A Simple Analysis for Exp-concave Empirical Minimization with Arbitrary Convex Regularizer

Tianbao Yang[*], Zhe Li[*], Lijun Zhang[♮]

[*]The University of Iowa, [♮]Nanjing University

April 9, 2018

AISTATS 2018

Outline

Outline

Problem

- Motivated by solving the stochastic composite optimization problem by Empirical Minimization:

$$\mathbf{w}_* = \arg\min_{\mathbf{w}\in\mathcal{W}} \left[P(\mathbf{w}) \triangleq \mathrm{E}_{\mathbf{z}\sim\mathcal{P}}[f(\mathbf{w},\mathbf{z})] + R(\mathbf{w})\right] \qquad (1)$$

Problem

- Motivated by solving the stochastic composite optimization problem by Empirical Minimization:

stochastic component

$$\mathbf{w}_* = \arg\min_{\mathbf{w}\in\mathcal{W}} \left[P(\mathbf{w}) \triangleq \mathrm{E}_{\mathbf{z}\sim\mathcal{P}}[f(\mathbf{w},\mathbf{z})] + R(\mathbf{w}) \right] \qquad (1)$$

Problem

- Motivated by solving the stochastic composite optimization problem by Empirical Minimization:

deterministic component

$$\mathbf{w}_* = \arg\min_{\mathbf{w}\in\mathcal{W}} \left[P(\mathbf{w}) \triangleq \mathrm{E}_{\mathbf{z}\sim\mathcal{P}}[f(\mathbf{w},\mathbf{z})] + R(\mathbf{w}) \right] \qquad (1)$$

Problem

- Motivated by solving the stochastic composite optimization problem by Empirical Minimization:

$$\mathbf{w}_* = \arg\min_{\mathbf{w}\in\mathcal{W}} \left[P(\mathbf{w}) \triangleq \mathrm{E}_{\mathbf{z}\sim\mathcal{P}}[f(\mathbf{w},\mathbf{z})] + R(\mathbf{w}) \right] \qquad (1)$$

  - $\mathcal{W}$ is compact and bounded convex set

Problem

- Motivated by solving the stochastic composite optimization problem by Empirical Minimization:

$$\mathbf{w}_* = \arg\min_{\mathbf{w}\in\mathcal{W}} \left[P(\mathbf{w}) \triangleq \mathrm{E}_{\mathbf{z}\sim\mathcal{P}}[f(\mathbf{w},\mathbf{z})] + R(\mathbf{w}) \right] \qquad (1)$$

  - $\mathcal{W}$ is compact and bounded convex set
  - $f(\mathbf{w},\mathbf{z})$: smooth and β-exp-concave function of $\mathbf{w}$ for any $\mathbf{z}$, Lipschitz continuous over $\mathcal{W}$.

Problem

- Motivated by solving the stochastic composite optimization problem by Empirical Minimization:

$$\mathbf{w}_* = \arg\min_{\mathbf{w}\in\mathcal{W}} \left[P(\mathbf{w}) \triangleq \mathrm{E}_{\mathbf{z}\sim\mathcal{P}}[f(\mathbf{w},\mathbf{z})] + R(\mathbf{w}) \right] \qquad (1)$$

  - $\mathcal{W}$ is compact and bounded convex set
  - $f(\mathbf{w},\mathbf{z})$: smooth and β-exp-concave function of $\mathbf{w}$ for any $\mathbf{z}$, Lipschitz continuous over $\mathcal{W}$.
  - No assumption on $R(\mathbf{w})$ except for convexity and boundness over $\mathcal{W}$.

Problem

- Study the convergence of the empirical minimizer of (1)

$$\widehat{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathcal{W}} \left[P_n(\mathbf{w}) \triangleq \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{w},\mathbf{z}_i) + R(\mathbf{w}) \right] \qquad (2)$$

  - where $\mathbf{z}_1,\ldots,\mathbf{z}_n$ are i.i.d samples from $\mathbb{P}$.

