Privacy Amplification by Iteration for ADMM with (Strongly) Convex Objective Functions

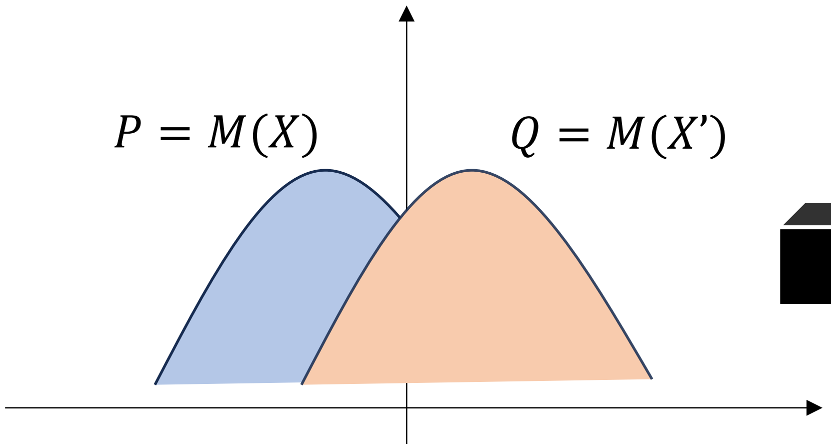
T-H. Hubert Chan, Hao Xie, Mengshi Zhao

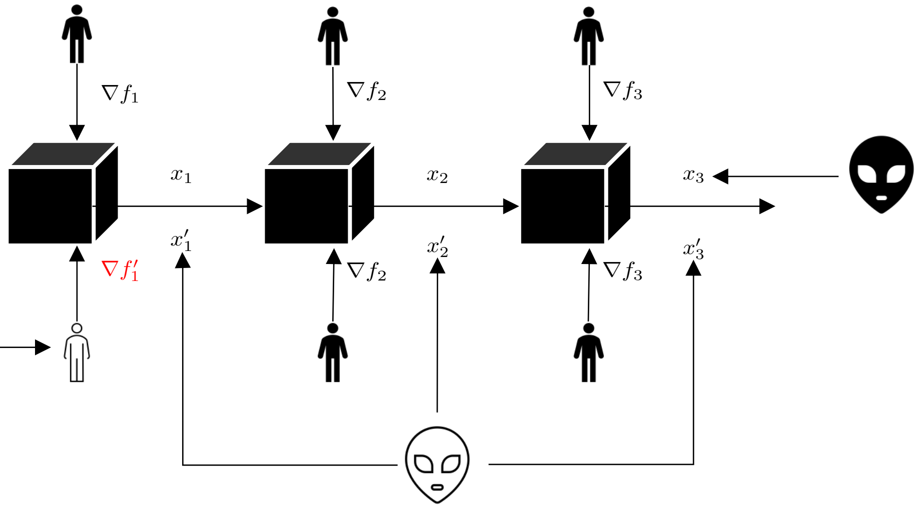
Department of Computer Science

The University of Hong Kong

Key Concepts

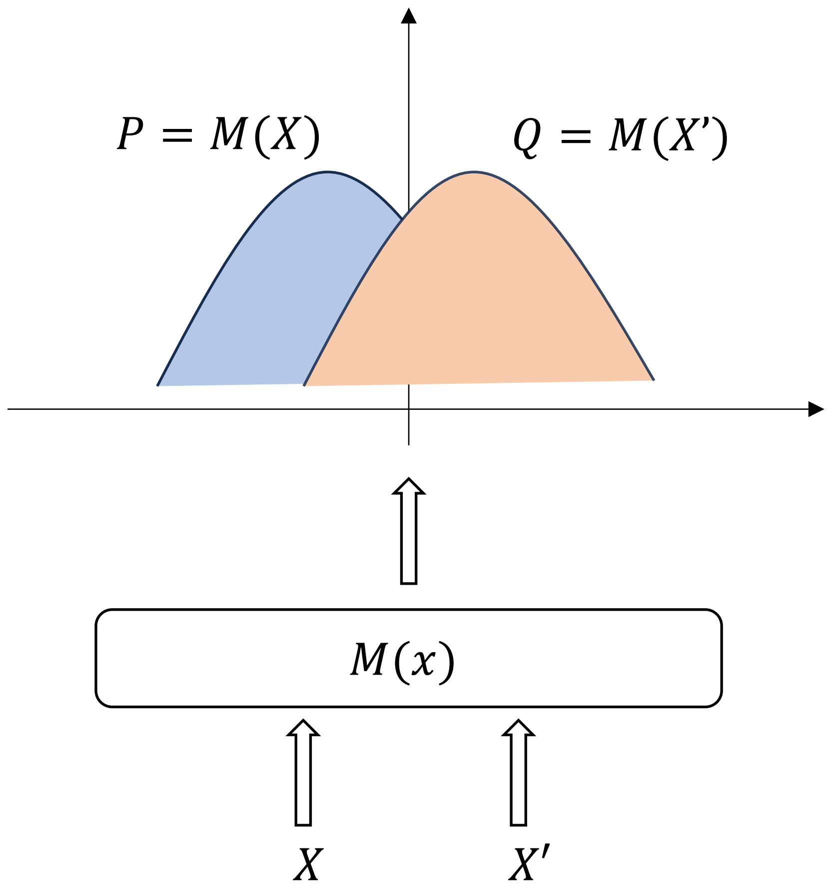
- Differential Privacy
- DP Amplification By Iteration
- ADMM





$$
\begin{aligned}
\min_{x,y} \quad & \mathfrak{f}(x) + g(y) \\
\text{s.t.} \quad & Ax + By = c \in \mathcal{R}^m \\
& x \in \mathcal{R}^n, y \in \mathcal{R}^\ell
\end{aligned}
$$

Renyi Differential Privacy



- The distribution of the output P on input X has nearly the similar distribution as the output Q on input X'.

- **Rényi divergence** of order $\alpha > 1$:

$$\mathrm{D}_\alpha(P\|Q) := \frac{1}{\alpha-1} \ln E_{x \leftarrow Q} \left(\frac{P(x)}{Q(x)} \right)^\alpha .$$

$$\mathrm{D}_\alpha(\mathcal{N}(x, \sigma^2 I_n) \| \mathcal{N}(x', \sigma^2 I_n)) = \frac{\|x - x'\|^2}{2\alpha\sigma^2} .$$

- Smaller Renyi-divergence means more similar.

