Techniques for Optimization Methods in Applications

Rui Diao

Academy of Mathematics and Systems Science, Chinese Academy of Sciences

August 31, 2013

Why this topic

- In large scale real problems, the techniques are crucial for the efficiency of algorithms.
- Many engineers did not pay much attention to optimization techniques.
- Interests of the audience.

Techniques

- Non-monotone methods
- Subspace methods
- Inexact methods (For example, Truncated Newton method)
- Penalty methods
- ...

We mainly focus on the first two kinds of techniques.

What is non-monotone method and Why

What is non-monotone method

- In monotone methods, x_{k+1} is always “better” than x_k.
- non-monotone methods do not need x_{k+1} “better” than x_k, but good on some aspects.

Why non-monotone method

- monotone is too conservative.
- monotone is too sensitive to the initial point.
- non-monotone may jump out from the local minimum.

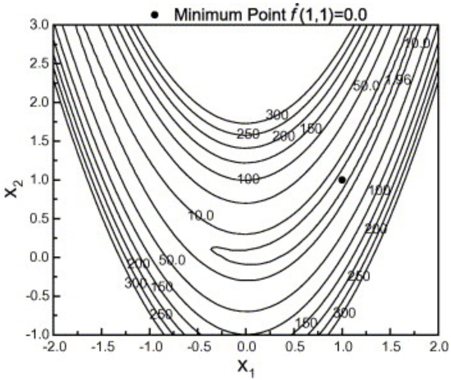

Figure: Rosenbroke Function

$$f(x, y) = (1 - x)^2 + 100(y - x^2)^2 \quad (1)$$

Monotone Line Search

Each iteration computes a search direction d_k and then decides how far to move along this direction.

$$x_{k+1} = x_k + \alpha_k d_k \tag{2}$$

α_k is called step length. Three conditions are widely used in line search.

$$f(x_k + \alpha_k d_k) \leq f(x_k) + c_1 \alpha_k \nabla f_k^T d_k \tag{3}$$

$$\nabla f(x_k + \alpha_k d_k)^T d_k \geq c_2 \nabla f_k^T d_k \tag{4}$$

$$|\nabla f(x_k + \alpha_k d_k)^T d_k| \leq c_2 |\nabla f_k^T d_k| \tag{5}$$

- Armijo line search : (3)
- Wolfe line search : (3) & (4)
- strong Wolfe line search : (3) & (5)

Non-monotone Line Search

quasi-Newton method without line search

$$x_{k+1} = x_k + H_k d_k = x_k - H_k \nabla f_k \tag{6}$$

H_k satisfies quasi-Newton condition (secant condition) :

$$H_k y_k = s_k \qquad \text{or} \qquad H_k^{-1} s_k = y_k \tag{7}$$

where $y_k = \nabla f_k - \nabla f_{k-1}$ and $s_k = x_k - x_{k-1}$.
We choose a step length α_k instead of H_k with similar property :

$$\min \|\alpha_k y_k - s_k\| \quad \text{or} \quad \min \|\frac{1}{\alpha_k} s_k - y_k\| \tag{8}$$

$$\alpha_k = \frac{s_k^T y_k}{y_k^T y_k} \quad \text{or} \quad \alpha_k = \frac{s_k^T s_k}{s_k^T y_k} \tag{9}$$

Non-monotone Line Search (Cont.)

Barzilai-Borwein method is as fast as many traditional methods in theory, but faster in practice.

References

- J. Barzilai & J. M. Borwein, Two point step size gradient methods, IMA Journal of Numerical Analysis, 8(1): 141-148, 1988
- Y.H. Dai & L.Z. Liao, R-linear convergence of the Barzilai and Borwein gradient method, IMA Journal of Numerical Analysis, 22(1): 1-10, 2002
- Y.F. Wang & S.Q. Ma, Projected Barzilai-Borwein method for large-scale nonnegative image restoration, Inverse Problems in Science and Engineering, 15(6): 559-583, 2007.

Non-monotone Line Search (Cont.)

Another way is to slightly change the Armijo line search condition to some formula like

$$f(x_k + \alpha_k d_k) \leq Z_k + c_1 \alpha_k \nabla f_k^T d_k. \tag{10}$$

Z_k is objective function value of previous iteration, or average objective function value of previous iterations.

References

- L. Grippo, F. Lampariello & S. Lucidi, A nonmonotone line search technique for Newton's method, SIAM Journal of Numerical Analysis, 23(1): 707-716, 1986.
- H.C. Zhang & W. Hager, A nonmonotone line search technique and its application to unconstrained optimization, SIAM Journal of Optimization, 14(4): 1043-1056, 2004.

