Structural Optimization for Large-Scale Problems

Lecture 3: Second-order methods. Systems of nonlinear equations

Yurii Nesterov

Minicourse: November 15, 16, 22, 23, 2024 (SDS, Shenzhen)

Outline

Historical remarks

§1 历史方法

① 牛顿法

Problem: $f(x) \quad \to \quad \min : \quad x \in \mathbb{R}^n$

is treated as a non-linear system $f'(x) = 0$.

Newton method: $x_{k+1} = x_k - [f''(x_k)]^{-1} f'(x_k)$.

Standard objections:

- The method is not always well defined ($\det f''(x_k) = 0$).
- Possible divergence.
- Possible convergence to saddle points or even to local maximums.
- Chaotic global behavior.

Pre-History (see Ortega, Rheinboldt [1970].)

② 对牛顿法的改进

- *Bennet [1916]:* Newton's method in general analysis.
- *Levenberg [1944]:* Regularization. If $f''(x) \not\succ 0$, then use $d = G^{-1} f'(x)$ with $G = f''(x) + \gamma I \succ 0$. (See also *Marquardt [1963]*.)
- *Kantorovich [1948]:* Proof of local quadratic convergence. Assumptions:
  a) $f \in C^3(\mathbb{R}^n)$.
  b) $\|f''(x) - f''(y)\| \leq L_2 \|x - y\|$.
  c) $f''(x^*) \succ 0$.
  d) $x_0 \approx x^*$.

Global convergence: Use line search (good advice).

Global performance: Not addressed.

Modern History (see in Conn, Gould and Toint [2000])

③ Trust region approach

Main idea: *Trust Region Approach*.

1. Use some norm $\|\cdot\|_k$ for defining trust region
$$\mathcal{B}_k = \{x \in \mathbb{R}^n : \ \|x - x_k\|_k \le r_k\}.$$ (确保每次移动不超过 r)

2. Denote $m_k(x) = f(x_k) + \langle f'(x_k), x - x_k \rangle + \frac{1}{2}\langle G_k(x - x_k), x - x_k \rangle$.
Variants: $G_k = f''(x_k)$, $G_k = f''(x_k) + \gamma_k I \succ 0$, etc. (局部用 m(x) 逼近)

3. Compute the trial point $\hat{x}_k = \arg\min_{x \in \mathcal{B}_k} m_k(x)$.

4. Compute the ratio $\rho_k = \frac{f(x_k) - f(\hat{x}_k)}{f(x_k) - m_k(\hat{x}_k)}$. $= \frac{\text{f的下落值}}{\text{m的下落值}}$ (确保逼近的准确性)

5. In accordance to ρ_k, either accept $x_{k+1} = \hat{x}_k$, or update the value r_k and repeat the steps above.

Comments

Advantages:

- More parameters $\quad\Rightarrow\quad$ Flexibility
- Convergence to a point, which satisfies second-order necessary optimality condition:

$$f'(x^*) = 0, \quad f''(x^*) \succeq 0.$$

- In case of Euclidean norm, the auxiliary problem is one-dimensional:

$$\text{Find } \lambda > 0\text{: } \|(G_k + \lambda I)^{-1} f'(x_k)\| = r_k.$$

Disadvantages:

- Complicated strategies for parameters' coordination.
- For certain $\|\cdot\|_k$ the auxiliary problem is difficult.
- Line search abilities are quite limited.
- Unselective theory.
- Global complexity issues are not addressed.

Development of numerical schemes

Classical style: $\boxed{\text{Problem formulation} \Rightarrow \text{Method}}$

Examples:

- Gradient and Newton methods in optimization.
- Runge-Kutta method for ODE, etc.

2. **Modern style:**

$$\boxed{\left.\begin{array}{l}\text{Problem formulation}\\ \text{Problem class}\end{array}\right\} \quad \Rightarrow \quad \text{Method}}$$

Examples:

- Non-smooth convex minimization.
- Smooth minimization: $\min\limits_{x \in Q} f(x)$, with $f \in C^{1,1}$.

  Gradient mapping (Nemirovsky&Yudin 77):

$$x_+ \quad = \quad T(x) \equiv \arg\min_{y \in Q} \; m_1(y),$$

$$m_1(y) \quad \equiv \quad f(x) + \langle f'(x), y - x \rangle + \tfrac{L_1}{2} \|y - x\|^2.$$

Justification: $f(y) \leq m_1(y)$ for all $y \in Q$.

