16. Newton’s method

- Kantorovich theorem
- inexact Newton method
Newton’s method for nonlinear equations

Newton iteration for solving a nonlinear equation \( f(x) = 0 \):

\[
x^{(k+1)} = x^{(k)} - f'(x^{(k)})^{-1} f(x^{(k)})
\]

- \( f : \mathbb{R}^n \to \mathbb{R}^n \) is a vector valued function \( f(x) = (f_1(x), \ldots, f_n(x)) \)
- \( f'(x) \) is the \( n \times n \) Jacobian matrix at \( x \):

\[
(f'(x))_{ij} = \frac{\partial f_i}{\partial x_j}(x), \quad i, j = 1, \ldots, n
\]

- \( x^{(k+1)} \) is the solution of the linearized equation at \( x^{(k)} \):

\[
f(x^{(k)}) + f'(x^{(k)})(x - x^{(k)}) = 0
\]

we denote the iterates by the simpler notation \( x_k = x^{(k)} \) if the meaning is clear
Matrix norm

in this lecture, *operator norms* are used for square matrices:

\[
\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}
\]

the same (arbitrary) vector norm is used for \(\|Ax\|\) and \(\|x\|\)

**Properties** (\(A, B\) are \(n \times n\) matrices and \(x\) is an \(n\)-vector)

- *identity matrix:* \(\|I\| = 1\)
- *matrix-vector product:* \(\|Ax\| \leq \|A\|\|x\|\)
- *submultiplicative property:* \(\|AB\| \leq \|A\|\|B\|\)
- *perturbation lemma:* if \(A\) is invertible and \(\|A^{-1}B\| < 1\), then \(A + B\) is invertible,

\[
\|(A + B)^{-1}\| \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}B\|}
\]
Proof of perturbation lemma

• $A + B$ is invertible: if $(A + B)x = 0$, then

$$
\|x\| = \|A^{-1}Bx\| \leq \|A^{-1}B\||x|
$$

if $\|A^{-1}B\| < 1$ this is only possible if $x = 0$

• $Y = (A + B)^{-1}$ satisfies $(I + A^{-1}B)Y = A^{-1}$; therefore

$$
\|Y\| = \|A^{-1} - A^{-1}BY\|
\leq \|A^{-1}\| + \|A^{-1}BY\|
\leq \|A^{-1}\| + \|A^{-1}B\||Y|
$$

from which the inequality in the lemma follows
Outline

- Kantorovich theorem
- inexact Newton method
Assumptions in Kantorovich theorem

- \( f : \mathbb{R}^n \rightarrow \mathbb{R}^n \) is continuously differentiable on an open convex set \( D \)
- the Jacobian matrix \( f'(x_0) \) at the starting point \( x_0 \in D \) is invertible
- the Jacobian is Lipschitz continuous on \( D \): there exists a positive \( \gamma \) such that
  \[
  \|f'(x_0)^{-1}(f'(x) - f'(y))\| \leq \gamma \|x - y\| \quad \text{for all } x, y \in D
  \]
- the norm \( \eta := \|f'(x_0)^{-1}f(x_0)\| \) of the first Newton step is bounded by
  \[
  \eta \gamma \leq \frac{1}{2}
  \]
- \( D \) contains the ball \( B(x_0, r) = \{ x \mid \|x - x_0\| \leq r \} \), where
  \[
  r = \frac{1 - \sqrt{1 - 2\gamma \eta}}{\gamma}
  \]
Kantorovich theorem

\[ x_{k+1} = x_k - f'(x_k)^{-1} f(x_k), \quad k = 0, 1, \ldots \]

under the assumptions on the previous page:

• the iteration is well defined, i.e., the Jacobian matrices \( f''(x_k) \) are invertible

• the iterates remain in \( B(x_0, r) \)

• the iterates converge to a solution \( x^* \) of \( f(x) = 0 \)

• the following error bound holds:

\[ \|x_k - x^*\| \leq \frac{(2\gamma \eta)^{2^k-1}}{2^{k-1}} \eta \]
Comments

• this is the affine-invariant version of the theorem: invariant under transformation

\[ \tilde{f}(x) = Af(x), \quad A \text{ nonsingular} \]

• existence of a solution is not assumed but follows from the theorem

• complete theorem includes uniqueness of solution in a larger region

• theorem explains very fast local convergence; for example, if \( \gamma \eta = 0.4 \)

