# Lecture 3: Second-order methods. Solving systems of nonlinear equations

Yurii Nesterov, CORE/INMA (UCL)

June 18, 2013 (Senege)

5th Traditional School on Control, Information, and Optimization

# Outline

# Historical remarks

**Problem:** $f(x) \quad \to \quad \min : \quad x \in R^n$
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].)

- *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) $f \in C^3(R^n)$.
  - 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])

**Main idea:** *Trust Region Approach.*

1. By some norm $\| \cdot \|_k$ define the trust region
$\mathcal{B}_k = \{x \in R^n : \; \|x - x_k\|_k \leq \Delta_k\}$.

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.

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

4. Compute the ratio $\rho_k = \frac{f(x_k) - f(\hat{x}_k)}{f(x_k) - m_k(\hat{x}_k)}$.

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

# Comment

**Advantages:**

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

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

**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:** Problem formulation $\Rightarrow$ Method

Examples:

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

**2. Modern style:** $\left.\begin{array}{l}\text{Problem formulation}\\ \text{Problem class}\end{array}\right\} \Rightarrow$ 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)*:

$$\begin{aligned} x_+ &= T(x) \equiv \arg\min_{y \in Q} \; m_1(y), \\ m_1(y) &\equiv f(x) + \langle f'(x), y - x \rangle + \tfrac{L_1}{2} \| y - x \|^2. \end{aligned}$$

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

# Using the second-order model

**Problem:** $f(x) \quad \min : \; x \in R^n$.

**Assumption:** Let $\mathcal{F}$ be an open convex set. Then

$$\|f''(x) - f''(y)\| \leq L_2 \|x - y\| \quad \forall x, y \in \mathcal{F},$$

$$\mathcal{L}(x_0) = \{x \in R^n : \; f(x) \leq f(x_0)\} \subset \mathcal{F}.$$

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.** For any $x, y \in \mathcal{F}$

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

**Corollary:** For any $x$ and $y$ from $\mathcal{F}$,

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

# Cubic regularization

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

$$T_M(x) \in \mathrm{Arg} \min_y \; \hat{f}_M(x,y),$$

where "Arg" indicates that $T_M(x)$ is a *global* minimum.

**Computability:** If $\| \cdot \|$ is a Euclidean norm, then $T_M(x)$ can be computed from a *convex* problem.

For $r \in \mathcal{D} \equiv \{ r \in R : f''(x) + \frac{M}{2} r I \succ 0, r \geq 0 \}$, denote
$v(r) \; = \; -\frac{1}{2} \langle (f''(x) + \frac{Mr}{2} I)^{-1} f'(x), f'(x) \rangle - \frac{M}{12} r^3$.

**Lemma.** For $M > 0$, $\quad \min\limits_{h \in R^n} \hat{f}_M(x, x+h) \; = \; \sup\limits_{r \in \mathcal{D}} v(r)$.

If the *sup* is attained at $r^* : f''(x) + \frac{Mr^*}{2} I \succ 0$, then
$h^* = -(f''(x) + \frac{Mr^*}{2} I)^{-1} f'(x)$
where $r^* > 0$ is a unique solution to $r = \| (f''(x) + \frac{Mr}{2} I)^{-1} f'(x) \|$.

# Simple properties

**1.** Denote $r_M(x) = \|x - T_M(x)\|$. Then

$$f'(x) + f''(x)(T_M(x) - x) + \frac{Mr_M(x)}{2}(T_M(x) - x) = 0,$$

$$f''(x) + \tfrac{1}{2} M r_M(x) I \succeq 0.$$

**2.** We have $\langle f'(x), x - T_M(x) \rangle \geq 0$, and

$$f(x) - \bar{f}_M(x) \geq \frac{M}{12} r_M^3(x),$$

$$r_M^2(x) \geq \frac{2}{L+M} \|f'(x)\|.$$

**3.** If $M \geq L$ then $\bar{f}_M(x) \geq f(T_M(x))$.

**4.** $\bar{f}_M(x) \leq \min\limits_y \left[ f(y) + \frac{L+M}{6} \|y - x\|^3 \right]$.

Compare with *prox-method*: $x_+ = \min\limits_y [f(y) + \frac{1}{2} M \|y - x\|^2]$.

