

新能源与储能容量配置优化研究

报告人: 杜洪博

导 师: 寇彩霞

北京邮电大学 理学院

2023 年 12 月 2 日

- ① 问题背景与研究意义
- ② 研究现状
- ③ Algorithm and Convergence analysis
- ④ Numerical Experiments
- ⑤ Conclusion

- ① 问题背景与研究意义
- ② 研究现状
- ③ Algorithm and Convergence analysis
- ④ Numerical Experiments
- ⑤ Conclusion

问题背景与研究意义

问题背景与研究意义

- ① 问题背景与研究意义
- ② 研究现状
- ③ Algorithm and Convergence analysis
- ④ Numerical Experiments
- ⑤ Conclusion

研究现状

Let

$$\bar{X} = \nabla f_{S_k}(\omega^k), \bar{Y} = \nabla f_{S_k}(\phi_i^k), \text{ or } \bar{Y} = \nabla f_{S_k}(\psi_i^k)$$

(SAGA/SVRG)

$$g_k = \bar{X} - \bar{Y} + E(\bar{Y})$$

$$E(g_k) = E(\bar{X} - \bar{Y} + E(\bar{Y})) = E(\bar{X})$$

$$\text{Var}(g_k) = \text{Var}(\bar{X} - \bar{Y} + E(\bar{Y})) = \text{Var}(\bar{X} - \bar{Y})$$

(SAG)

$$g_k = \frac{\bar{X} - \bar{Y}}{n} + E(\bar{Y})$$

$$E(g_k) = E\left(\frac{\bar{X} - \bar{Y}}{n} + E(\bar{Y})\right) = \frac{1}{n}E(\bar{X}) + \left(1 - \frac{1}{n}\right)E(\bar{Y})$$

$$\text{Var}(g_k) = \text{Var}\left(\frac{\bar{X} - \bar{Y}}{n} + E(\bar{Y})\right) = \frac{1}{n^2} \text{Var}(\bar{X} - \bar{Y})$$

Gradient Estimate with Unbiasedness and Minimal Variance

Estimator:

$$\theta_\gamma = \bar{X} - \gamma(\bar{Y} - E(\bar{Y}))$$

Expected Value:

$$E(\theta_\gamma) = E(\bar{X})$$

Variance:

$$\begin{aligned} Var(\theta_\gamma) &= Var(\bar{X} - \gamma(\bar{Y} - E(\bar{Y}))) \\ &= Var(\bar{X}) + \gamma^2 Var(\bar{Y}) - 2\gamma Cov(\bar{X}, \bar{Y}) \end{aligned} \quad (1)$$

Best Param:

$$\gamma^* = \frac{Cov(\bar{X}, \bar{Y})}{Var(\bar{Y})} \approx \frac{s_{XY}}{s_Y^2} \quad (2)$$

$$\begin{aligned} s_{XY} &= \frac{1}{|S|-1} \sum_{j \in S_k} (X_j - \bar{X})(Y_j - \bar{Y}) \\ s_Y^2 &= \frac{1}{|S|-1} \sum (Y_j - \bar{Y})^2 \end{aligned} \quad (3)$$

Estimation of γ^*

$$\begin{aligned}
 Cov(\bar{X}, \bar{Y}) &= Cov\left(\frac{1}{|S|} \sum_{j \in S_k} X_j, \frac{1}{|S|} \sum_{j \in S_k} Y_j\right) = \frac{1}{|S|^2} Cov\left(\sum_{j \in S_k} X_j, \sum_{j \in S_k} Y_j\right) \\
 &= \frac{1}{|S|^2} \sum_{j \in S_k} Cov(X_j, Y_j) = \frac{1}{|S|} Cov(X, Y) \approx \frac{1}{|S|} s_{XY}
 \end{aligned} \tag{4}$$

$$\begin{aligned}
 Var(\bar{Y}) &= Var\left(\frac{1}{|S|} \sum_{j \in S_k} Y_j\right) = \frac{1}{|S|^2} Var\left(\sum_{j \in S_k} Y_j\right) \\
 &= \frac{1}{|S|^2} \sum_{j \in S_k} Var(Y_j) = \frac{1}{|S|} Var(Y) \approx \frac{1}{|S|} s_Y^2
 \end{aligned} \tag{5}$$

$$\begin{aligned}
 s_{XY} &= \frac{1}{|S| - 1} \sum_{j \in S_k} (X_j - \bar{X})(Y_j - \bar{Y}) \\
 s_Y^2 &= \frac{1}{|S| - 1} \sum_{j \in S_k} (Y_j - \bar{Y})^2
 \end{aligned} \tag{6}$$