Private Stochastic gradient descent

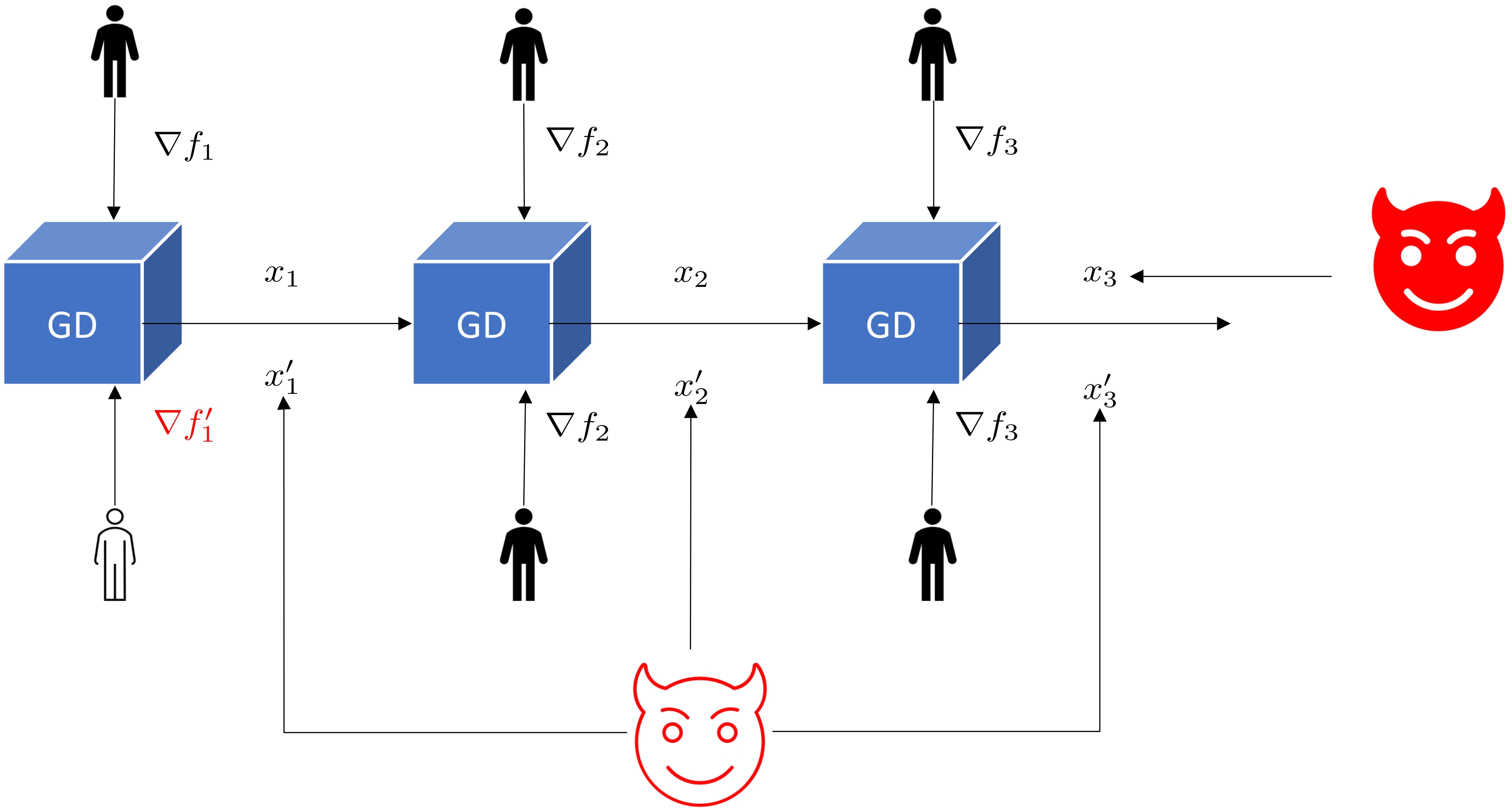
(Feldman, V.; Mironov, I.; Talwar, K.; and Thakurta, FOCS, 2018)

- Stochastic gradient descent:
  - A distribution $\mathcal{D}$. Objective function $\mathfrak{f} = E_{f \leftarrow \mathcal{D}}(f)$.
  - $\min \mathfrak{f}(x)$.
- T users, T samples in $\mathcal{D}$. $f_1, f_2, \dots, f_T$.
- For t=0,,2,...,T-1, do

$$x_{t+1} \leftarrow \Pi_K(x_t - \eta(\nabla f_{t+1}(x_t) + N(0, \sigma^2 I_d)))$$

- Two types of randomness:
  - Sampling: For convergence of the solution.
  - Gaussian noise: For achieving local DP.
- Two types of adversaries:
  - Observe all intermediate x_t. (Local DP)
  - Only obverse the final output x_T. (Privacy Amplification)
- Amplification: preserve local DP while the intermediate step can be guaranteed for the first user.

Privacy Amplification for Gradient Descent
(Feldman, V.; Mironov, I.; Talwar, K.; and Thakurta, FOCS, 2018)



Privacy Amplification for Gradient Descent
(Feldman, V.; Mironov, I.; Talwar, K.; and Thakurta, FOCS, 2018)

- Main theorem:
  - Two scenarios.
  - Same input X_0.
  - T iterations of noisy stochastic GD.
  - In the two scenarios, the first functions f_1 and f_1' are different. Other functions are the same.