§2 unconstrained convex problem: second-order model with cubic regularization (Cubic Newton Method)

Using the second-order model

Problem: $\boxed{f(x) \quad \to \quad \min : \; x \in \mathbb{R}^n}$ Let $f \in C^3$ be convex.

Hessian Lipschitz cont.

Assumption: $\|f''(x) - f''(y)\| \le L_2 \|x - y\|$ $\forall x, y \in \mathbb{R}^n$. Define

$$m_2(x, y) = f(x) + \langle f'(x), y - x \rangle + \tfrac{1}{2}\langle f''(x)(y - x), y - x \rangle,$$

$$m_2'(x, y) = f'(x) + f''(x)(y - x).$$

Lemma 1:

泰勒展开

$$|f(y) - m_2(x, y)| \le \tfrac{1}{6} L_2 \|y - x\|^3, \quad \|f'(y) - m_2'(x, y)\| \le \tfrac{1}{2} L_2 \|y - x\|^2$$

Proof: By Taylor formula, we have

$$f(y) - m_2(x, y) = \tfrac{1}{2} \int_0^1 (1 - \tau)^2 D^3 f(x + \tau(y - x))[y - x]^3 d\tau,$$

$$f'(y) - m_2'(x, y) = \int_0^1 (1 - \tau) D^2 f(x + \tau(y - x))[y - x]^2 d\tau. \; \Box$$

Corollary: For any x and y from $\mathbb{R}^n$,

descent lemma (若 L_2 已知)

$$\boxed{f(y) \le m_2(x, y) + \tfrac{1}{6} L_2 \|y - x\|^3}$$

Cubic regularization

现实中L_2通常未知，因此人为添加某 constant M

For $M > 0$, denote $\boxed{\hat{f}_M(x, y) = m_2(x, y) + \frac{1}{6}M\|y - x\|^3}$ and

$$\boxed{T_M(x) = \arg\min_y \ \hat{f}_M(x, y),} \quad \boxed{\bar{f}_M(x) \stackrel{\text{def}}{=} \hat{f}_M(x, T_M(x)).}$$

Note that $\boxed{\bar{f}_M(x) \leq \min_y \left[f(y) + \frac{L_2+M}{6}\|y - x\|^3\right]}$

First-order optimality condition

$$\boxed{f'(x) + f''(x)(T - x) + \frac{M}{2}r(T - x) = 0, \quad r = \|T - x\|} \quad (*)$$

Thus, we need to solve the equation （为了求出 $T_M(x)$）

$$\boxed{\|f''(x)(T - x) + \frac{M}{2}rI)^{-1}f'(x)\| = r}$$

(Compare with Trust Region.)

Main properties

Denote $r_M(x) = \|x - T_M(x)\|$. Multiplying $(*)$ by $T_M - x$, we get:

$$\langle f'(x), T_M - x\rangle + \langle f''(x)(T_M - x), T - x\rangle + \frac{M}{2} r^3 = 0.$$

If $M \geq L_2$, then

$$f(T_M) \leq \bar{f}_M(x) = f(x) - \frac{1}{2}\langle f''(x)(T_M - x), T_M - x\rangle - \frac{M}{3} r_M^3,$$

Note that by $(*)$ we have

$$\begin{aligned} \langle f'(T_M), x - T_M\rangle &= \langle f'(T_M) - m_2'(x, T_M) - \frac{M}{2} r_M(T_M - x), x - T_M\rangle \\ &\geq \frac{M-L_2}{2} r_M^3. \end{aligned}$$

On the other hand,

$$\|f'(T_M)\| = \|f'(T_M) - m_2'(x, T_M) - \frac{M}{2} r_M(T_M - x)\| \leq \frac{L_2+M}{2} r_M^2.$$

Therefore,

descent lemma (monotonicity)

$$\boxed{f(x) - f(T_M) \geq \langle f'(T_M), x - T_M\rangle \geq \frac{M-L_2}{2}\left[\frac{2}{L_2+M}\|f'(T_M)\|\right]^{3/2}}$$

For $M = 3L_2$, we get $\langle f'(T_M), x - T_M\rangle \geq \frac{1}{2\sqrt{2L_2}}\|f'(T_M)\|^{3/2}$.