<table>
<thead>
<tr>
<th>( k )</th>
<th>((2\gamma \eta)^{2^k-1}/2^{k-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.0000000000000000</td>
</tr>
<tr>
<td>1</td>
<td>0.8000000000000000</td>
</tr>
<tr>
<td>2</td>
<td>0.2560000000000000</td>
</tr>
<tr>
<td>3</td>
<td>0.0524288000000000</td>
</tr>
<tr>
<td>4</td>
<td>0.0043980465111040</td>
</tr>
<tr>
<td>5</td>
<td>0.0000618970019643</td>
</tr>
<tr>
<td>6</td>
<td>0.000000245199287</td>
</tr>
<tr>
<td>7</td>
<td>0.0000000000000077</td>
</tr>
</tbody>
</table>

Newton's method 16.7
Newton method for quadratic scalar equation

we first examine the convergence of Newton’s method applied to \( g(t) = 0 \), where

\[
g(t) = \frac{\gamma}{2} t^2 - t + \eta \quad \text{with } \gamma > 0, \eta > 0, \ h := \gamma \eta \leq \frac{1}{2}
\]

- the roots will be denoted by \( t^\star \) and \( t^{**} \)

\[
t^\star = \frac{1 - \sqrt{1 - 2h}}{\gamma}, \quad t^{**} = \frac{1 + \sqrt{1 - 2h}}{\gamma}
\]

- the Newton iteration started at \( t_0 = 0 \) is

\[
t_{k+1} = \frac{(\gamma/2)t_k^2 - \eta}{\gamma t_k - 1}
\]
Newton iteration can be written as $t_{k+1} = \psi(t_k)$ where

$$\psi(t) = \frac{(\gamma/2)t^2 - \eta}{\gamma t - 1}$$

for $\gamma = 1$, $\eta = h = 3/8$ and $\gamma = 1$, $\eta = h = 1/2$. 

Newton's method 16.9
Recursions

to derive simple error bounds, we define $g_k(\tau)$ as $g(t)$ scaled and centered at $t_k$:

$$g_k(\tau) = \frac{g(\tau + t_k)}{-g'(t_k)} = \frac{\gamma_k}{2}\tau^2 - \tau + \eta_k, \quad k = 0, 1, \ldots$$

- coefficients $\gamma_k$, $\eta_k$, and $h_k = \gamma_k \eta_k$ satisfy the recursions (see next page)

$$
\gamma_{k+1} = \frac{\gamma_k}{1 - h_k}, \quad \eta_{k+1} = \frac{h_k \eta_k}{2(1 - h_k)}, \quad h_{k+1} = \frac{h_k^2}{2(1 - h_k)^2}
$$

with $\gamma_0 = 0$, $\eta_0 = \eta$, $h_0 = \gamma \eta$

- we denote the smallest root of $g_k$ by $r_k$:

$$r_k = t^* - t_k = \frac{1 - \sqrt{1 - 2h_k}}{\gamma_k} = \frac{2\eta_k}{1 + \sqrt{1 - 2h_k}}$$

- Newton step for $g_k$ at $\tau = 0$ is equal to Newton step for $g$ at $t = t_k$:

$$\eta_k = t_{k+1} - t_k$$
Proof of recursions

• since \( g \) is quadratic,

\[
\frac{\gamma_k \tau^2 - \tau + \eta_k}{2} = \frac{g(t_k + \tau)}{-g'(t_k)} = \frac{(g''(t_k)/2) \tau^2 + g'(t_k) \tau + g(t_k)}{-g'(t_k)}
\]

• recursion for \( \gamma_k \):

\[
\gamma_{k+1} = \frac{g''(t_{k+1})}{-g'(t_{k+1} + \eta_k)} = \frac{g''(t_k)}{-g'(t_k) - g''(t_k) \eta_k} = \frac{\gamma_k}{1 - \gamma_k \eta_k} = \frac{\gamma_k}{1 - h_k}
\]

• recursion for \( \eta_k \):

\[
\eta_{k+1} = \frac{g(t_k + \eta_k)}{-g'(t_k + \eta_k)} = \frac{(g''(t_k)/2) \eta_k^2}{-g'(t_k) - g''(t_k) \eta_k} = \frac{\gamma_k \eta_k^2}{2(1 - \gamma_k \eta_k)} = \frac{h_k \eta_k}{2(1 - h_k)}
\]

• recursion for \( h_k \) follows from \( h_{k+1} = \gamma_{k+1} \eta_{k+1} \)
Error bounds

• Newton step $\eta_k = t_{k+1} - t_k$:

$$
\eta_k \leq \frac{(2h)^{2^k-1}}{2^k} \eta
$$

(see next page)

• error $r_k = t^* - t_k$:

$$
r_k = \frac{2\eta_k}{1 + \sqrt{1 - 2h_k}} \leq 2\eta_k \leq \frac{(2h)^{2^k-1}}{2^{k-1}} \eta
$$
Proof of bound on $\eta_k$

