16. Newton's method

- Kantorovich theorem
- inexact Newton method

Newton's method for nonlinear equations

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

$$x^{(k+1)} = x^{(k)} - f'(x^{(k)})^{-1} f(x^{(k)})$$

- $f : \mathbf{R}^n \to \mathbf{R}^n$ 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

$$\begin{aligned}
\|Y\| &= \|A^{-1} - A^{-1}BY\| \\
&\leq \|A^{-1}\| + \|A^{-1}BY\| \\
&\leq \|A^{-1}\| + \|A^{-1}B\|\|Y\|
\end{aligned}$$

  from which the inequality in the lemma follows

Outline

- **Kantorovich theorem**
- inexact Newton method

Assumptions in Kantorovich theorem

- $f : \mathbf{R}^n \to \mathbf{R}^n$ 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 γ 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 $x^\star$ of $f(x) = 0$
- the following error bound holds:

$$\|x_k - x^\star\| \le \frac{(2\gamma\eta)^{2^k-1}}{2^{k-1}}\eta$$

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 $\gamma\eta = 0.4$

k	$(2\gamma\eta)^{2^k-1}/2^{k-1}$
0	2.000000000000000
1	0.800000000000000
2	0.256000000000000
3	0.052428800000000
4	0.004398046511104
5	0.000061897001964
6	0.000000245199287
7	0.000000000000077

Newton method for quadratic scalar equation

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

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

- the roots will be denoted by $t^\star$ and $t^{\star\star}$

$$t^\star = \frac{1 - \sqrt{1 - 2h}}{\gamma}, \qquad t^{\star\star} = \frac{1 + \sqrt{1 - 2h}}{\gamma}$$

- the Newton iteration started at $t_0 = 0$ is

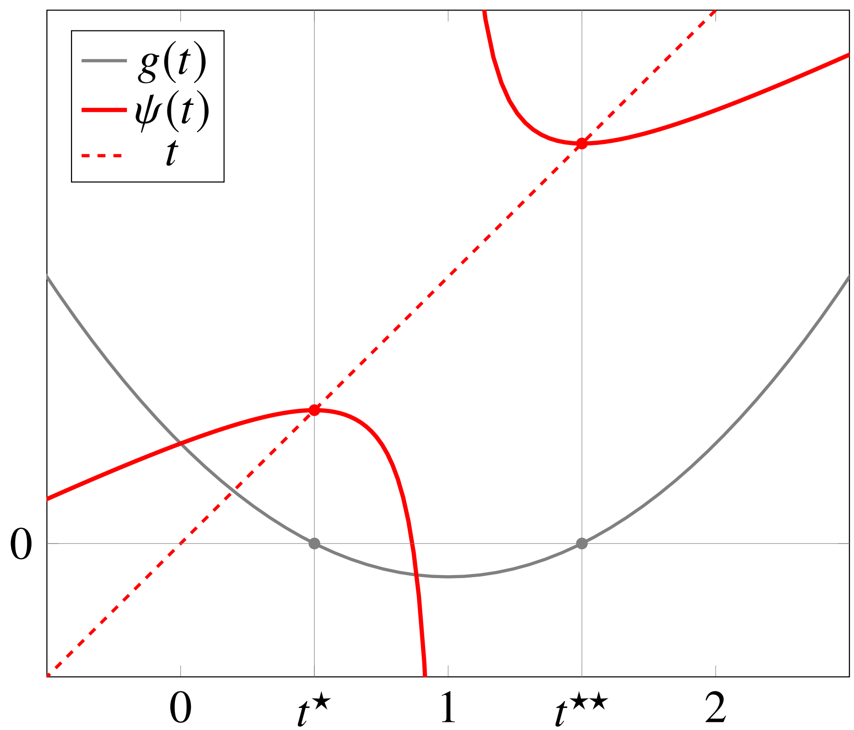
$$t_{k+1} = \frac{(\gamma/2)t_k^2 - \eta}{\gamma t_k - 1}$$

Iteration map

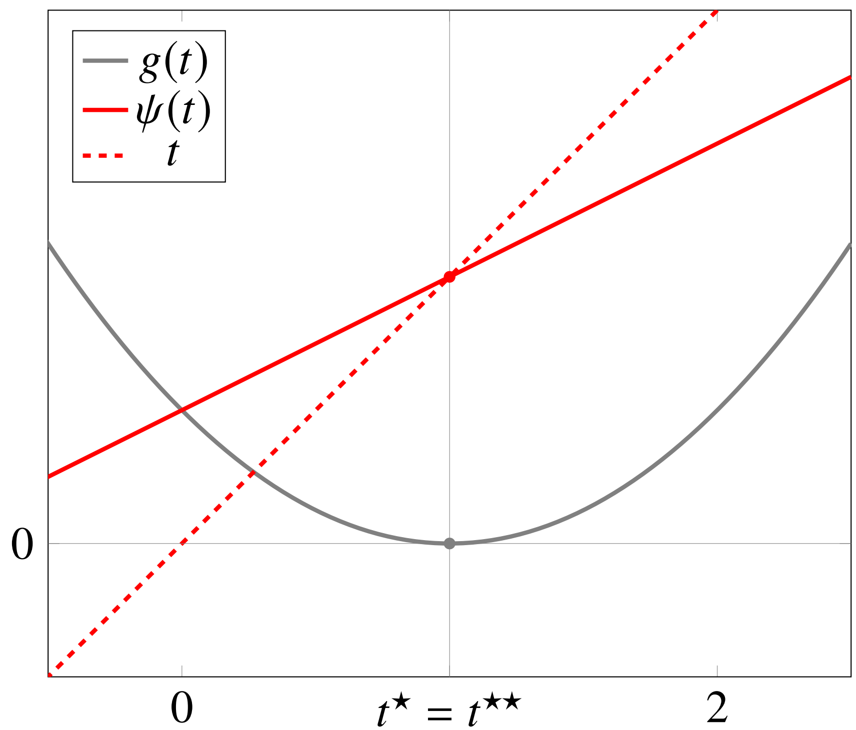
Newton iteration can be written as $t_{k+1} = \psi(t_k)$ where