The Stochastic Conjugate Gradient Algorithm with Variance Reduction

CGVR(Ran Xin, Usman A. Khan, and Soummya Kar, 2020)

$$g_k = \nabla f_{S_k}(\omega^k) - \nabla f_{S_k}(\psi_j^k) + \frac{1}{n} \sum_{i=1}^n \nabla f_i(\psi_i^k)$$

$$d_k = -g_k + \beta_k d_{k-1}$$

SCGA(Caixia Kou and Han Yang, 2022)

$$g_k = \nabla f_{S_k}(\omega^k) - \nabla f_{S_k}(\phi_j^k) + \frac{1}{n} \sum_{i=1}^n \nabla f_i(\phi_i^k)$$

$$d_k = -g_k + \beta_k d_{k-1}$$

- ① 问题背景与研究意义
- ② 研究现状
- ③ Algorithm and Convergence analysis
- ④ Numerical Experiments
- ⑤ Conclusion

Algorithm 1: SCGA with the minimal variance stochastic gradient estimate

```

1 Initialization: .....
2 Iteration:
3 for  $k = 1, 2 \dots$  do
4     .....
5     for  $j : S_k$  do
6         Compute  $\nabla f_j(\omega_k)$  and store it into matrix  $\nabla f_{[S_k]}(\omega_k)$ 
7         Select  $\nabla f_j(\omega_{[k-1]})$  in  $\nabla f(\omega_{[k-1]})$  and store it into matrix
             $\nabla f_{[S_k]}(\omega_{[k-1]})$ 
8     for  $r = 1, 2 \dots d$  do
9         Compute the sample covariance of  $\nabla f_{[S_k]}^{(r)}(\omega_k)$  and
             $\nabla f_{[S_k]}^{(r)}(\omega_{[k-1]})$ , the sample variance of  $\nabla f_{[S_k]}^{(r)}(\omega_{[k-1]})$  using
            (6)
10        Compute  $\gamma^{*(r)}$  using (2)
11    Compute

```

Algorithm 2: CGVR with the minimal variance stochastic gradient estimate

```

1 Initialization: Given  $x_0 \in R^d$ , compute  $h_0 = \frac{1}{n} \sum_{i=1}^n \nabla f_i(x_0)$ :
2 Iteration: for  $\underline{l = 1, 2 \dots T}$  do
3     .....
4     for  $\underline{k = 1, 2 \dots m}$  do
5         .....
6         for  $\underline{j: S_k}$  do
7             Compute  $\nabla f_j(\omega_k), \nabla f_j(\omega_0)$ 
8             Store  $\nabla f_{[S_k]}(\omega_k) \leftarrow \nabla f_j(\omega_k), \nabla f_{[S_k]}(\omega_0) \leftarrow \nabla f_j(\omega_0)$ 
9         for  $\underline{r = 1, 2 \dots d}$  do
10             Compute the sample covariance  $\nabla f_{[S_k]}^{(r)}(\omega_k)$  and
11                  $\nabla f_{[S_k]}^{(r)}(\omega_0)$ , the sample variance of  $\nabla f_{[S_k]}^{(r)}(\omega_0)$  using (6)
12             Compute  $\gamma^{*(r)}$  using (2)
13         Compute  $\nabla f_{S_k}(\omega_k), \nabla f_{S_k}(\omega_0)$ 
14         Compute  $g_k = \nabla f_{S_k}(\omega_k) - \gamma^*(\nabla f_{S_k}(\omega_0) - \mu_{k-1})$ 
  
```

Assumptions

Assumption 1 (μ -strong convexity and L -smoothness)

$f_i, 1 \leq i \leq n$ is strongly convex and has Lipschitz continuous gradients, i.e.,

$$\mu I \prec \nabla^2 f_i(w) \prec LI \quad (7)$$

For $\omega \in R^d$, μ is strong convexity constant and L is Lipschitz constant.

Assumption 2 (low and upper bounds of step size) every step size α_k in Alg1 and Alg2 algorithm satisfies $\alpha_1 \leq \alpha_k \leq \alpha_2$

Assumption 3 (upper bound of scalar β_k) There exists constants β such that

$$\beta_k \leq \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \leq \beta \quad (8)$$

Related Lemma

Lemma

Under Assumption1, we have

$$2\mu(f(\omega) - f(\omega^*)) \leq \|\nabla f(w)\|^2 \leq 2L(f(\omega) - f(\omega^*)) \quad (9)$$