Globar rate of convergence, I

Let us assume that we know $M \geq L_2$. Let us choose $x_0 \in \mathbb{R}^n$ and consider the following method: $\boxed{x_{k+1} = T_M(x_k)}$ (locally method)

Assumption: The level sets of the objective function are bounded:

$$\|x - x^*\| \leq D, \quad \forall x: \; f(x) \leq f(x_0).$$

函数值为f(x0)的点到x*的距离最大值

Since the method is monotone, we have $\|x_k - x^*\| \leq D$, $k \geq 0$.

Therefore,

$$\begin{aligned} f(x_{k+1}) &\leq \min_y \left[f(y) + \tfrac{L_2+M}{6}\|y - x_k\|^3 \right] \\ &\leq \min_{0 \leq \alpha \leq 1} \left[f((1-\alpha)x_k + \alpha x^*)) + \tfrac{(L_2+M)\alpha^3}{6}\|x_k - x^*\|^3 \right] \\ &\leq \min_{0 \leq \alpha \leq 1} \left[f(x_k) - \alpha(f(x_k) - f^*) + \tfrac{(L_2+M)\alpha^3}{6} D^3 \right]. \end{aligned}$$

The optimal solution of this problem is $\alpha_k^* = \min\left\{1, \sqrt{\frac{2(f(x_k)-f^*)}{(L_2+M)D^3}}\right\}$.

If $\alpha_k^* < 1$, then $f(x_{k+1}) \leq f(x_k) - \frac{2}{3}\sqrt{\frac{2}{(L_2+M)D^3}}(f(x_k) - f^*)^{3/2}$.

Globar rate of convergence, II

Denote $\mu_k = f(x_k) - f^*$ and $\gamma = \frac{(L_2+M)D^3}{6}$. For $k = 0$, we have:

$$\mu_1 \leq \max_{\mu_0} \left\{ \gamma, \mu_0 - \frac{2}{3\sqrt{3\gamma}} \mu_0^{3/2} \right\} = \gamma.$$

Thus, $\alpha_k^* < 1$ for all $k \geq 1$, and we conclude that

$$\frac{1}{\mu_{k+1}^{1/2}} - \frac{1}{\mu_k^{1/2}} = \frac{\mu_k^{1/2} - \mu_{k+1}^{1/2}}{\mu_k^{1/2}\mu_{k+1}^{1/2}} = \frac{\mu_k - \mu_{k+1}}{\mu_k^{1/2}\mu_{k+1}^{1/2}(\mu_k^{1/2} + \mu_{k+1}^{1/2})} \geq \frac{\mu_k - \mu_{k+1}}{2\mu_k^{3/2}} \geq \frac{1}{3\sqrt{3\gamma}}.$$

This means that $\frac{1}{\mu_k^{1/2}} \geq \frac{1}{\sqrt{\gamma}} + \frac{k-1}{3\sqrt{3\gamma}}$.

Thus, we have proved the following bound: $\boxed{f(x_k) - f^* \leq \frac{(L_2+M)D^3}{6\left(1 + \frac{k-1}{3\sqrt{3}}\right)^2}}$

Local quadratic convergence: Assume $f''(x) \geq \mu I$ with $\mu > 0$.

Then $x - T_M = [f''(x) + \frac{M}{2} r_M I]^{-1} f'(x)$. Therefore,

$$r_M = \|[f''(x) + \tfrac{M}{2} r_M I]^{-1} f'(x)\| \leq \tfrac{1}{\mu} \|f'(x)\|.$$

Hence $\boxed{\|f'(T_M)\| \leq \frac{L_2+M}{2\mu^2} \|f'(x)\|^2}$ 局部二阶收敛（不差于牛顿法）

§3 Accelerated Cubic Newton Method

Accelerated Newton: Cubic prox-function

x_0: starting point

Denote $d(x) = \frac{1}{3}\|x - x_0\|^3$.

Lemma. Cubic prox-function is *uniformly convex*: for all $x, y \in \mathbb{R}^n$,

$$\langle d'(x) - d'(y), x - y \rangle \geq \frac{1}{2}\|x - y\|^3,$$

$$d(x) - d(y) - \langle d'(y), x - y \rangle \geq \frac{1}{6}\|x - y\|^3.$$

Moreover, its Hessian is Lipschitz continuous:

$$\|d''(x) - d''(y)\| \leq 2\|x - y\|, \; x, y \in \mathbb{R}^n.$$

Remark. In our constructions, we are going to use $d(x)$ instead of the standard *strongly convex* prox-functions.