Problem

- Study the convergence of the empirical minimizer of (1)

$$\widehat{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathcal{W}} \left[P_n(\mathbf{w}) \triangleq \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{w}, \mathbf{z}_i) + R(\mathbf{w}) \right] \quad (2)$$

  - where $\mathbf{z}_1, \ldots, \mathbf{z}_n$ are i.i.d samples from $\mathbb{P}$.

- Goal: to establish the fast convergence rate of the empirical minimizer in terms of $P(\widehat{\mathbf{w}}) - P(\mathbf{w}_*)$.

Main Theorem:

$$\widehat{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathcal{W}} \left[P_n(\mathbf{w}) \triangleq \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{w}, \mathbf{z}_i) + R(\mathbf{w}) \right]$$

$P(\widehat{\mathbf{w}}) - \min_{\mathbf{w}\in\mathcal{W}} P(\mathbf{w}) \leq O(\frac{d\log n + d\log(1/\delta)}{n})$ with with high probability $1-\delta$, where $P(\mathbf{w}) \triangleq \mathrm{E}_{\mathbf{z}\sim\mathcal{P}}[f(\mathbf{w}, \mathbf{z})] + R(\mathbf{w})$

Main Results

Main Theorem:

$$\widehat{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathcal{W}} \left[P_n(\mathbf{w}) \triangleq \frac{1}{n}\sum_{i=1}^n f(\mathbf{w},\mathbf{z}_i) + R(\mathbf{w}) \right]$$

$P(\widehat{\mathbf{w}}) - \min_{\mathbf{w}\in\mathcal{W}} P(\mathbf{w}) \leq O(\frac{d\log n + d\log(1/\delta)}{n})$ with with high probability $1-\delta$, where $P(\mathbf{w}) \triangleq \mathrm{E}_{\mathbf{z}\sim\mathcal{P}}[f(\mathbf{w},\mathbf{z})] + R(\mathbf{w})$

Corollary:

$F(\widetilde{\mathbf{w}}) - \min_{\mathbf{w}\in\mathcal{W}} F(\mathbf{w}) \leq O(\frac{d\log n + d\log(1/\delta)}{n})$ with with high probability $1-\delta$, where $F(\mathbf{w}) \triangleq \mathrm{E}_{\mathbf{z}\sim\mathcal{P}}[f(\mathbf{w},\mathbf{z})]$

$$\widetilde{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathcal{W}} \left[\widehat{F}_n(\mathbf{w}) \triangleq \frac{1}{n}\sum_{i=1}^n f(\mathbf{w},\mathbf{z}_i) + \frac{1}{n} g(\mathbf{w}) \right]$$

Outline

Comparison with Related Works

- The three recent studies [1, 2, 3] focus on establishing fast rates in terms of risk minimization without a regularizer

$$\min_{\mathbf{w}\in\mathcal{W}} F(\mathbf{w}) \triangleq \mathrm{E}_{\mathbf{z}}[f(\mathbf{w},\mathbf{z})] \tag{3}$$

Comparison with Related Works

- The three recent studies [1, 2, 3] focus on establishing fast rates in terms of risk minimization without a regularizer

$$\min_{\mathbf{w}\in\mathcal{W}} F(\mathbf{w}) \triangleq \mathrm{E}_{\mathbf{z}}[f(\mathbf{w},\mathbf{z})] \tag{3}$$

  - Koren & Levy [1] studied the convergence of a regularized empirical risk minimizer by

$$\widetilde{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathcal{W}} \left[\widehat{F}_n(\mathbf{w}) \triangleq \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{w},\mathbf{z}_i) + \frac{1}{n} g(\mathbf{w})\right] \tag{4}$$

Comparison with Related Works

- The three recent studies [1, 2, 3] focus on establishing fast rates in terms of risk minimization without a regularizer

$$\min_{\mathbf{w}\in\mathcal{W}} F(\mathbf{w}) \triangleq \mathrm{E}_{\mathbf{z}}[f(\mathbf{w},\mathbf{z})] \tag{3}$$

  - Koren & Levy [1] studied the convergence of a regularized empirical risk minimizer by

strongly convex

$$\widetilde{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathcal{W}} \left[\widehat{F}_n(\mathbf{w}) \triangleq \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{w},\mathbf{z}_i) + \frac{1}{n} g(\mathbf{w})\right] \tag{4}$$