  - Then

$$D_\alpha(X_T \| X_T') \leq O(\frac{1}{T}).$$

Generalization of GD

- Decomposing the Objective function: i: adopting different optimization ALG to each part. ii: for distributed learning.
- ADMM (Proximal version): Cyffers, E.; Bellet, A.; and Basu, D., ICML, 2023.
- ADMM (First-order approximation): our model.
- GD as a special case of ADMM: $g = 0, A = 0, B = 0, c = 0$.

$$
\begin{aligned}
\min_{x,y} \quad & \mathfrak{f}(x) + g(y) \\
\text{s.t.} \quad & Ax + By = c \in \mathcal{R}^m \\
& x \in \mathcal{R}^n, y \in \mathcal{R}^\ell
\end{aligned}
$$

Alternating direction method of multipliers

- Optimize over x:
  - Linear Approximation: $x_{t+1} = \arg\min \mathfrak{f}(x_t) + \langle \nabla \mathfrak{f}(x_t), x - x_t \rangle + g(y_t) + \frac{1}{2\eta}\|x - x_t\|^2 + \mathcal{L}(x, y_t, \lambda_t)$
  - Proximal ADMM: $x_{t+1} = \arg\min \mathfrak{f}(x) + g(y_t) + \mathcal{L}(x, y_t, \lambda_t)$
- Optimize over y: $y_{t+1} = \arg\min \mathfrak{f}(x_{t+1}) + g(y) + \mathcal{L}(x_{t+1}, y, \lambda_t)$
- Update λ: $\lambda_{t+1} = \lambda_t - \beta(Ax_{t+1} + By_{t+1} + c)$

$$\mathcal{L}(x, y, \lambda) = \frac{\beta}{2}\|Ax + By - c\|^2 - \langle \lambda, Ax + By - c \rangle$$

Noisy Stochastic ADMM

$$\begin{aligned} \min_{x,y} \quad & \mathfrak{f}(x) + g(y) \\ \text{s.t.} \quad & Ax + By = c \in \mathcal{R}^m \\ & x \in \mathcal{R}^n, y \in \mathcal{R}^\ell \end{aligned}$$

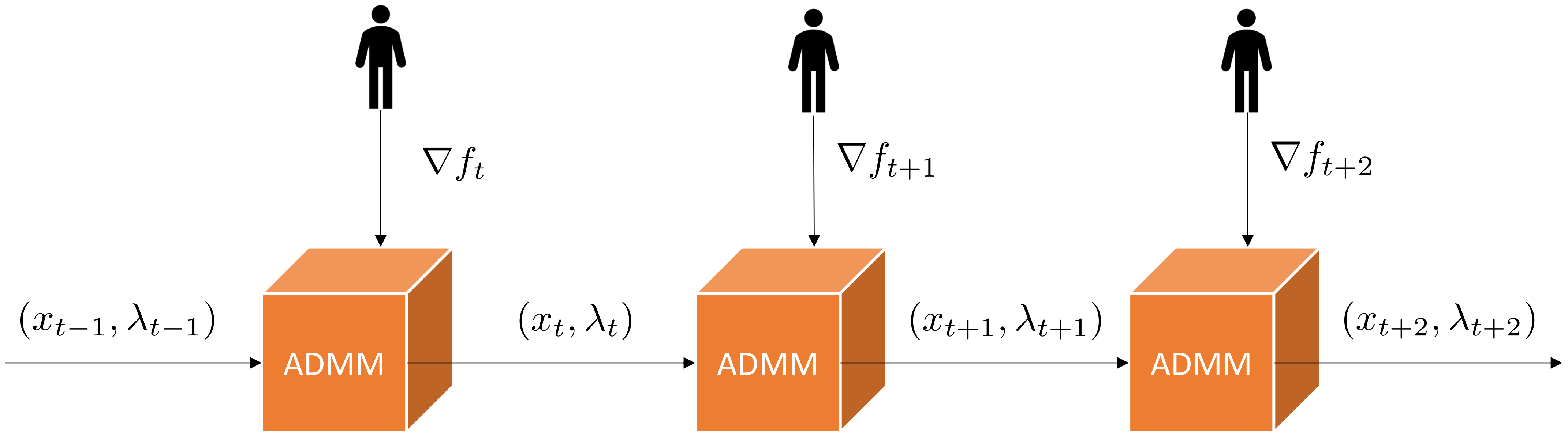
- Optimize over x: $\mathfrak{f} = E_{f \leftarrow \mathcal{D}}(f)$.

$$x_{t+1} = \arg\min f_{t+1}(x_t) + \langle \nabla f_{t+1}(x_t), x - x_t \rangle + g(y_t) + \mathcal{L}(x, y_t, \lambda_t)$$