Trust Region Methods

In each iteration, we solve a approximation subproblem with an extra constraint to obtain x_{k+1}:

$$\|x - x_k\| \leq \Delta_k \tag{11}$$

Δ_k is a scalar, which is called Trust Region Radius. We define

$$\rho_k = \frac{\text{Descent of Original Problem}}{\text{Descent of Subproblem}} = \frac{f(x_k) - f(x_{k+1})}{m_k(x_k) - m_k(x_{k+1})}, \tag{12}$$

where $f(x)$ is the original objective function and $m(x)$ is the objective function of subproblem.
In general, we accept x_{k+1} if and only if $\rho_k \geq \eta > 0$ (η is a given parameter), otherwise we adjust trust region radius. This makes our method monotone.

Non-monotone Trust Region

Non-monotone trust region methods have different ways to define ρ_k or η, but the key point is the same.

- possible to accept x_{k+1} when worse than x_k.
- x_{k+1} is better than previous several iterates on some aspects.

This two criterion make our method non-monotone but convergence.

References

- N.Y. Deng, Y. Xiao and F.J. Zhou, Non-monotonic trust region algorithms, Journal of Optimization Theory and Applications 76(2): 259-285, 1993.
- P.L. Toint, Non-monotone trust-region algorithms for nonlinear optimization subject to convex constraints, Mathematical Programming 77(3): 69-94, 1997.

This technique is also successfully applied to petroleum and petrochemical planning system.

What is subspace method and Why

What is subspace method

Optimal solution in a large space
→ A sequence of optimal or nearly optimal solutions in subspaces

Why subspace method

- The original problem is too difficult to solve.
- Very easy to find an optimal or nearly optimal solution in some special subspaces.
- In real problems, prior knowledge tells us the subspaces.

Subspace Methods

Subspace techniques are suitable for large scale problems.
There are ideas of subspace in many standard optimization methods.

- The step of gradient method is in one dimensional space Span$\{\nabla f_k\}$.

$$x_{k+1} = x_k - \alpha_k \nabla f_k \tag{13}$$

- The step of conjugate gradient method is in two dimensional space Span$\{\nabla f_k, d_{k-1}\}$.

$$x_{k+1} = x_k + \alpha_k d_k = x_k + \alpha_k(-\nabla f_k + \beta_k d_{k-1}) \tag{14}$$

- The step of limited memory quasi-Newton algorithms is in $2m+1$ dimensional space Span$\{\nabla f_k, s_k, \ldots, s_{k-m+1}, y_k, \ldots, y_{k-m+1}\}$

Subspace Methods (Cont.)

A model subspace algorithm for unconstrained optimization.

Step 1 Given x_1, define $\mathcal{S}_1$, $\overline{Q}_1(d)$, $\epsilon > 0$, $k := 1$.

Step 2 Solve a subspace subproblem:

$$\min_{d \in \mathcal{S}_k} \overline{Q}_k(d) \tag{15}$$

obtaining d_k. If $\|d_k\| \leq \epsilon$ then stop.

Step 3 Carry out line search to obtain $\alpha_k > 0$, set

$$x_{k+1} = x_k + \alpha_k d_k \tag{16}$$

Step 4 Generate $\mathcal{S}_{k+1}$ and $\overline{Q}_{k+1}(d)$.

Step 5 $k := k + 1$, go to Step 2.

$\overline{Q}_k(d)$ is often chosen as an approximation to the original objective function. $\mathcal{S}_k$ is the subspace. The search direction d_k is the “best” in subspace $\mathcal{S}_k$.

Subspace Methods (Cont.)

- To choose d_k in a small subspace is easy, but d_k performs bad (for example, Span$\{\nabla f\}$).
- To choose d_k in a large subspace is hard, but d_k performs better (for example, the whole space).

References

- Y.X. Yuan, Subspace techniques for nonlinear optimization, in: Some Topics in Industrial and Applied Mathematics, volume 8 of Series on Concrete and Applicable Mathematics, pp. 206-218, Higher Education Press, Beijing, 2007.
- Y.X. Yuan, Subspace methods for large scale nonlinear equations and nonlinear least squares, Optimization and Engineering, 10(2): 207-218, 2009.
- Z.K. Zhang, Subspace technique in derivative-free optimization methods, chapter 5 of On Derivative-Free Optimization Methods, PhD thesis.

Summary

- Non-monotone methods performs well in some real problems. Some of them have convergence theory for a large family of functions, but others are not.
- Subspace methods are powerful tools to deal with large scale problems. However, it is not easy to choose a good subspace.
- Some methods without nice convergence property perform better (maybe no one knows why), but the methods with convergence property are more robust.
- How to find methods with fast convergence property and also fast in practice ?