- The three recent studies [1, 2, 3] focus on establishing fast rates in terms of risk minimization without a regularizer

$$\min_{\mathbf{w}\in\mathcal{W}} F(\mathbf{w}) \triangleq \mathrm{E}_{\mathbf{z}}[f(\mathbf{w},\mathbf{z})] \tag{3}$$

  - Koren & Levy [1] studied the convergence of a regularized empirical risk minimizer by

$$\widetilde{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathcal{W}} \left[\widehat{F}_n(\mathbf{w}) \triangleq \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{w},\mathbf{z}_i) + \frac{1}{n} g(\mathbf{w})\right] \tag{4}$$

  - Mehta [2] targeted on the original risk mimization as (3)

Comparison with Related Works

- The three recent studies [1, 2, 3] focus on establishing fast rates in terms of risk minimization without a regularizer

$$\min_{\mathbf{w}\in\mathcal{W}} F(\mathbf{w}) \triangleq \mathrm{E}_{\mathbf{z}}[f(\mathbf{w},\mathbf{z})] \tag{3}$$

  - Koren & Levy [1] studied the convergence of a regularized empirical risk minimizer by

$$\widetilde{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathcal{W}} \left[\widehat{F}_n(\mathbf{w}) \triangleq \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{w},\mathbf{z}_i) + \frac{1}{n} g(\mathbf{w})\right] \tag{4}$$

  - Mehta [2] targeted on the original risk mimization as (3)
  - Gonen & Shalev-Shawartz [3] focused on the risk minimization with generalized linear model:

$$\min_{\mathbf{w}\in\mathcal{W}} L(\mathbf{w}) \triangleq \mathrm{E}_{(\mathbf{x},y)\sim D}[\phi_y(\mathbf{w}^\top\mathbf{x})], \tag{5}$$

Comparison with Related Works

Difference of fast rates between our work and the related works [1, 2, 3]

Related Work	**Ours**
$\min_{\mathbf{w}\in\mathcal{W}} F(\mathbf{w})$ [1, 2, 3]	$\min_{\mathbf{w}\in\mathcal{W}} F(\mathbf{w}) + R(\mathbf{w})$
*s-convex regularizer [1]	convex regularizer
Expectation [1, 3]	High probability [2]

*s-convex: strongly convex

- Our result is more general.

[1]. T. Koren and K. Y. Levy. Fast rates for exp- concave empirical risk minimization.

[2]. N. A. Mehta. Fast rate with high probability in exp-concave statistical learning.

[3]. A. Gonen and S. Shalev-Shwartz. Average stability is invariant to data preconditioning. implications to exp-concave empirical risk minimization.

Outline

Assumption 1

- $\mathcal{W}$ is a closed and bounded convex set, i.e., there exists R such that $\|\mathbf{w}\|_2 \leq R$ for all $\mathbf{w} \in \mathcal{W}$.

Assumption 1

- $\mathcal{W}$ is a closed and bounded convex set, i.e., there exists R such that $\|\mathbf{w}\|_2 \leq R$ for all $\mathbf{w} \in \mathcal{W}$.
- $f(\mathbf{w}, \mathbf{z})$ is a G-Lipschitz continuous, L-smooth and β-exp-concave function of $\mathbf{w} \in \mathcal{W}$ for any $\mathbf{z} \in \mathcal{Z}$.

Assumption 1

- $\mathcal{W}$ is a closed and bounded convex set, i.e., there exists R such that $\|\mathbf{w}\|_2 \leq R$ for all $\mathbf{w} \in \mathcal{W}$.
- $f(\mathbf{w}, \mathbf{z})$ is a G-Lipschitz continuous, L-smooth and β-exp-concave function of $\mathbf{w} \in \mathcal{W}$ for any $\mathbf{z} \in \mathcal{Z}$.
- $R(\mathbf{w})$ is a convex function.