# Cubic regularization of Newton method

Consider the process: $x_{k+1} = T_L(x_k), \quad k = 0, 1, \dots$.
Note that $f(x_{k+1}) \leq f(x_k)$.

**Saddle points.** Let $f'(x^*) = 0$ and $f''(x^*) \not\succeq 0$. Then $\exists \epsilon, \delta > 0$ such that

$$\boxed{\|x - x^*\| \leq \epsilon, \; f(x) \geq f(x^*)} \Rightarrow \boxed{f(T_L(x)) \leq f(x^*) - \delta}$$

**Local convergence.** If $\mathcal{L}(x_0)$ is bounded, then
$X^* \equiv \lim\limits_{k \to \infty} \{x_k\} \neq \emptyset$.
For any $x^* \in X^*$ we have $f(x^*) = f^*$, $f'(x^*) = 0$, $f''(x^*) \succeq 0$.

**Global convergence:** $g_k \equiv \min\limits_{1 \leq i \leq k} \|f'(x_i)\| \leq O\left(\frac{1}{k^{2/3}}\right)$.

For gradient method we can guarantee only $g_k \leq O\left(\frac{1}{k^{1/2}}\right)$.

**Local rate of convergence:** Quadratic.

# Global performance: Star-convex functions

**Def.** For any $x^* \in X^*$ and any $x \in \mathcal{F}$, $\alpha \in [0,1]$ we have $f(\alpha x^* + (1-\alpha)x) \leq \alpha f(x^*) + (1-\alpha) f(x)$.

**Th 1.** Let diam $\mathcal{F} \leq D$. Then

1. If $f(x_0) - f^* \geq \frac{3}{2} L D^3$, then $f(x_1) - f^* \leq \frac{1}{2} L D^3$.
2. If $f(x_0) - f^* \leq \frac{3}{2} L D^3$, then $f(x_k) - f^* \leq \frac{3LD^3}{2(1+\frac{1}{3}k)^2}$.

Let $X^*$ be non-degenerate: $f(x) - f^* \geq \frac{\gamma}{2} \rho^2(x, X^*)$. Denote $\bar{\omega} = \frac{1}{L^2}(\frac{\gamma}{2})^3$.

**Th 2.** Denote $k_0$ the first number for which $f(x_{k_0}) - f^* \leq \frac{4}{9}\bar{\omega}$.

If $k \leq k_0$, then $f(x_k) - f^* \leq \left[ (f(x_0) - f^*)^{1/4} - \frac{k}{6}\sqrt{\frac{2}{3}}\bar{\omega}^{1/4} \right]^4$.

For $k \geq k_0$ we have $f(x_{k+1}) - f^* \leq \frac{1}{2}(f(x_k) - f^*)\sqrt{\frac{f(x_k)-f^*}{\bar{\omega}}}$.

**NB** The Hessian $f''(x^*)$ can be degenerate!

# Global performance: Gradient-dominated functions

**Definition.** For any $x \in \mathcal{F}$ and $x^* \in X^*$ we have

$$f(x) - f(x^*) \le \tau_f \|f'(x)\|^p$$

with $\tau_f > 0$ and $p \in [1,2]$ (*degree of domination*).

**Example 1.** *Convex functions*:

$$f(x) - f^* \le \langle f'(x), x - x^* \rangle \le R\|f'(x)\|$$

for $\|x - x^*\| \le R$. Thus, $p = 1$, $\tau_f = \frac{1}{2}D$.

**Example 2.** *Strongly convex functions*: $\forall x, y \in R^n$

$$f(x) \le f(y) + \langle f'(y), x - y \rangle + \frac{1}{2\gamma}\|f'(x) - f'(y)\|^2.$$

Thus, $f(x) - f^* \le \frac{1}{2\gamma}\|f'(x)\|^2 \quad \Rightarrow \quad p = 2,\ \tau_f = \frac{1}{2\gamma}$.

# Gradient dominated functions, II

**Example 3.** *Sum of squares.* Consider the system

$$g(x) = 0 \in R^m, \quad x \in R^n.$$

Assume that $m \leq n$ and the Jacobian $J(x) = (g_1'(x), \ldots, g_m'(x))$ is *uniformly non-degenerate*:

$$\sigma \equiv \inf_{x \in \mathcal{F}} \lambda_{\min}(J^T(x)J(x)) > 0.$$

Consider the function $f(x) = \sum\limits_{i=1}^{m} g_i^2(x)$. Then

$$f(x) - f^* \leq \frac{1}{2\sigma} \|f'(x)\|^2.$$

Thus, $p = 2$ and $\tau_f = \frac{1}{2\sigma}$.