Where $\omega \in R^d$, ω^ is the unique minimizer*

Lemma

Consider that Alg1 and Alg2 (CG) algorithm, where step size α_k satisfies strong Wolfe condition with $0 < \sigma_2 < \frac{1}{2}$ and β_k satisfies $|\beta_k| \leq \beta_k^{FR}$, then it generates descent directions d_k satisfying

$$-\frac{1}{1 - \sigma_2} \leq \frac{\langle g_k, d_k \rangle}{\|g_k\|^2} \leq \frac{2\sigma_2 - 1}{1 - \sigma_2} \quad (10)$$

Convergence of Alg1

Theorem

Let Assumption 1,2,3 hold. If the bound of the step-size in Alg1 satisfies:

$$0 < \alpha_1 < \frac{1 - \beta}{2L\sigma_1} \quad (11)$$

Then we have: $\forall k > 0$

$$E(f(\omega_k)) - f(\omega^*) \leq C\xi^k(E(f(\omega_0)) - f(\omega^*)) \quad (12)$$

where $\xi = \frac{(1-\sigma_1)(1-\beta)+2L\alpha\sigma_1\sigma_2(1-\beta^m)}{2\mu\sigma_1m(1-\sigma_2)(1-\beta)} < 1$, ω^ is the unique minimizer of f*

Convergence of Alg2

Theorem

Let Assumption 1,2,3 hold. Let

$$m > \frac{(1 - \sigma_1) + 2L\alpha_2\sigma_1\sigma_2\beta}{2\mu\sigma_1\alpha_1(1 - \sigma_2)}$$

Then we have: $\forall l > 0$

$$E(f(x_l)) - f(\omega^*) \leq \xi^l (E(f(x_0)) - f(\omega^*)) \quad (13)$$

where

$$\xi = \frac{(1 - \sigma_1)(1 - \beta) + 2L\alpha\sigma_1\sigma_2(1 - \beta^m)}{2\mu\sigma_1 m(1 - \sigma_2)(1 - \beta)} < 1$$

and ω^* is the unique minimizer of f

- ① 问题背景与研究意义
- ② 研究现状
- ③ Algorithm and Convergence analysis
- ④ Numerical Experiments**
- ⑤ Conclusion

Datasets

Table: Summary of data sets used in numerical experiments

dataset	d	n	type
A9a	123	32561	binary classification
ljcnn1	22	49990	binary classification
Protein	74	145751	binary classification
Quantum	78	50000	binary classification
W8a	300	49749	binary classification
Covtype	54	581012	binary classification
YearPredictionMSD	90	463715	regression
Pyrim	27	74	regression
Bodyfat	24	252	regression
Triazines	60	180	regression
Eunite2001	16	336	regression

Test Function

The ridge regression model are presented as follows:

$$\min_{\omega} \frac{1}{n} \sum_{i=1}^n (y_i - x_i \omega)^2 + \lambda \|\omega\|^2 \quad (14)$$

where $x_i \in R^d$ is denoted the feature vector of the i-th data sample, $y_i \in R$ is denoted the actual value of the i-th data sample, and λ is the regularization parameter.

Numerical Results

Figure: Variance comparison of stochastic gradient estimates: $g_{S_l}^k(\gamma = \gamma^*)$ and $g_{S_l}^k(\gamma = 1)$

$$g_{S_l}^k(\gamma) = \nabla f_{S_l}(w_{101}) - \gamma(\nabla f_{S_l}(w_k) - \frac{1}{n} \sum_{i=1}^n \nabla f_i(w_k)) \quad l = 1, \dots, 100 \quad (15)$$

$$Var(g_{S_l}^k(\gamma)) = \frac{1}{100} \sum_{l=1}^{100} (g_{S_l}^k(\gamma) - \frac{1}{100} \sum_{l=1}^{100} g_{S_l}^k(\gamma))^2, \quad k = 0, 1, 2, \dots, 100 \quad (16)$$

Numerical Results

Figure: performance profiles of SCGA,CGVR,Alg1,Alg2 on the six data sets of binary classification (x-axis is times of iteration, y-axis is loss value in terms of \log_{10})

Numerical Results

Figure: performance profiles of SCGA,CGVR,Alg1,Alg2 on the six data sets of regression (x-axis is times of iteration, y-axis is loss value in terms of \log_{10})

- ① 问题背景与研究意义
- ② 研究现状
- ③ Algorithm and Convergence analysis
- ④ Numerical Experiments
- ⑤ Conclusion

Conclusion

- ① Propose the minimal variance stochastic gradient estimate
- ② Propose two improved algorithms of SCGA and CGVR: Alg1 and Alg2
- ③ Prove the linear convergence rate of the new algorithms under strong convexity and smoothness
- ④ From a series of experiments, compared with SCGA and CGVR, Alg1 and Alg2 is competitive algorithms.

Thanks for your attention!