- since $h_0 \leq 1/2$ the recursion for $h_k$ shows that

$$2h_k = \frac{h_{k-1}^2}{(1 - h_{k-1})^2} \leq (2h_{k-1})^2$$

- applying this recursively we obtain $2h_k \leq (2h_0)^{2^k}$

- from the recursion for $\eta_k$ (and $h_k \leq 1/2$):

$$\eta_k = \frac{h_{k-1}\eta_{k-1}}{2(1 - h_{k-1})} \leq h_{k-1}\eta_{k-1}$$

- applying this recursively and using the bound on $h_k$ we obtain the bound on $\eta_k$:

$$\eta_k \leq h_{k-1} \cdots h_1 h_0 \eta_0$$

$$= 2^{-k} (2h_0)^{2^{k-1}} (2h_0)^{2^{k-2}} \cdots (2h_0)^2 (2h_0) \eta_0$$

$$= 2^{-k} (2h_0)^{2^k-1} \eta_0$$
Summary of proof of Kantorovich theorem

to prove the Kantorovich theorem (pp. 16.5–16.6), we show that

$$\|x_{k+1} - x_k\| \leq t_{k+1} - t_k$$

where $t_k$ are the iterates in Newton’s method, started at $t_0 = 0$, for

$$\frac{\gamma}{2} t^2 - t + \eta = 0$$

- $t_k$ is called a majorizing sequence for the sequence $x_k$
- the bounds for $t_k$ on page 16.12 provide bounds and convergence results for $x_k$
Consequences of majorization

\[ t_0 = 0, \quad \|x_{k+1} - x_k\| \leq t_{k+1} - t_k \quad \text{for } k \geq 0 \]

- by the triangle inequality, if \( k \geq j \),

\[ \|x_k - x_j\| \leq \sum_{i=j}^{k-1} \|x_{i+1} - x_i\| \leq \sum_{i=j}^{k-1} (t_{i+1} - t_i) = t_k - t_j \]

- the inequality shows that \( x_k \) is a Cauchy sequence, so it converges

- taking \( j = 0 \) shows that \( x_k \) remains in the set \( B(x_0, r) \) (defined on page 16.5):

\[ \|x_k - x_0\| \leq t_k - t_0 \leq t^* = r \]

- taking limits for \( k \to \infty \) shows the error bound on page 16.6:

\[ \|x^* - x_j\| \leq t^* - t_j = r_j \leq \frac{(2h)^{2j-1}}{2^{j-1} \eta} \]
Details of proof of Kantorovich theorem

we prove that the following inequalities hold for \( k = 0, 1, \ldots \)

\[
\|f'(x_{k+1})^{-1} f'(x_k)\| \leq \frac{1}{1 - h_k} \tag{1}
\]

\[
\|f'(x_k)^{-1} (f'(x) - f'(y))\| \leq \gamma_k \|x - y\| \quad \text{for all } x, y \in D \tag{2}
\]

\[
\|f'(x_k)^{-1} f(x_k)\| \leq \eta_k \tag{3}
\]

\[
B(x_{k+1}, r_{k+1}) \subseteq B(x_k, r_k) \tag{4}
\]

- \( \gamma_k, \eta_k, h_k, r_k \) are the sequences defined on page 16.10
- for \( k = 0 \), inequalities (2) and (3) hold by assumption, since \( \gamma_0 = \gamma, \eta_0 = \eta \)
- (3) is the majorization inequality \( \|x_{k+1} - x_k\| \leq \eta_k = t_{k+1} - t_k \)
Proof by induction: suppose (2) and (3) holds at $k = i$, and (4) holds for $k < i$

- $x_{i+1} \in D$ because $B(x_i, r_i) \subseteq \cdots \subseteq B(x_0, r_0) \subseteq D$ and $\|x_{i+1} - x_i\| \leq \eta_i \leq r_i$

- the inequality (2) at $k = i$ implies that

$$\| f'(x_i)^{-1} f'(x_{i+1}) - I \| = \| f'(x_i)^{-1} (f'(x_{i+1}) - f'(x_i)) \|$$

$$\leq \gamma_i \| x_{i+1} - x_i \|$$

$$\leq \gamma_i \eta_i$$

$$\leq h_i$$

invertibility of $f'(x_{i+1})$ and (1) at $k = i$ follow from the perturbation lemma