Prox-function $d(x)$**:** a differentiable ① strongly convex function:

$d(y) \geq d(x) + \langle \nabla d(x), y - x \rangle + \frac{1}{2}\|x - y\|^2, \quad x, y \in \text{rint}\, Q.$

② 最小值为0

Let $d(x)$ attain its minimum on Q at x_0, and $d(x_0) = 0$.

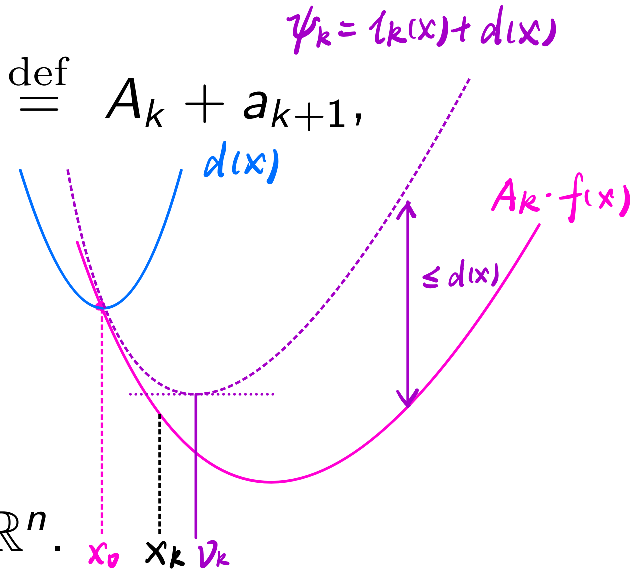
Thus, $\quad d(x) \geq \frac{1}{2}\|x - x_0\|^2, \quad x \in Q.$

Linear Estimating Functions

(Compare with 1st-order methods)

We recursively update the following sequences.

- Sequence of estimating functions $\psi_k(x) = \ell_k(x) + d(x)$, $k \geq 0$, where $\ell_k(x)$ are linear and $\ell_0(x) \equiv 0$.
- A minimizing sequence $\{x_k\}_{k=0}^{\infty}$.
- Sequence of scaling parameters $\{A_k\}_{k=0}^{\infty}$: $A_{k+1} \stackrel{\text{def}}{=} A_k + a_{k+1}$, with $A_0 = 0$.

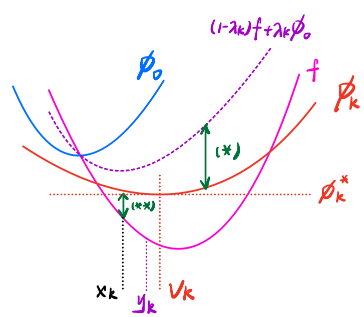


These objects have to satisfy the following relations:

$$(*): \quad \begin{array}{lrcl} ① & A_k f(x_k) & \leq & \psi_k^* \equiv \min_x \psi_k(x), \\ ② & \psi_k(x) & \leq & A_k f(x) + d(x), \quad \text{for all } x \in \mathbb{R}^n. \end{array}$$

$$(\Rightarrow \underbrace{A_k}_{\uparrow \text{ as } k\uparrow} \underbrace{(f(x_k) - f(x^*))}_{\downarrow \text{ as } k\uparrow} \leq d(x^*).)$$

For $k = 0$, we have $A_0 = 0$ and $\psi_0^* = 0$.



One interation

Denote $v_k = \arg\min_x \ \psi_k(x)$.

For some $a_{k+1} > 0$ and $M = 3L_2$, define

此时还用确定$\{A_k\}$

$$\alpha_k \ = \ \frac{a_{k+1}}{A_k + a_{k+1}} \in (0, 1),$$

$$y_k \ = \ (1 - \alpha_k) x_k + \alpha_k v_k,$$

$$x_{k+1} \ = \ T_M(y_k),$$

$$\psi_{k+1}(x) = \psi_k(x) + a_{k+1}[f(x_{k+1}) + \langle f'(x_{k+1}), x - x_{k+1} \rangle].$$

Theorem 1. Let us choose $M = 3L_2$ and $A_0 = 0$.