Theoretical Results

Recall: $P(\mathbf{w}) \triangleq \mathrm{E}_{\mathbf{z}\sim\mathcal{P}}[f(\mathbf{w},\mathbf{z})] + R(\mathbf{w})$

Theorem 1

For the stochastic composite minimization problem (1), we consider the empirical minimizer $\widehat{\mathbf{w}}$ by solving (2). Under Assumption 1, with probability at least $1-\delta$, we have

$$P(\widehat{\mathbf{w}}) - P(\mathbf{w}_*) \leq O\left(\frac{d\log n}{n} + \frac{d\log(1/\delta)}{n\sigma}\right).$$

Theoretical Results

Recall: $P(\mathbf{w}) \triangleq \mathrm{E}_{\mathbf{z}\sim\mathcal{P}}[f(\mathbf{w},\mathbf{z})] + R(\mathbf{w})$

Theorem 1

For the stochastic composite minimization problem (1), we consider the empirical minimizer $\widehat{\mathbf{w}}$ by solving (2). Under Assumption 1, with probability at least $1-\delta$, we have

$$P(\widehat{\mathbf{w}}) - P(\mathbf{w}_*) \leq O\left(\frac{d\log n}{n} + \frac{d\log(1/\delta)}{n\sigma}\right).$$

Remarks:

- When $R(\mathbf{w}) = 0$, directly obtain a fast rate with high probability of the empirical risk minimizer for the exp-concave risk minimization.
- Linear dependence on dimensionality d is unavoidable [4].

[4]. V. Feldman. Generalization of ERM in stochastic convex optimization: The dimension strikes back.

Theoretical Results

Recall: $F(\mathbf{w}) \triangleq \mathrm{E}_{\mathbf{z}\sim\mathcal{P}}[f(\mathbf{w},\mathbf{z})]$

$$\widetilde{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathcal{W}} \left[\widehat{F}_n(\mathbf{w}) \triangleq \frac{1}{n}\sum_{i=1}^n f(\mathbf{w},\mathbf{z}_i) + \frac{1}{n}g(\mathbf{w})\right]$$

Theoretical Results

Recall: $F(\mathbf{w}) \triangleq \mathrm{E}_{\mathbf{z}\sim\mathcal{P}}[f(\mathbf{w},\mathbf{z})]$

$$\widetilde{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathcal{W}} \left[\widehat{F}_n(\mathbf{w}) \triangleq \frac{1}{n}\sum_{i=1}^n f(\mathbf{w},\mathbf{z}_i) + \frac{1}{n}g(\mathbf{w})\right]$$

Theorem 2

Under Assumption 1 (i), (ii), and that $g(\mathbf{x})$ is bounded over $\mathcal{W}$ such that $\sup_{\mathbf{w},\mathbf{w}'\in\mathcal{W}} |g(\mathbf{w}) - g(\mathbf{w}')| \leq B$, with probability at least $1-\delta$, we have

$$F(\widetilde{\mathbf{w}}) - \min_{\mathbf{w}\in\mathcal{W}} F(\mathbf{w}) \leq O\left(\frac{d\log n}{n} + \frac{d\log(1/\delta)}{n\sigma}\right).$$

Remarks:

Recall: $F(\mathbf{w}) \triangleq \mathrm{E}_{\mathbf{z}\sim\mathcal{P}}[f(\mathbf{w},\mathbf{z})]$

$$\widetilde{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathcal{W}} \left[\widehat{F}_n(\mathbf{w}) \triangleq \frac{1}{n}\sum_{i=1}^n f(\mathbf{w},\mathbf{z}_i) + \frac{1}{n}g(\mathbf{w})\right]$$

Theorem 2

Under Assumption 1 (i), (ii), and that $g(\mathbf{x})$ is bounded over $\mathcal{W}$ such that $\sup_{\mathbf{w},\mathbf{w}'\in\mathcal{W}} |g(\mathbf{w}) - g(\mathbf{w}')| \leq B$, with probability at least $1-\delta$, we have

$$F(\widetilde{\mathbf{w}}) - \min_{\mathbf{w}\in\mathcal{W}} F(\mathbf{w}) \leq O\left(\frac{d\log n}{n} + \frac{d\log(1/\delta)}{n\sigma}\right).$$

Remarks:

- Address the open problem raised in [1] about high probability bound for strongly regularized empirical risk minimizer.

