Paper :

Toulis, P., & Airoldi, E. M. (2017). Asymptotic and finite-sample properties of estimators based on stochastic gradients. *The Annals of Statistics*, *45*(4), 1694-1727.

Theorem2.3

$$\theta_n^{im} = \theta_{n-1}^{im} + \gamma_n \nabla \log f\left(Y_n; X_n, \theta_n^{im}\right),$$

$$\overline{\theta_n^{im}} = \frac{1}{n}\sum_{i=1}^{n} \theta_i^{im},$$

where $\gamma_n = \gamma_1 n^{-\gamma}, \gamma \in [0.5,1)$. Then, $\overline{\theta_n^{im}}$ converges to true parameters $\theta_\star$ in probability and is asymptotically efficient, i.e.,

$$n\mathrm{Var}(\overline{\theta_n^{im}}) \rightarrow \mathcal{L}(\theta_\star)^{-1},$$

where $\mathcal{L}(\theta_\star) = E(\hat{\mathcal{L}}_n(\theta_\star)) = E(-\nabla^2 \log f(Y_n; X_n, \theta_n)\ \ )$, a fisher information matrix.

In simulation, set linear model with $Y_i = X_i^T \theta_\star + \varepsilon_i$, where $\theta_\star \in R^5, X_i \sim N(5,1)$, $\varepsilon_i \sim N(0,1), i = 1, \dots, n$. Then θ_n^{im} function will be

$$\theta_n^{im} = \theta_{n-1}^{im} + \gamma_n X_n\left(Y_n - X_n^T \theta_n^{im}\right)$$

And it can be rewrite to the equation below:

$$\theta_n^{im} = (I + \gamma_n X_n X_n^T)^{-1}\theta_{n-1}^{im} + (I + \gamma_n X_n X_n^T)^{-1}\gamma_n X_n Y_n$$

Let n = 10000, and do the implicit SGD iteration procedure 100 times to get each $\overline{\theta_n^{im}}$. Then, construct $n\mathrm{Var}(\overline{\theta_n^{im}})$ to compare with $\mathcal{L}(\theta_\star)^{-1} = E(XX^T)^{-1}$.