It can “bypass” ambient dimension
- In fact, GCD adapts to problems geometry

Thank you!

For more details, preprint online:

`https://arxiv.org/abs/2207.01560`