# Gradient dominated functions: rate of convergence

**Theorem 3.** Let $p = 1$. Denote $\hat{\omega} = \frac{2}{3}L(6\tau_f)^3$. Let $k_0$ be defined as $f(x_{k_0}) - f^* \leq \xi^2\hat{\omega}$ for some $\xi > 1$. Then for $k \leq k_0$ we have

$$\ln\left(\tfrac{1}{\hat{\omega}}(f(x_k) - f^*)\right) \leq \left(\tfrac{2}{3}\right)^k \ln\left(\tfrac{1}{\hat{\omega}}(f(x_0) - f^*)\right).$$

Otherwise, $f(x_k) - f^* \leq \hat{\omega} \cdot \frac{\xi^2(2+\frac{3}{2}\xi)^2}{(2+(k+\frac{3}{2})\cdot\xi)^2}$.

**Theorem 4.** Let $p = 2$. Denote $\tilde{\omega} = \frac{1}{(144L)^2\tau_f^3}$. Let $k_0$ be defined as $f(x_{k_0}) - f^* \leq \tilde{\omega}$. Then for $k \leq k_0$ we have

$$f(x_k) - f^* \leq (f(x_0) - f^*) \cdot e^{-k\sigma}$$

with $\sigma = \frac{\tilde{\omega}^{1/4}}{\tilde{\omega}^{1/4}+(f(x_0)-f^*)^{1/4}}$. Otherwise,

$$f(x_{k+1}) - f^* \leq \tilde{\omega} \cdot \left(\frac{f(x_k)-f^*}{\tilde{\omega}}\right)^{4/3}.$$

**NB:** Superlinear convergence without direct nondegeneracy assumption for the Hessian.

# Transformations of convex functions

Let $u(x) : R^n \to R^n$ be non-degenerate. Denote by $v(u)$ its inverse: $v(u(x)) \equiv x$.

Consider the function $f(x) = \phi(u(x))$, where $\phi(u)$ is a convex function. Denote

$$\begin{aligned} \sigma &= \max_u \{\|v'(u)\| : \phi(u) \leq f(x_0)\}, \\ D &= \max_u \{\|u - u^*\| : \phi(u) \leq f(x_0)\}. \end{aligned}$$

**Theorem 5.**

1. If $f(x_0) - f^* \geq \frac{3}{2}L(\sigma D)^3$, then $f(x_1) - f^* \leq \frac{1}{2}L(\sigma D)^3$.
2. If $f(x_0) - f^* \leq \frac{3}{2}L(\sigma D)^3$, then $f(x_k) - f^* \leq \frac{3L(\sigma D)^3}{2(1+\frac{1}{3}k)^2}$.

**Example.**

$$\begin{aligned} u_1(x) &= x_1, \quad u_2(x) = x_2 + \phi_1(x_1), \ \ldots \\ u_n(x) &= x_n + \phi_{n-1}(x_1, \ldots, x_{n-1}), \end{aligned}$$

where $\phi_i(\cdot)$ are *arbitrary* functions.

# Accelerated Newton: Cubic prox-function

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

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

$$\begin{aligned} \langle d'(x) - d'(y), x - y \rangle &\geq \tfrac{1}{2}\|x - y\|^3, \\ d(x) - d(y) - \langle d'(y), x - y \rangle &\geq \tfrac{1}{6}\|x - y\|^3, \end{aligned}$$

Moreover, its Hessian is Lipschitz continuous:

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

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

# Linear estimate functions (Compare with 1st-order methods)

We recursively update the following sequences.

- Sequence of estimate functions $\psi_k(x) = l_k(x) + \frac{N}{2} d(x)$, $k \geq 1$, where $l_k(x)$ are linear, and $N > 0$.
- A minimizing sequence $\{x_k\}_{k=1}^{\infty}$.
- A sequence of scaling parameters $\{A_k\}_{k=1}^{\infty}$: $A_{k+1} \stackrel{\mathrm{def}}{=} A_k + a_k$, $k \geq 1$.