Let the coefficients $\{a_k\}_{k \geq 1}$ satisfy the following condition:

$$\boxed{(A_k + a_{k+1})^2 \geq \frac{64}{9} L_2 a_{k+1}^3} \qquad (k \geq 0)$$

Then, for all $k \geq 0$, we have $\boxed{A_k f(x_k) \leq \psi_k^* \ \equiv \ \min_x \psi_k(x)}$

$\Rightarrow$ 如此构造的$\{A_k\}$可使(*)回满足

Proof of Theorem 1

Note that

$$\begin{aligned}\psi_k(v_{k+1}) \quad &\geq \quad \psi_k^* + \tfrac{1}{6}\|v_{k+1} - v_k\|^3 \;\geq\; A_k f(x_k) + \tfrac{1}{6}\|v_{k+1} - v_k\|^3 \\ &\geq \quad A_k[f(x_{k+1}) + \langle f'(x_{k+1}), x_k - x_{k+1}\rangle] + \tfrac{1}{6}\|v_{k+1} - v_k\|^3.\end{aligned}$$

Therefore, $\psi_{k+1}^* = \psi_k(v_{k+1}) + a_{k+1}[f(x_{k+1}) + \langle f'(x_{k+1}), v_{k+1} - x_{k+1}\rangle]$

$$\geq A_{k+1} f(x_{k+1}) + \langle f'(x_{k+1}), a_{k+1} v_{k+1} + A_k x_k - A_{k+1} x_{k+1}\rangle + \tfrac{1}{6}\|v_{k+1} - v_k\|^3.$$

Note that $\quad a_{k+1}\langle f'(x_{k+1}), v_{k+1} - v_k\rangle + \frac{1}{6}\|v_{k+1} - v_k\|^3$

$$\geq -\tfrac{2\sqrt{2}}{3}\left[a_{k+1}\|f'(x_{k+1})\|\right]^{3/2}, \text{ and}$$

$$\begin{aligned}\langle f'(x_{k+1}), a_{k+1} v_k + A_k x_k - A_{k+1} x_{k+1}\rangle &= A_{k+1}\langle f'(x_{k+1}), y_k - x_{k+1}\rangle \\ &\geq \tfrac{A_{k+1}}{2\sqrt{2L_2}}\|f'(x_{k+1})\|^{3/2}.\end{aligned}$$

Hence, we get inequality $A_{k+1} \geq \frac{8}{3}\sqrt{L_2}\, a_{k+1}^{3/2}$. $\square$

Global rate of convergence

找到了一个合适的$\{A_k\}$

Th.2. Sequence $A_k = \frac{1}{3L_2}\left(\frac{k}{4}\right)^3$, $k \geq 0$, satisfies condition of Theorem 1.

Hence, for any $k \geq 1$, we have $\boxed{f(x_k) - f(x^*) \;\leq\; \frac{3L_2}{2}\|x_0 - x^*\|^3 \left(\frac{4}{k}\right)^3}$ $O(\frac{1}{k^3})$

Proof. For $B_{k+1} \stackrel{\text{def}}{=} (k+1)^3$ and $b_{k+1} \stackrel{\text{def}}{=} B_{k+1} - B_k$, $k \geq 0$, we have

$$B_{k+1}^{2/3} = (k+1)^2 \geq \tfrac{1}{3}[3k^2 + 3k + 1] = \tfrac{1}{3} b_{k+1}.$$

Let us define $A_{k+1} = \alpha B_{k+1}$ with some $\alpha > 0$. Then

$$A_{k+1} \geq \alpha \left[\tfrac{1}{3\alpha} a_{k+1}\right]^{3/2}.$$

Thus, we need to choose $\frac{1}{3\sqrt{3\alpha}} = \frac{8}{3}\sqrt{L_2}$. □

Accelerated CNM

Initialization: Set $\psi_0(x) = d(x)$. Define $A_k = \frac{1}{3L_2}\left(\frac{k}{4}\right)^3$, $k \geq 0$.

Iteration k, ($k \geq 0$): $v_k = \arg\min_{x \in \mathbb{R}^n} \psi_k(x)$,

$a_{k+1} := A_{k+1} - A_k$

$y_k = \frac{A_k}{A_{k+1}} x_k + \frac{a_{k+1}}{A_{k+1}} v_k, \quad x_{k+1} = T_{3L_2}(y_k)$,

$\psi_{k+1}(x) = \psi_k(x) + a_{k+1}[f(x_{k+1}) + \langle f'(x_{k+1}), x - x_{k+1}\rangle]$.