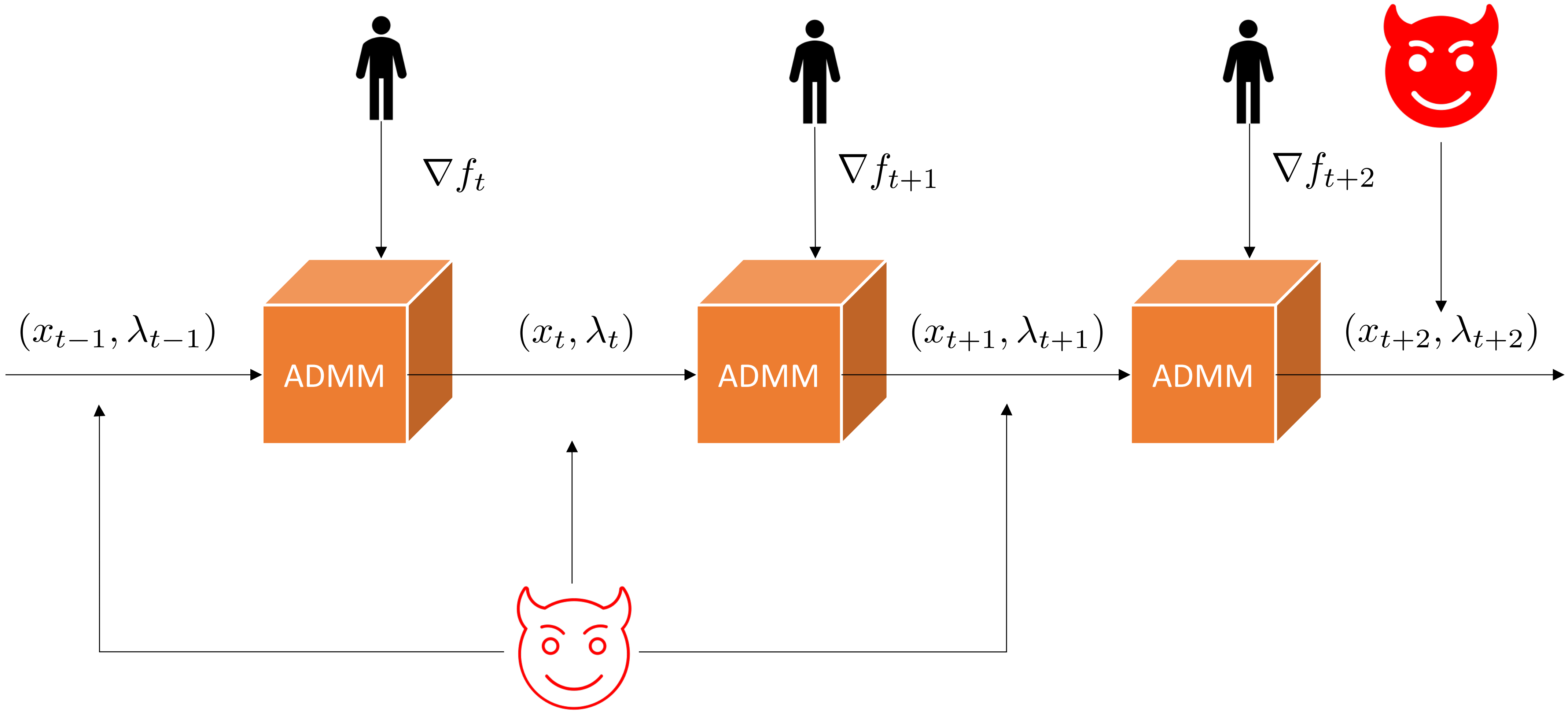
$$\tilde{x}_{t+1} = x_{t+1} + \mathcal{N}(0, \sigma^2 I)$$

- T users, T samples in $\mathcal{D}$. $f_1, f_2, \ldots, f_T$.

Stochastic ADMM



Stochastic ADMM



Privacy Amplification for Noisy Stochastic ADMM

- Main theorem:
  - Two scenarios.
  - Same input (X_0, λ_0).
  - T iterations of noisy stochastic ADMM.
  - In the two scenarios, the first functions f_1 and f_1' are different. Other functions are the same.

