Numerical Analysis

Lecture 3: The Solutions of Nonlinear Systems

Instructor: Prof. Qiwei Zhan
Zhejiang University

Fixed Points for Functions of Several Variables

Nonlinear Systems

A system of nonlinear equations has the form

$$\begin{aligned} f_1(x_1, x_2, \ldots, x_n) &= 0, \\ f_2(x_1, x_2, \ldots, x_n) &= 0, \\ &\vdots \\ f_n(x_1, x_2, \ldots, x_n) &= 0, \end{aligned}$$

which can be represented by

$$\mathbf{F}(\mathbf{x}) = \mathbf{0}$$

The functions $f_1, f_2, \ldots, f_n$ are called the coordinate functions of $\mathbf{F}$.

Outline

Vector Norms

- Let $\mathbb{R}^n$ denote the set of all n-dimensional column vectors with real-number components.
- To define a distance in $\mathbb{R}^n$ we use the notion of a norm, which is the generalization of the absolute value on $\mathbb{R}$, the set of real numbers.

Definition: Vector Norm

A vector norm on $\mathbb{R}^n$ is a function, $||\cdot||$, from $\mathbb{R}^n$ into $\mathbb{R}$ with the following properties:

1. $||\mathbf{x}|| \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$
2. $||\mathbf{x}|| = 0$ iff $\mathbf{x} = 0$
3. $||\alpha\mathbf{x}|| = |\alpha|||\mathbf{x}||$ for all $\alpha \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$
4. $||\mathbf{x}+\mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$ for all $||\mathbf{x}||, ||\mathbf{y}|| \in \mathbb{R}^n$.

Vector Norms

Definition: L_1, L_2, and L_∞ Norms

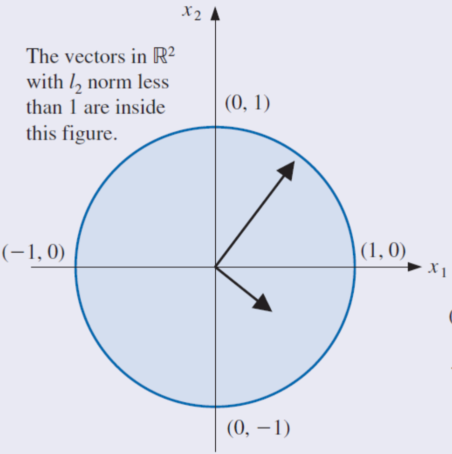
The norms for the vector $\mathbf{x} = (x_1, x_2, \cdots, x_n)^\top$ are defined by:

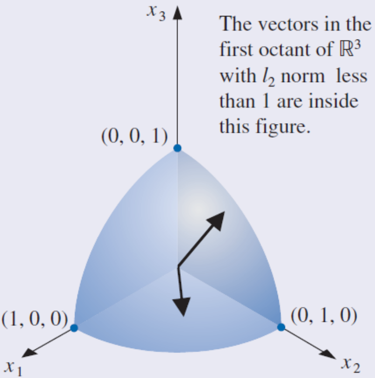
- L_1 Norm
$$||\mathbf{x}||_1 = \sum_{i=1}^{n} |x_i|$$
- L_2 Norm
$$||\mathbf{x}||_2 = \left\{ \sum_{i=1}^{n} x_i^2 \right\}^{1/2}$$
- L_∞ Norm
$$||\mathbf{x}||_\infty = \max_{1 \le i \le n} |x_i|$$

Note that each of these norms reduces to the absolute value in the case $n = 1$.

Vector Norms

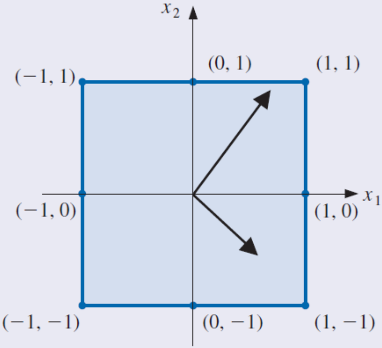
Boundary in $\mathbb{R}^2$ and $\mathbb{R}^3$ with $L_2 < 1$



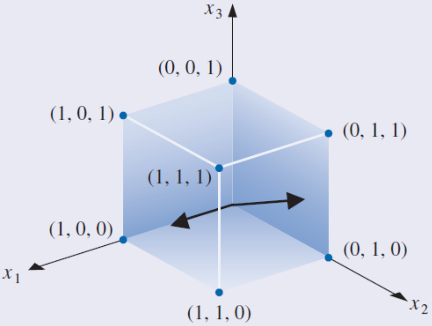


Vector Norms

Boundary in $\mathbb{R}^2$ and $\mathbb{R}^3$ with $L_\infty < 1$



The vectors in $\mathbb{R}^2$ with l_∞ norm less than 1 are inside this figure.



The vectors in the first octant of $\mathbb{R}^3$ with l_∞ norm less than 1 are inside this figure.