Theoretical Results

Recall: $F(\mathbf{w}) \triangleq \mathrm{E}_{\mathbf{z}\sim\mathcal{P}}[f(\mathbf{w},\mathbf{z})]$

$$\widetilde{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathcal{W}} \left[\widehat{F}_n(\mathbf{w}) \triangleq \frac{1}{n}\sum_{i=1}^n f(\mathbf{w},\mathbf{z}_i) + \frac{1}{n}g(\mathbf{w})\right]$$

Theorem 2

Under Assumption 1 (i), (ii), and that $g(\mathbf{x})$ is bounded over $\mathcal{W}$ such that $\sup_{\mathbf{w},\mathbf{w}'\in\mathcal{W}} |g(\mathbf{w}) - g(\mathbf{w}')| \leq B$, with probability at least $1-\delta$, we have

$$F(\widetilde{\mathbf{w}}) - \min_{\mathbf{w}\in\mathcal{W}} F(\mathbf{w}) \leq O\left(\frac{d\log n}{n} + \frac{d\log(1/\delta)}{n\sigma}\right).$$

Remarks:

- Address the open problem raised in [1] about high probability bound for strongly regularized empirical risk minimizer.
- Extend the fast rate to any regularized empirical risk minimizer as long as the regularizer is convex.

Outline

Analysis Technique

- Step 1: by using the convexity of $P(\mathbf{w})$, the optimality condition of $\widehat{\mathbf{w}}$, and Cauchy-Schwarz inequality:

$$
\begin{aligned}
&P(\widehat{\mathbf{w}}) - P(\mathbf{w}_*) \\
&\leq \left\|G(\widehat{\mathbf{w}}, \mathbf{w}_*) - G_n(\widehat{\mathbf{w}}, \mathbf{w}_*)\right\|_2 \left\|\widehat{\mathbf{w}} - \mathbf{w}_*\right\|_2 \\
&+ \left\|\Delta_n(\mathbf{w}_*)\right\|_{H^{-1}} \left\|\widehat{\mathbf{w}} - \mathbf{w}_*\right\|_H
\end{aligned}
$$

where

$$
\begin{aligned}
&G(\mathbf{w}, \mathbf{w}_*) = \nabla P(\mathbf{w}) - \nabla P(\mathbf{w}_*), \\
&G_n(\mathbf{w}, \mathbf{w}_*) = \nabla P_n(\mathbf{w}) - \nabla P_n(\mathbf{w}_*), \\
&\Delta_n(\mathbf{w}) = \nabla P(\mathbf{w}) - \nabla P_n(\mathbf{w}), \\
&H = I + \frac{\sigma}{\alpha}\mathrm{E}[\nabla f(\mathbf{w}^*, \mathbf{z})\nabla f(\mathbf{w}^*, \mathbf{z})^\top],
\end{aligned}
$$

σ is parameter related to β-exp-concave and α is artificial parameter introduced.

Analysis Technique

- Step 1: by using the convexity of $P(\mathbf{w})$, the optimality condition of $\widehat{\mathbf{w}}$, and Cauchy-Schwarz inequality:

$$
\begin{aligned}
& P(\widehat{\mathbf{w}}) - P(\mathbf{w}_*) \\
& \leq \|G(\widehat{\mathbf{w}}, \mathbf{w}_*) - G_n(\widehat{\mathbf{w}}, \mathbf{w}_*)\|_2 \|\widehat{\mathbf{w}} - \mathbf{w}_*\|_2 \\
& + \|\Delta_n(\mathbf{w}_*)\|_{H^{-1}} \|\widehat{\mathbf{w}} - \mathbf{w}_*\|_H
\end{aligned}
$$

Local norm

where

$$
\begin{aligned}
& G(\mathbf{w}, \mathbf{w}_*) = \nabla P(\mathbf{w}) - \nabla P(\mathbf{w}_*), \\
& G_n(\mathbf{w}, \mathbf{w}_*) = \nabla P_n(\mathbf{w}) - \nabla P_n(\mathbf{w}_*), \\
& \Delta_n(\mathbf{w}) = \nabla P(\mathbf{w}) - \nabla P_n(\mathbf{w}), \\
& H = I + \tfrac{\sigma}{\alpha}\mathrm{E}[\nabla f(\mathbf{w}^*, \mathbf{z})\nabla f(\mathbf{w}^*, \mathbf{z})^\top],
\end{aligned}
$$

σ is parameter related to β-exp-concave and α is artificial parameter introduced.