  - Then

$$\mathsf{D}_\alpha((\widetilde{x}_T, \lambda_T) \| (\widetilde{x}'_T, \lambda'_T)) \leq O(\tfrac{1}{T}).$$

Privacy Amplification By Coupling
(Balle, B.; Barthe, G.; and Gaboardi, M., NeurIPS, 2019)

- **In order to apply the method, those are required:**
  - Non-expansion for ADMM? One iteration without noise is non-expansion.

$$\|(x_{t+1} - x'_{t+1}, \lambda_{t+1} - \lambda'_{t+1})\|^2 \leq \|(x_t - x'_t, \lambda_t - \lambda'_t)\|^2.$$

  - One-step Privacy for ADMM? There is an upper bound for the divergence for noisy ADMM.

$$\mathrm{D}_\alpha \left((\widetilde{x}_{t+2}, \lambda_{t+2}) \middle\| (\widetilde{x}'_{t+2}, \lambda'_{t+2}) \right) \leq C_{\mathcal{K}} \left\| (\widetilde{x}_t - \widetilde{x}'_t, \lambda_t - \lambda'_t) \right\|^2 .$$

  - Naively doing this is impossible.

Challenge 1: Non-expansion for ADMM

- **Usual Norm may not be non-expansion:** one iteration signifies a transition in the (x, λ)-space.

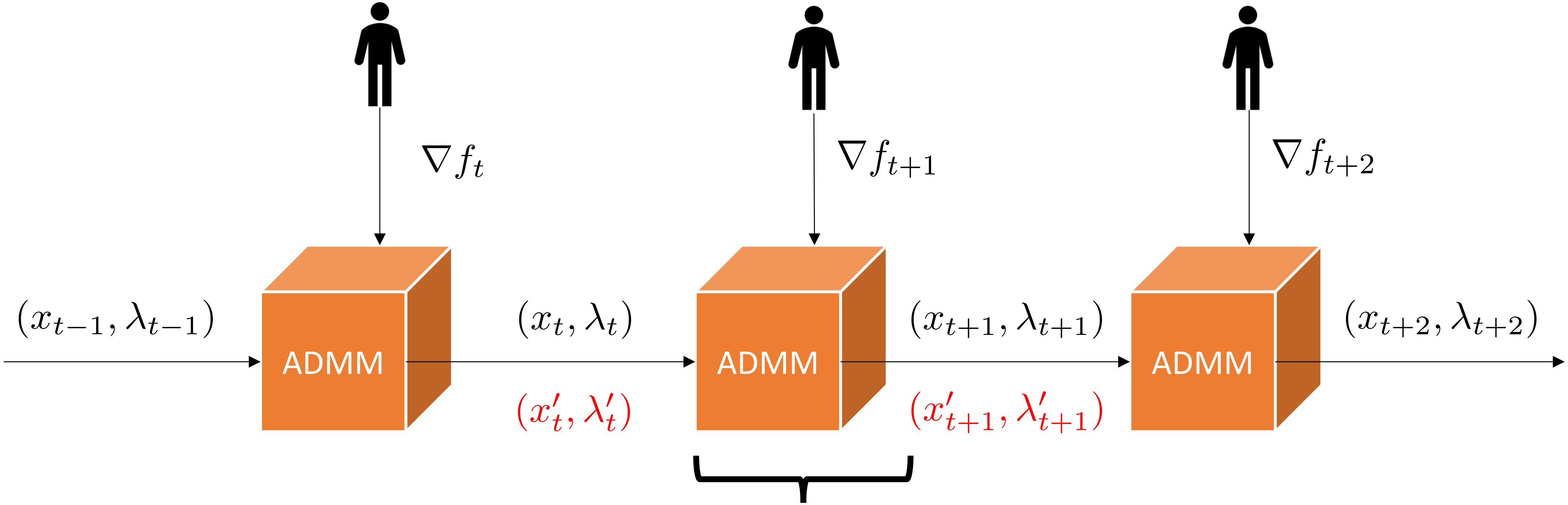
- **Customized Norm:**

$$\|(x,\lambda)\|_*^2 := \|x\|^2 + \frac{\eta}{\beta} \cdot \|\lambda - \beta A x\|^2.$$