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

$\gamma = 1, \eta = h = 3/8$



$\gamma = 1, \eta = h = 1/2$



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

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

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

- we denote the smallest root of g_k by r_k:

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

- 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,

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

- recursion for γ_k:

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

- recursion for η_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 $\eta_k = t_{k+1} - t_k$:

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

(see next page)

- error $r_k = t^\star - t_k$:

$$r_k = \frac{2\eta_k}{1+\sqrt{1-2h_k}} \le 2\eta_k \le \frac{(2h)^{2^k-1}}{2^{k-1}}\eta$$

Proof of bound on η_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 η_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 η_k:

$$\begin{aligned} \eta_k &\leq h_{k-1} \cdots h_1 h_0 \eta_0 \\ &= 2^{-k}(2h_0)^{2^{k-1}}(2h_0)^{2^{k-2}} \cdots (2h_0)^2(2h_0)\eta_0 \\ &= 2^{-k}(2h_0)^{2^k-1}\eta_0 \end{aligned}$$

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

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

- 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):

$$\|x_k - x_0\| \leq t_k - t_0 \leq t^\star = r$$

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

$$\|x^\star - x_j\| \leq t^\star - t_j = r_j \leq \frac{(2h)^{2^j-1}}{2^{j-1}}\eta$$

Details of proof of Kantorovich theorem

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

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

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

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

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

- γ_k, η_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

$$\begin{aligned}
\|f'(x_i)^{-1} f'(x_{i+1}) - I\| &= \|f'(x_i)^{-1}(f'(x_{i+1}) - f'(x_i))\| \\
&\leq \gamma_i \|x_{i+1} - x_i\| \\
&\leq \gamma_i \eta_i \\
&\leq h_i
\end{aligned}$$

  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 $k = i$:

$$\begin{aligned}
\|f'(x_{i+1})^{-1}(f'(x) - f'(y))\| &\leq \|f'(x_{i+1})^{-1} f'(x_i)\| \|f'(x_i)^{-1}(f'(x) - f'(y))\| \\
&\leq \frac{\gamma_i}{1 - h_i} \|x - y\| \\
&= \gamma_{i+1} \|x - y\|
\end{aligned}$$

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

$$
\begin{aligned}
\|f'(x_{i+1})^{-1} f(x_{i+1})\| &= \left\| f'(x_{i+1})^{-1} \left(\int_0^1 f'(x_i + tv) v\, dt + f(x_i) \right) \right\| \\
&= \left\| f'(x_{i+1})^{-1} \int_0^1 \left(f'(x_i + tv) - f'(x_i) \right) v\, dt \right\| \\
&\le \|v\| \int_0^1 \left\| f'(x_{i+1})^{-1} (f'(x_i + tv) - f'(x_i)) \right\| dt \\
&\le \frac{\gamma_{i+1}}{2} \|v\|^2 \\
&\le \frac{\gamma_{i+1} \eta_i^2}{2} \\
&= \eta_{i+1}
\end{aligned}
$$

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

$$
\|x - x_{i+1}\| \le r_{i+1} \quad \Longrightarrow \quad \|x - x_i\| \le \|x - x_{i+1}\| + \|x_{i+1} - x_i\| \le r_{i+1} + \eta_i = r_i
$$

Limit

it remains to show that the limit $x^\star$ solves the equation

- by the assumptions on page 16.5,

$$\begin{aligned}
\|f'(x_0)^{-1}f(x_k)\| &= \|f'(x_0)^{-1}f'(x_k)(x_{k+1}-x_k)\| \\
&= \|(f'(x_0)^{-1}(f'(x_k)-f'(x_0))+I)\,(x_{k+1}-x_k)\| \\
&\leq \left(\|(f'(x_0)^{-1}(f'(x_k)-f'(x_0))\|+1\right)\|x_{k+1}-x_k\| \\
&\leq (\gamma r+1)\|x_{k+1}-x_k\|
\end{aligned}$$

- since $\|x_{k+1}-x_k\| \longrightarrow 0$ and f is continuous,

$$\|f'(x_0)^{-1}f(x^\star)\| = \lim_{k\to\infty}\|f'(x_0)^{-1}f(x_k)\| = 0$$

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*
- α_k is limit on relative error in the Newton equation
- provides a stopping condition in iterative method for solving Newton equation
- α_k is constant or adjusted adaptively

Local convergence

Assumptions

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

Local convergence result

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

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

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

	α_k constant	$\alpha_k = 0$	$\alpha_k \searrow 0$
convergence:	linear	quadratic	superlinear

References

Newton method

- J. E. Dennis, Jr., and R. B. Schabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* (1996).
- P. Deuflhard, *Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms* (2011).
- C. T. Kelley, *Iterative Methods for Linear and Nonlinear Equations* (1995).
- J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables* (2000).

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

- P. Deuflhard, *Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms* (2011), theorem 2.1.
- T. Yamamoto, *A unified derivation of several error bounds for Newton's process*, Journal of Computational and Applied Mathematics (1985).

Inexact Newton method

- C. T. Kelley, *Iterative Methods for Linear and Nonlinear Equations* (1995), chapter 6.
- J. Nocedal and S. J. Wright, *Numerical Optimization* (2006), chapter 7.