Remark:

Instead of recursive computation of $\psi_k(x)$, we can update only one vector:

$$s_0 = 0, \quad s_{k+1} = s_k + a_{k+1} f'(x_{k+1}), \quad k \geq 1.$$

Then v_k can be computed by an explicit expression.

§4 Second-order method 的 nondegeneracy (为了确定 CNM 的 complexity)

Global non-degeneracy

non-degenerate problem 的定义

Standard setting: for convex $f \in C^2(\mathbb{R}^n)$, define positive constants σ_1 and L_1 such that

$$\sigma_1 \|h\|^2 \leq \langle f''(x)h, h\rangle \leq L_1 \|h\|^2$$

for all $x, y, h \in \mathbb{R}^n$. The value $\gamma_1(f) = \frac{\sigma_1}{L_1}$ is called the *condition number* of f.

(Compatible with definition in Linear Algebra.) How good is the gradient for approaching the optima?

Geometric interpretation: $\frac{\langle f'(x), x - x^* \rangle}{\|f'(x)\| \cdot \|x - x^*\|} \geq \frac{2\sqrt{\gamma_1(f)}}{1+\gamma_1(f)}$, $x \in \mathbb{R}^n$.

Complexity: (1st-order methods)

PGM: $O\left(\frac{1}{\gamma_1(f)} \cdot \ln \frac{1}{\epsilon}\right)$, **FGM:** $O\left(\frac{1}{\sqrt{\gamma_1(f)}} \cdot \ln \frac{1}{\epsilon}\right)$.

It *does not work* for 2nd-order schemes:

$$f(x_k) - f^* \leq \frac{3\mathbf{L}_2}{2} \|x_0 - x^*\|^3 \left(\frac{4}{k}\right)^3.$$

Hessian 的 Lip-const

Global 2nd-order non-degeneracy

Assumption: for any $x, y \in \mathbb{R}^n$, function $f \in C^2(\mathbb{R}^n)$ satisfies inequalities

$$\begin{aligned} \|f''(x) - f''(y)\| &\leq L_2 \|x - y\|, \\ \langle f'(x) - f'(y), x - y \rangle &\geq \sigma_2 \|x - y\|^3, \end{aligned}$$

where $\sigma_2 > 0$. We call the value $\gamma_2(f) = \frac{\sigma_2}{L_2} \in (0, 1)$ the *2nd-order condition number* of function f.

(Invariant w.r.t. addition of convex quadratic functions.)

Example: $\gamma_2(d) = \frac{1}{4}$.

Justification: $\frac{\sigma_2}{3} \|x_k - x^*\|^3 \leq f(x_k) - f^* \leq \frac{3L_2}{2} \|x_0 - x^*\|^3 \left(\frac{4}{k}\right)^3$.

Hence, in $O\left(\frac{1}{[\gamma_2(f)]^{1/3}}\right)$ iterations we halve the distance to x^*.

Complexity bound: (Accelerated CNM with restart)

$$O\left(\frac{1}{[\gamma_2(f)]^{1/3}} \cdot \ln \frac{1}{\epsilon}\right) \text{ iterations.}$$

§5 解非线性方程组的方法

Solving the systems of nonlinear equations

1. Standard Gauss-Newton method

Problem: Find $x \in \mathbb{R}^n$ satisfying the system $F(x) = 0 \in \mathbb{R}^m$.

Assumption: $\forall x, y \in \mathbb{R}^n \quad \|F'(x) - F'(y)\| \le L\|x - y\|$.

Gauss-Newton method: Choose a merit function $\phi(u) \ge 0$, $\phi(0) = 0$, $u \in \mathbb{R}^m$.

Compute $x_+ \in \operatorname{Arg}\min_y \left[\phi(F(x) + F'(x)(y - x))\right]$.

Usual choice: $\phi(u) = \sum_{i=1}^{m} u_i^2$. (Justification: *Why not?*)

(根据先前的讨论,这只是一个 idea of method)

Remarks

- Local quadratic convergence ($m \ge n$, non-degeneracy and $F(x^*) = 0$ (?)).
- If $m < n$, then the method is not well-defined.
- No global complexity results.