- **Given two different input (x_t, λ_t) and (x'_t, λ'_t),**

$$\|(x_{t+1} - x'_{t+1}, \lambda_{t+1} - \lambda'_{t+1})\|_*^2 \leq \|(x_t - x'_t, \lambda_t - \lambda'_t)\|_*^2.$$

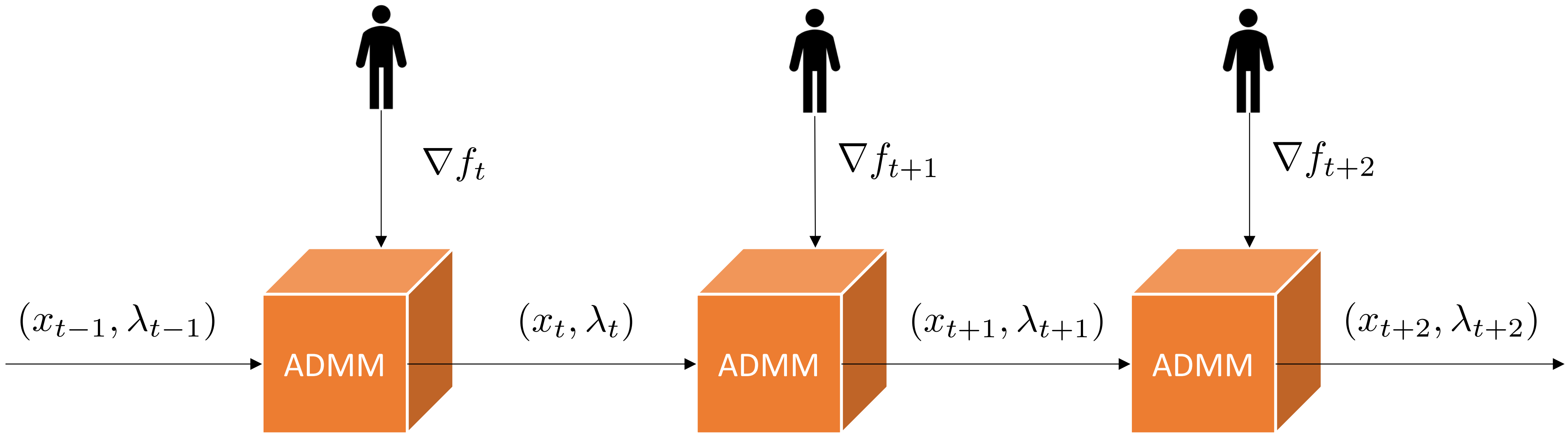
Challenge 2: One-step Privacy

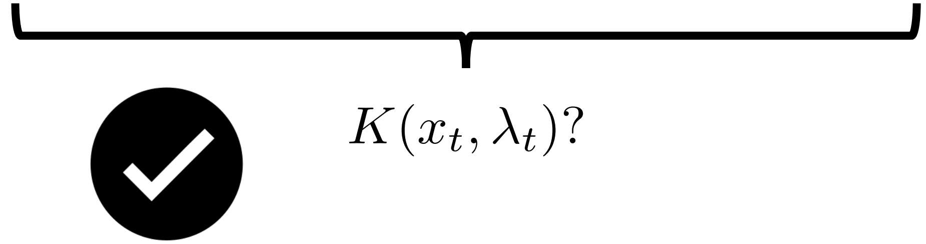


Noises added on x_t and λ_t are not independent.

$$\mathrm{D}_\alpha((x_{t-1}, \lambda_{t-1}) \| (x'_{t-1}, \lambda'_{t-1})) = +\infty$$

