

Optimization Method Coding Project 2: Conjugate Gradient Methods

Li, Ziyao

1500017776 / Yuanpei College

leeezy@pku.edu.cn

Abstract

This work focuses on the realization of Conjugate Gradient and Global BB Methods. Three test functions are evaluated in this work, namely Generalized BARD Function, Generalized CUBE Function and s303-305 Function. Numerical Experiments are conducted on three formerly mentioned problems. Generally FR and PRP methods takes less iterations than other CG and GBB (Global BB) methods, while adopting a non-monotone line search method, GBB do significantly outperform CGs with regard to Generalized CUBE Function.

1 Conjugate Gradient Methods

Generally the Conjugate Gradient Methods introduce the direction of last iteration when calculating a descend direction, i.e.

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

The differences of the variety of CG Methods diverge over the calculation of the proportion, i.e. β in the above formula. Five different CG methods are examined in this report with formulas:

$$\begin{aligned}\beta_{k-1}^{FR} &= \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} \\ \beta_{k-1}^{PRP} &= \frac{g_k^T (g_k - g_{k-1})}{g_{k-1}^T g_{k-1}} \\ \beta_{k-1}^{PRP+} &= \max(\beta_{k-1}^{PRP}, 0) \\ \beta_{k-1}^{CD} &= \frac{g_k^T g_k}{d_{k-1}^T g_{k-1}} \\ \beta_{k-1}^{DY} &= \frac{g_k^T g_k}{d_{k-1}^T (g_k - g_{k-1})}\end{aligned}$$

2 Numerical Performance

Performance of six different methods are tested. Relevant results are shown in Table 1, 2 and 3. The general convergence criterion, without specific notifications, are defined as

$$\begin{aligned}|f_k - f_{k-1}| &\leq \epsilon \\ \|g_k\|_\infty &\leq \epsilon \\ \epsilon &= 10^{-8}.\end{aligned}$$

The problem Generalized Bard Function is illly defined as the variables are on the denominators while the optimal is derived near 0. The function suffers from magnificent fluctuations in the neighborhood of the optimal solution. Figure 1 is an demonstration of the illness of the problem. Besides, as calculating the Generalized Bard Function itself can be rather time-costing (with the efforts of calculating and summing 15 independent variables), only smaller scale problems are experimented, i.e. $n = 10^2, 10^3, 10^4$.

Even so, the iterations observed are still significantly excessive compared to those in other two problems and the calculation time can be unacceptable for such scale problems. Conjugate Descend (CD) serves to be the least appealing methods among all three. This is partly because that an exact line search strategy is used in the experiment, while the CD method is proposed to accept weaker line search criterions instead of to improve efficiency.

The Generalized Cube Function and s303-305 Function do have better numerical properties, which is shown in the good-looking iteration data, while CD is still not a good algorithm solving these problems. Nevertheless, simply by adding a mechanism to restart the algorithm ($c = 10$ in the experiments), the performance become much better.

GEN-BARD	$n = 10^2$		$n = 10^3$		$n = 10^4$	
Method	iter	feva	iter	feva	iter	feva
FR	168	9,724	170	9,816	200	11,986
PRP	228	13,061	212	12,214	190	13,294
PRP+	228	13,061	212	12,214	188	11,962
CD	933(2)	51,324	617(2)	33,901	-	-
CD*	441(2)	24,607	711(2)	40,045	603(2)	33,890
DY	168	9,719	198	11,449	249	14,511
GBB	1,803(2)	6,082	1,664(2)	5,655	1,122(2)	3,787

-: Algorithm does not converge after 3000 iterations.

*: A restart strategy is applied to improve the method. By default the period is 5 iterations.

Table 1: Numerical performance of six methods optimizing Generalized Cube Function. The function was defined in *Extended CUTer*. A convergence criterion of $\epsilon = 10^{-3}$ is adopted if no specific remark. Iterations with bracketed numbers (b) indicate that the experiment is under a weaker convergence criterion with $-\log_{10} \epsilon = b$.

GEN-CUBE	$n = 10^3$		$n = 10^4$		$n = 10^5$	
Method	iter	feva	iter	feva	iter	feva
FR	23	1128	23	1128	22	1079
PRP	23	1128	23	1128	22	1079
PRP+	23	1128	23	1128	22	1079
CD	-	-	-	-	-	-
CD*	42	2059	52	2549	62	3039
DY	23	1128	23	1128	22	1079
GBB	41	129	41	129	45	140

-: Algorithm does not converge after 1000 iterations.

*: A restart strategy is applied to improve the method. By default the period is 10 iterations.

Table 2: Numerical performance of six methods optimizing Generalized Cube Function. The function was defined in *Extended CUTer* with a correction to x_0 ($x_0 = (0.9, 0.9, \dots, 0.9)^T$), since the original initial point is already at the global minimum, which could be a negligence in the original material.

s303-305	$n = 10^3$		$n = 10^4$		$n = 10^5$	
Method	iter	feva	iter	feva	iter	feva
FR	7	2164	9	2625	38	5074
PRP	3	929	3	782	5	1097
PRP+	3	929	5	1393	5	1097
CD	617	189,434	235	72,003	971(6)	297,952
CD*	80	24,614	116	35,497	251	76,913
DY	7	2164	4	1086	27	8146
GBB	-	-	-	-	-	-

-: Algorithm does not converge after 1000 iterations.

*: A restart strategy is applied to improve the method. By default the period is 10 iterations.

Table 3: Numerical performance of six methods optimizing s303-305 Function. The function was defined in *Extended CUTer*. Iterations with bracketed numbers (b) indicate that the experiment is under a weaker convergence criterion with $-\log_{10} \epsilon = b$.

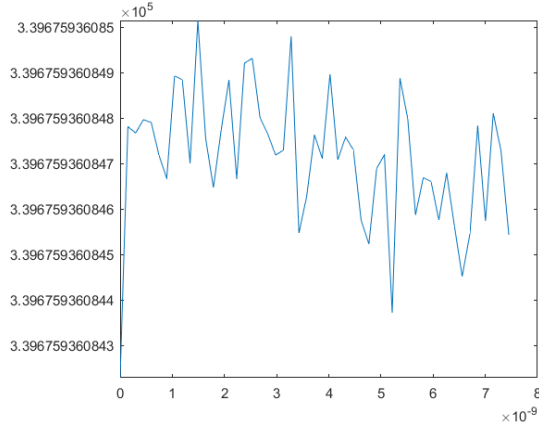


Figure 1: A Figure of function value versus step length during some iteration of the Generalized Bard Function with $n = 10^6$. The function suffers from giant fluctuations, and even a step larger than 10^{-9} cannot be derived.

BB method is a strong method with competent function-evaluation time performance, but this is partly because ELS is conducted in other experiments and an ILS method (the non-monotone line search method) is attempted here. Seeing a method only taking gradient as descending direction can converge so fast in Generalized Cube Function is still comforting, though.

Between all CG methods, no single method outperform others in all three problems. All of them except CD have robust and good performances over large scale problems.