- Step 2: By using concentration inequality [5], union bound and covering number of $\mathcal{W}$, with probability at least $1-\delta$,

$$\|G(\widehat{\mathbf{w}}, \mathbf{w}_*) - G_n(\widehat{\mathbf{w}}, \mathbf{w}_*)\|_2 \leq O(\frac{d \log n}{n}) + O(\sqrt{\frac{d(P(\widehat{\mathbf{w}}) - P(\mathbf{w}_*))}{n}})$$

- Step 2: By using concentration inequality [5], union bound and covering number of $\mathcal{W}$, with probability at least $1-\delta$,

Variance

$$\|G(\widehat{\mathbf{w}}, \mathbf{w}_*) - G_n(\widehat{\mathbf{w}}, \mathbf{w}_*)\|_2 \leq O\left(\frac{d\log n}{n}\right) + O\left(\sqrt{\frac{d(P(\widehat{\mathbf{w}}) - P(\mathbf{w}_*))}{n}}\right)$$

- Step 2: By using concentration inequality [5], union bound and covering number of $\mathcal{W}$, with probability at least $1-\delta$,

$$\|G(\widehat{\mathbf{w}}, \mathbf{w}_*) - G_n(\widehat{\mathbf{w}}, \mathbf{w}_*)\|_2 \leq O(\frac{d \log n}{n}) + O(\sqrt{\frac{d(P(\widehat{\mathbf{w}}) - P(\mathbf{w}_*))}{n}})$$

$$\|\Delta_n(\mathbf{w}_*)\|_{H^{-1}} \leq \frac{2G\log(2/\delta)}{n} + \sqrt{\frac{2\alpha d \log(2/\delta)}{n\sigma}}$$

- Step 2: By using concentration inequality [5], union bound and covering number of $\mathcal{W}$, with probability at least $1-\delta$,

$$\|G(\widehat{\mathbf{w}},\mathbf{w}_*) - G_n(\widehat{\mathbf{w}},\mathbf{w}_*)\|_2 \leq O(\frac{d\log n}{n}) + O(\sqrt{\frac{d(P(\widehat{\mathbf{w}}) - P(\mathbf{w}_*))}{n}})$$

$$\|\Delta_n(\mathbf{w}_*)\|_{H^{-1}} \leq \frac{2G\log(2/\delta)}{n} + \sqrt{\frac{2\alpha d\log(2/\delta)}{n\sigma}}$$

[5]. S. Smale and D. X. Zhou. Learning theory estimates via integral operators and their approximations.

Analysis Technique

- Step 3: using Young's inequality and do some linear algebra:

$$P(\widehat{\mathbf{w}}) - P(\mathbf{w}_*) \leq O\left(\frac{d \log n}{n} + \frac{d \log(1/\delta)}{n\sigma}\right).$$

with probability at least $1 - \delta$.

Outline

Conclusion

- Developed a simple analysis of fast rates for empirical minimization with exp-concave loss functions and an arbitrary convex regularizer.

Conclusion

- Developed a simple analysis of fast rates for empirical minimization with exp-concave loss functions and an arbitrary convex regularizer.
- Exploited the covering number of a finite-dimensional bounded set and a concentration inequality of random vectors.

Conclusion

- Developed a simple analysis of fast rates for empirical minimization with exp-concave loss functions and an arbitrary convex regularizer.
- Exploited the covering number of a finite-dimensional bounded set and a concentration inequality of random vectors.
- Induced an unified fast rate results for exp-concave empirical risk minimization without and with any convex regularizer.