Functions of Several Variables

Menglin Liu

Group Meeting

June 18, 2010

3.3.4 Marquardt's Method

- Cauchy's
  - When $x^{(0)}$ is far from $x^{(*)}$, good reduction

$$x^{(k+1)} = x^{(k)} - \alpha^{(k)} \nabla f\left(x^{(k)}\right)$$

- Newton's
  - Ideal search directions near the solution

$$x^{(k+1)} = x^{(k)} - \mathbf{H}^{(k)} \nabla f\left(x^{(k)}\right)$$

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} s\left(x^{(k)}\right)$$

$$s\left(x^{(k)}\right) = -\left[\mathbf{H}^{(k)} + \lambda^{(k)}\mathbf{I}\right]^{-1} \nabla f\left(x^{(k)}\right) \qquad \alpha^{(k)} = 1$$

Step 1. Define $x^{(0)}$ = initial estimate of x^*

M = maximum number of iterations allowed

ε = convergence criterion

Step 2. Set $k = 0$. $\lambda^{(0)} = 10^4$.

Step 3. Calculate $\nabla f(x^{(k)})$.

Step 4. Is $\|\nabla f(x^{(k)})\| < \varepsilon$?

Yes: Go to step 11.

No: Continue.

Step 5. Is $k \geqslant M$?

Yes: Go to step 11.

No: Continue.

Step 6. Calculate $s(x^{(k)}) = -[\mathbf{H}^{(k)} + \lambda^{(k)}\mathbf{I}]^{-1}\,\nabla f(x^{(k)})$.

Step 7. Set $x^{(k+1)} = x^{(k)} + s(x^{(k)})$.

Step 8. Is $f(x^{(k+1)}) < f(x^{(k)})$?

Yes: Go to step 9.

No: Go to step 10.

Step 9. Set $\lambda^{(k+1)} = \frac{1}{2}\lambda^{(k)}$ and $k = k + 1$. Go to step 3.

Step 10. Set $\lambda^{(k)} = 2\lambda^{(k)}$. Go to step 6.

Step 11. Print results and stop.

termination criteria

- Advantages
  - Simplicity, descent property
  - excellent convergence rate near x^*
  - Absence of a line search
- Disadvantages
  - Need to calculate $\mathbf{H}^{(k)}$

3.3.5 Conjugate Gradient Methods

- *Quadratically convergent*:

  terminates in approximately N steps when applied to a quadratic function

  - Employ gradient information to generate *conjugate directions*

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} s\left(x^{(k)}\right)$$

$$s^{(k)} = -g^{(k)} + \sum_{i=0}^{k-1} \gamma^{(i)} s^{(i)}$$

$$s^{(k)} = -g^{(k)} + \left[\frac{||g^{(k)}||^2}{||g^{(k-1)}||^2}\right] s^{(k-1)}$$

- *f(x)* quadratic
  - *N-1* directions and *N* line searches
- *f(x)* not quadratic
  - Additional directions and line searches
  - Restart every *N* or *N+1* steps

3.3.6 quasi-Newton Methods

- Based on properties of quadratic functions
- Mimic Newton's method using only first-order

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} s\left(x^{(k)}\right)$$

$$s\left(x^{(k)}\right) = -\mathbf{A}^{(k)} \nabla f\left(x^{(k)}\right)$$

$$\Delta g = \mathbf{C} \Delta x$$

$$\mathbf{A}^{(k+1)} = \mathbf{A}^{(k)} + \mathbf{A}_c^{(k)}$$

$$\mathbf{A}^{(k)} = \mathbf{A}^{(k-1)} + \frac{\Delta x^{(k-1)} \, \Delta x^{(k-1)\mathrm{T}}}{\Delta x^{(k-1)\mathrm{T}} \, \Delta g^{(k-1)}} - \frac{\mathbf{A}^{(k-1)} \, \Delta g^{(k-1)} \, \Delta g^{(k-1)\mathrm{T}} \, \mathbf{A}^{(k-1)}}{\Delta g^{(k-1)\mathrm{T}} \, \mathbf{A}^{(k-1)} \, \Delta g^{(k-1)}}$$

3.3.7 Trust Regions

- Gradient-based methods
  - Direction search
  - Line search
- Trust region methods
  - Form a "trustworthy" approximation of $f(x)$
  - Find the minimum of the approximation
- Marquardt's methods
  - λ determines the approximation
  - Approximation becomes quadratic in the limit

3.3.8 Gradient-based Algorithm

A general algorithm including all of the methods

Step 1. Define M = maximum number of allowable iterations

N = number of variables

$x^{(0)}$ = initial estimate of x^*

ε_1 = overall convergence criteria

ε_2 = line search convergence criteria

Step 2. Set $k = 0$.

Step 3. Calculate $\nabla f(x^{(k)})$.

Step 4. Is $\|\nabla f(x^{(k)}\| \leqslant \varepsilon_1$?

Yes: Print "convergence: gradient"; go to 13.

No: Continue.

Step 5. Is $k > M$?

Yes: Print "termination: $k = M$"; go to 13.

No: Continue.

Step 6. Calculate $s(x^{(k)})$.

Step 7. Is $\nabla f(x^{(k)})s(x^{(k)}) < 0$?

Yes: Go to 9.

No: Set $s(x^{(k)}) = -\nabla f(x^{(k)})$. Print "restart: bad direction"; Go to 9.

Step 8. Find $\alpha^{(k)}$ such that $f(x^{(k)} + \alpha^{(k)}s(x^{(k)})) \rightarrow$ minimum using ε_2.

Step 9. Set $x^{(k+1)} = x^{(k)} + \alpha^{(k)}s(x^{(k)})$.

Step 10. Is $f(x^{(k+1)}) < f(x^{(k)})$?

Yes: Go to 11.

No: Print "termination: no descent"; go to 13.

Step 11. Is $\|\Delta x\|/\|x^{(k)}\| \leqslant \varepsilon_1$?

Yes: Print "termination: no progress"; go to 13.

No: Go to 12.

Step 12. Set $k = k + 1$. Go to 3.

Step 13. Stop.

3.3.9 Numerical Gradient Approximations

- So far, the gradient and Hessian matrix available.
- In practical, not really.

$$\left.\frac{\partial f(x)}{\partial x_i}\right|_{x=\bar{x}} = \frac{f(\bar{x} + \varepsilon e^{(i)}) - f(\bar{x})}{\varepsilon}$$

$$\left.\frac{\partial f(x)}{\partial x_i}\right|_{x=\bar{x}} = \frac{f(\bar{x} + \varepsilon e^{(i)}) - f(\bar{x} - \varepsilon e^{(i)})}{2\varepsilon}$$

3.4 Comparison of Methods and Numerical Results

- What is known about the efficiency comes from numerical experiments
- Results in publications
- [H72,SS71,CS75]

Summary

- Necessary and sufficient conditions for existence of a minimum of a function
- Survey of methods