These objects have to satisfy the following relations:

$$(*): \quad \begin{array}{rcl} A_k f(x_k) & \leq & \psi_k^* \equiv \min\limits_x \psi_k(x), \\ \psi_k(x) & \leq & A_k f(x) + (L_2 + \frac{1}{2}N) d(x), \; \forall x \in R^n, \end{array}$$

for all $k \geq 1$. $\left(\Rightarrow A_k(f(x_k) - f(x^*)) \leq (L + \frac{N}{2}) d(x^*).\right)$

For $k = 1$, we can choose $x_1 = T_{L_2}(x_0)$, $l_1(x) \equiv f(x_1)$, $A_1 = 1$.

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

For some $a_k > 0$ and $M \geq 2L_2$, define

$$\alpha_k \;=\; \tfrac{a_k}{A_k + a_k} \in (0,1),$$

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

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

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

**Theorem.** For $M = 2L_2$, $N = 12L_2$, and $a_k = \frac{(k+1)(k+2)}{2}$, $k \geq 1$, relations (*) hold recursively.

**Corollary.** For any $k \geq 1$ we have $f(x_k) - f(x^*) \;\leq\; \frac{14L_2\|x_0 - x^*\|^3}{k(k+1)(k+2)}$.

# Accelerated CNM

**Initialization:** Set $x_1 = T_{L_2}(x_0)$. Define $\psi_1(x) = f(x_1) + 6L_2 \cdot d(x)$.

**Iteration $k$, ($k \geq 1$):** $v_k = \arg\min\limits_{x \in R^n} \psi_k(x)$,

$y_k = \frac{k}{k+3} x_k + \frac{3}{k+3} v_k$, $x_{k+1} = T_{2L_2}(y_k)$,

$$\psi_{k+1}(x) = \psi_k(x) + \tfrac{(k+1)(k+2)}{2}[f(x_{k+1}) + \langle f'(x_{k+1}), x - x_{k+1} \rangle]$$

**Remark:**

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

$$s_1 = 0, \quad s_{k+1} = s_k + \tfrac{(k+1)(k+2)}{2} f'(x_{k+1}), \quad k \geq 1.$$

Then $v_k$ can be computed by an explicit expression.

# Global non-degeneracy

**Standard setting:** for convex $f \in C^2(R^n)$ define positive constants $\sigma_1$ and $L_1$ such that

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

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

(Compatible with definition in Linear Algebra.)

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

**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: $f(x_k) - f^* \;\leq\; \frac{14\,\mathbf{L}_2\,R^3}{k(k+1)(k+2)}$.

# Global 2nd-order non-degeneracy

**Assumption:** for any $x, y \in R^n$, function $f \in C^2(R^n)$ 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:** $\frac{\sigma_2}{3}\|x_k - x^*\|^3 \leq f(x_k) - f^* \leq \frac{14L_2\|x_0 - x^*\|^3}{k(k+1)(k+2)}$.

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)

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

# Open questions

**1.** Problem classes.

**2.** Lower complexity bounds and optimal methods.

**3.** Non-degenerate problems: geometric interpretation?

**4.** Complexity of strongly convex functions.
(1st-order schemes?)

**5.** Consequences for polynomial-time methods.

# Solving the systems of nonlinear equations

**1. Standard Gauss-Newton method**

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

**Assumption:** $\quad \forall x, y \in R^n \quad \|F'(x) - F'(y)\| \leq L\|x - y\|$.

**Gauss-Newton method:** Choose a merit function $\phi(u) \geq 0$, $\phi(0) = 0$, $u \in R^m$.

Compute $x_+ \in \mathrm{Arg}\min\limits_{y} [\phi(F(x) + F'(x)(y - x))]$.

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

**Remarks**

- Local quadratic convergence ($m \geq n$, 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 $x, y \in R^n$ we have

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

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

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

**Modified method:**

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

**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 $R^n$ and $R^m$ can be used.

# Testing CNM: Chebyshev oscilator

Consider $f(x) = \frac{1}{4}(1 - x^{(1)})^2 + \sum_{i=1}^{n-1} \left(x^{(i+1)} - p_2(x^{(i)})\right)^2$, with $p_2(\tau) = 2\tau^2 - 1$.

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

Hence, the equations for the "central path" is

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

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

**NB:** $f(x)$ 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), $T(n) = Const * (4.5)^n$.