Modified Gauss-Newton method

Lemma. For all $x, y \in \mathbb{R}^n$, we have

$$\|F(y) - F(x) - F'(x)(y - x)\| \leq \frac{1}{2}L\|y - x\|^2.$$

Corollary. Denote $f(y) = \|F(y)\|$. Then

$$f(y) \leq \|F(x) + F'(x)(y - x)\| + \frac{1}{2}L\|y - x\|^2.$$

Modified method:

$$x_{k+1} = \arg\min_{y} \left[\|F(x_k) + F'(x_k)(y - x_k)\| + \frac{1}{2}L\|y - x_k\|^2 \right].$$

Remarks

- The merit function is non-smooth.
- Nevertheless, $f(x_{k+1}) < f(x_k)$ unless x_k is a stationary point.
- Quadratic convergence for non-degenerate solutions.
- Global efficiency bounds.
- Problem of finding x_{k+1} is convex.
- Different norms in $\mathbb{R}^n$ and $\mathbb{R}^m$ can be used.

Testing CNM: Chebyshev oscilator

Consider $f(x) = \frac{1}{4}(1 - x^{(1)})^2 + \sum\limits_{i=1}^{n-1} \left(x^{(i+1)} - p_2(x^{(i)}) \right)^2$,

with $p_2(\tau) = 2\tau^2 - 1$.

Note that p_2 is a Chebyshev polynomial: $\quad p_k(\tau) = \cos(k \arccos(\tau))$.

Hence, the equations for the "central path" is

$$x^{(i+1)} = p_2(x^{(i)}) = p_4(x^{(i-1)}) = \ldots = p_{2^i}(x^{(1)}).$$

This is an exponential oscillation! However, all coefficients in function and derivatives are small.

NB: $\quad f(x)$ is unimodular and $x^* = (1, \ldots, 1)$.

In our experiments we usually take $x_0 = (-1, 1, \ldots, 1)$.

Drawback: $x_0 - 2\nabla f(x_0) = x^*$. Hence, sometimes we use $x_0 = (-1, 0.9. \ldots, 0.9)$.

Solving Chebyshev oscilator by CN: $\|\nabla f(x)\|_{(2)} \leq 10^{-8}$

n	Iter	DF	GNorm	NumF	Time (s)
2	14	$7.0 \cdot 10^{-19}$	$4.2 \cdot 10^{-09}$	18	0.032
3	33	$1.1 \cdot 10^{-24}$	$7.5 \cdot 10^{-12}$	51	0.031
4	82	$1.7 \cdot 10^{-20}$	$9.3 \cdot 10^{-10}$	148	0.047
5	207	$4.5 \cdot 10^{-19}$	$1.2 \cdot 10^{-09}$	395	0.078
6	541	$1.0 \cdot 10^{-17}$	$5.6 \cdot 10^{-09}$	1062	0.266
7	1490	$1.4 \cdot 10^{-18}$	$2.9 \cdot 10^{-09}$	2959	0.609
8	4087	$2.7 \cdot 10^{-17}$	$9.1 \cdot 10^{-09}$	8153	1.782
9	11205	$1.6 \cdot 10^{-16}$	$9.6 \cdot 10^{-09}$	22389	5.922
10	30678	$2.7 \cdot 10^{-15}$	$9.6 \cdot 10^{-09}$	61335	18.89
11	79292	$7.7 \cdot 10^{-14}$	$1.0 \cdot 10^{-08}$	158563	57.813
12	171522	$9.7 \cdot 10^{-13}$	$9.9 \cdot 10^{-09}$	343026	144.266
13	385353	$1.3 \cdot 10^{-11}$	$9.9 \cdot 10^{-09}$	770691	347.094
14	938758	$2.1 \cdot 10^{-11}$	$1.0 \cdot 10^{-08}$	1877500	1232.953
15	2203700	$7.8 \cdot 10^{-11}$	$1.0 \cdot 10^{-08}$	4407385	3204.359

Other methods

	Trust	region	Knitro	Minos	5.5	Snopt	
n	Inner	Iter	Iter	Iter	NFG	Iter$^{\#}$	NFG
3	129	50	30	44	120	106	78
4	431	123	80	136	309	268	204
5	1310	299	203	339	793	647	509
6	3963	722	531	871	2022	1417	1149*
7	12672	1921	1467	2291	5404	$***$	
8	40036	5234	4040	6109	14680		
9	120873	13907	11062	11939	28535		
10	358317	36837	29729*	$***$			
11	842368	78854	$***$				
12	2121780	182261					

Notation: * early termination, $(***)$ numerical difficulties/ inaccurate solution, $^{\#}$ needs an alternative starting point.

Trust region: very reliable, but $T(12) = 2577$ sec (Matlab), $T(n) = Const * (4.5)^n$.