- inequality (2) at $k = i + 1$ follows from (1) and (2) at $k = i$:

$$\| f'(x_{i+1})^{-1} (f'(x) - f'(y)) \| \leq \| f'(x_{i+1})^{-1} f'(x) \| \| f'(x_i)^{-1} (f'(x) - f'(y)) \|$$

$$\leq \frac{\gamma_i}{1 - h_i} \| x - y \|$$

$$= \gamma_{i+1} \| x - y \|$$
inequality (3) at \( k = i + 1 \) follows from (2) at \( k = i + 1 \): define \( v = x_{i+1} - x_i \),

\[
\| f'(x_{i+1})^{-1} f(x_{i+1}) \| = \left\| f'(x_{i+1})^{-1} \left( \int_0^1 f'(x_i + tv) v dt + f(x_i) \right) \right\|
\]

\[
= \left\| f'(x_{i+1})^{-1} \int_0^1 (f'(x_i + tv) - f'(x_i)) v dt \right\|
\]

\[
\leq \|v\| \int_0^1 \left\| f'(x_{i+1})^{-1} (f'(x_i + tv) - f'(x_i)) \right\| dt
\]

\[
\leq \frac{\gamma_{i+1}}{2} \|v\|^2
\]

\[
\leq \frac{\gamma_{i+1} \eta_i^2}{2}
\]

\[
= \eta_{i+1}
\]

inequality (4) at \( k = i \) follows from (3) at \( k = i + 1 \) and \( r_i = r_{i+1} + \eta_i \)

\[
\|x - x_{i+1}\| \leq r_{i+1} \implies \|x - x_i\| \leq \|x - x_{i+1}\| + \|x_{i+1} - x_i\| \leq r_{i+1} + \eta_i = r_i
\]
it remains to show that the limit \( x^* \) solves the equation

- by the assumptions on page 16.5,

\[
\| f'(x_0)^{-1} f(x_k) \| = \| f'(x_0)^{-1} f'(x_k) (x_{k+1} - x_k) \| \\
= \| (f'(x_0)^{-1} (f'(x_k) - f'(x_0)) + I) (x_{k+1} - x_k) \| \\
\leq \left( \| (f'(x_0)^{-1} (f'(x_k) - f'(x_0)) \| + 1 \right) \| x_{k+1} - x_k \| \\
\leq (\gamma r + 1) \| x_{k+1} - x_k \|
\]

- since \( \| x_{k+1} - x_k \| \to 0 \) and \( f \) is continuous,

\[
\| f'(x_0)^{-1} f(x^*) \| = \lim_{k \to \infty} \| f'(x_0)^{-1} f(x_k) \| = 0
\]
Outline

- Kantorovich theorem
- inexact Newton method
Inexact Newton method

inexact Newton method for solving nonlinear equation \( f(x) = 0 \):

\[
x_{k+1} = x_k + s_k \quad \text{where } s_k \approx -f'(x_k)^{-1} f(x_k)
\]

• \( s_k \) is an approximate solution of the Newton equation

\[
f'(x_k)s = -f(x_k)
\]

• goal is to reduce cost per iteration while retaining fast convergence

• in Newton-iterative methods, Newton equation is solved by iterative method

• an example is the Newton-CG method if \( f'(x) \) is symmetric positive definite
Forcing condition

accept the inexact Newton step \( s_k \) if

\[
\| f'(x_k) s_k + f(x_k) \| \leq \alpha_k \| f(x_k) \|
\]

- coefficient \( \alpha_k < 1 \) is called the **forcing term**

- \( \alpha_k \) is limit on relative error in the Newton equation

- provides a stopping condition in iterative method for solving Newton equation

- \( \alpha_k \) is constant or adjusted adaptively

Newton’s method
Local convergence

Assumptions

- the equation has a solution $x^*$ and $f'(x^*)$ is invertible
- $f'$ is Lipschitz continuous in a neighborhood of $x^*$

Local convergence result

- the iterates $x_k$ converge to $x^*$ if $x_0$ is sufficiently close to $x^*$
- an error bound of following type holds (for some $\kappa$ with $\kappa \bar{\alpha} < 1$ if $\alpha_k \leq \bar{\alpha} < 1$):

$$
\|x_{k+1} - x^*\| \leq \kappa \left( \|x_k - x^*\| + \alpha_k \right) \|x_k - x^*\|
$$

- this shows how the forcing term determines the rate of convergence

<table>
<thead>
<tr>
<th>$\alpha_k$ constant</th>
<th>$\alpha_k = 0$</th>
<th>$\alpha_k \searrow 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>convergence:</td>
<td>linear</td>
<td>quadratic</td>
</tr>
</tbody>
</table>
References

Newton method


Kantorovich theorem: the statement and proof of the theorem in the lecture follow


Inexact Newton method