Consider two iterations together

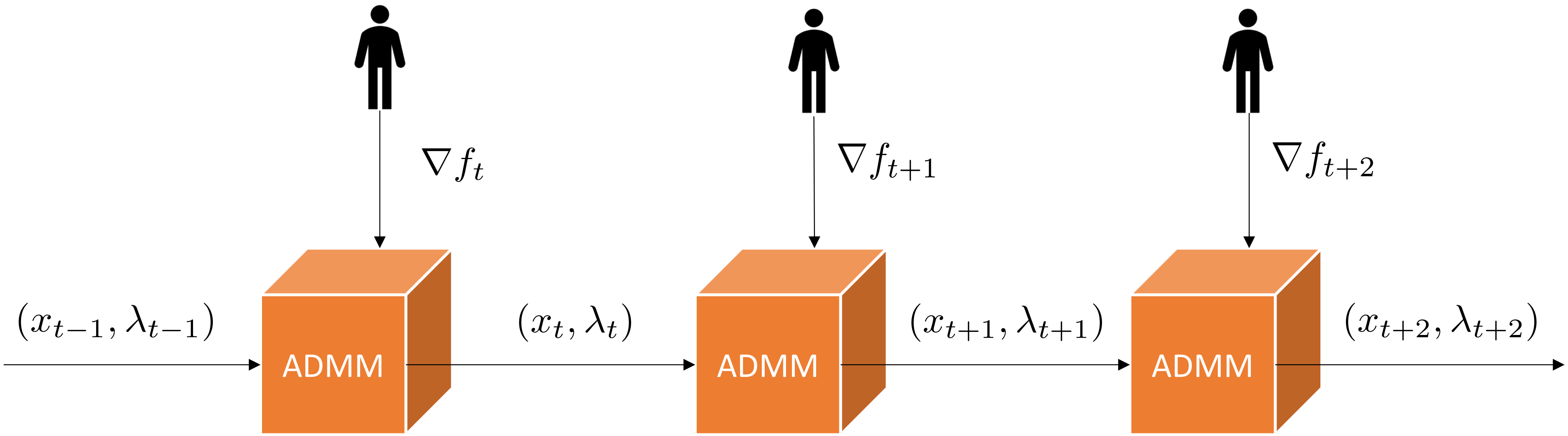




Produce (x_{t+1}, λ_{t+1}).
Produce masked $\tilde{x}_{t+1}$: $\widetilde{x}_{t+1} \leftarrow x_{t+1} + N_{t+1}$.

Produce (x_{t+2}, λ_{t+2}).
Produce masked $\tilde{x}_{t+2}$: $\widetilde{x}_{t+2} \leftarrow x_{t+2} + N_{t+2}$.
Return the pair.

Consider two iterations together



$$\mathsf{D}_\alpha\left((\widetilde{x}_{t+2}, \lambda_{t+2}) \middle\| (\widetilde{x}'_{t+2}, \lambda'_{t+2})\right) \leq C_{\mathcal{K}} \left\|(\widetilde{x}_t - \widetilde{x}'_t, \lambda_t - \lambda'_t)\right\|_*^2 .$$

Privacy Amplification for Stochastic ADMM

- Main theorem:
  - Two scenarios.
  - Same input (X_0, λ_0).
  - T iterations of noisy stochastic ADMM.
  - In the two scenarios, the first functions f_1 and f_1' are different. Other functions are the same.

  - Then

$$\mathrm{D}_\alpha((\widetilde{x}_T, \lambda_T) \| (\widetilde{x}'_T, \lambda'_T)) \leq O(\tfrac{1}{T}).$$

More Things In The Full Version

- **Strongly convex objective function:** the privacy amplification is improved to $\frac{L^T}{T}$, for some $0 < L < 1$.
- **Privacy for other users:** $O\left(\frac{log(T)}{T}\right)$-RDP.
  - Random permutation: the T users' data are processed in a uniformly random permutation.
  - Random stopping: The T users follow some deterministic arbitrary order. A random number N is sampled from {T+1/2, …, T} and only the first N users' data are used.
- **Privacy and Utility trade-off:** $O(\frac{1}{\sqrt{T}})$ convergence rate.
- **Experiments.**

Conclusion

- We have applied the coupling framework to achieve privacy amplification by iteration for ADMM.
- We have recovered the factor of $\frac{1}{T}$ in the Renyi divergence as the number T of iterations increases.
- We have performed experiments to evaluate the empirical performance of our methods in the full version.

Thanks!