Vector Norms

Establishing the Properties of a Vector Norm for L_∞

- It is easy to show that the first three properties of the vector norm hold for L_∞ norm.
- The only property that requires much demonstration is

$$||\mathbf{x}+\mathbf{y}||_\infty \leq ||\mathbf{x}||_\infty + ||\mathbf{y}||_\infty \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$

Vector Norms

Establishing the Properties of a Vector Norm for L_∞

- It is easy to show that the first three properties of the vector norm hold for L_∞ norm.
- The only property that requires much demonstration is

$$||\mathbf{x}+\mathbf{y}||_\infty \le ||\mathbf{x}||_\infty + ||\mathbf{y}||_\infty \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$

- In this case, if $\mathbf{x} = (x_1, x_2, \cdots, x_n)^\top$ and $\mathbf{y} = (y_1, y_2, \cdots, y_n)^\top$, then

$$\begin{aligned} ||\mathbf{x}+\mathbf{y}||_\infty &= \max_{1\le i\le n} |x_i + y_i| \le \max_{1\le i\le n} (|x_i| + |y_i|) \\ &\le \max_{1\le i\le n} |x_i| + \max_{1\le i\le n} |y_i| = ||\mathbf{x}||_\infty + ||\mathbf{y}||_\infty \end{aligned}$$

Vector Norms

Establishing the Properties of a Vector Norm for L_2

- It is easy to show that the first three properties of the vector norm hold for L_2 norm.
- But to show that
$$||\mathbf{x}+\mathbf{y}||_2 \leq ||\mathbf{x}||_2 + ||\mathbf{y}||_2 \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$
we need a famous inequality.

Theorem (Cauchy-Bunyakovsky-Schwarz Inequality for Sums)

For each $\mathbf{x} = (x_1, x_2, \cdots, x_n)^\top$ *and* $\mathbf{y} = (y_1, y_2, \cdots, y_n)^\top$ *in* $\mathbb{R}^n$,
$$\mathbf{x}^\top \mathbf{y} = \sum_{i=1}^{n} x_i y_i \leq \left\{ \sum_{i=1}^{n} x_i^2 \right\}^{1/2} \left\{ \sum_{i=1}^{n} y_i^2 \right\}^{1/2} = ||\mathbf{x}||_2 ||\mathbf{y}||_2$$

Vector Norms

Proof (1/2)

- If $\mathbf{y} = \mathbf{0}$ or $\mathbf{x} = \mathbf{0}$, the result is immediate.

Vector Norms

Proof (1/2)

- If $\mathbf{y} = \mathbf{0}$ or $\mathbf{x} = \mathbf{0}$, the result is immediate.
- Suppose $\mathbf{y} \neq \mathbf{0}$ and $\mathbf{x} \neq \mathbf{0}$. Note that, for each $\lambda \in \mathbb{R}$, we have

$$0 \leq ||\mathbf{x} - \lambda\mathbf{y}||_2^2 = \sum_{i=1}^{n}(x_i - \lambda y_i)^2 = \sum_{i=1}^{n} x_i^2 - 2\lambda \sum_{i=1}^{n} x_i y_i + \lambda^2 \sum_{i=1}^{n} y_i^2$$

so that

$$2\lambda \sum_{i=1}^{n} x_i y_i \leq \sum_{i=1}^{n} x_i^2 + \lambda^2 \sum_{i=1}^{n} y_i^2 = ||\mathbf{x}||_2^2 + \lambda^2 ||\mathbf{y}||_2^2$$

However $||\mathbf{x}||_2 > 0$ and $||\mathbf{y}||_2 > 0$, so we can let $\lambda = ||\mathbf{x}||_2 / ||\mathbf{y}||_2$ to give

$$\left(2\frac{||\mathbf{x}||_2}{||\mathbf{y}||_2}\right)\left(\sum_{i=1}^{n} x_i y_i\right) \leq ||\mathbf{x}||_2^2 + \frac{||\mathbf{x}||_2^2}{||\mathbf{y}||_2^2}||\mathbf{y}||_2^2 = 2||\mathbf{x}||_2^2$$

Vector Norms

Proof (2/2)

Hence

$$2\sum_{i=1}^{n} x_i y_i \leq 2||\mathbf{x}||_2^2 \frac{||\mathbf{y}||_2}{||\mathbf{x}||_2} = 2||\mathbf{x}||_2||\mathbf{y}||_2$$

and

$$\mathbf{x}^\top \mathbf{y} = \sum_{i=1}^{n} x_i y_i \leq ||\mathbf{x}||_2||\mathbf{y}||_2 = \left\{\sum_{i=1}^{n} x_i^2\right\}^{1/2} \left\{\sum_{i=1}^{n} y_i^2\right\}^{1/2}$$

Vector Norms

Proof: $||\mathbf{x}+\mathbf{y}||_2 \leq ||\mathbf{x}||_2 + ||\mathbf{y}||_2$

With the Cauchy-Bunyakovsky-Schwarz Inequality, we see that for each $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$\begin{aligned} ||\mathbf{x}+\mathbf{y}||_2^2 &= \sum_{i=1}^{n}(x_i+y_i)^2 \\ &= \sum_{i=1}^{n} x_i^2 + 2\sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n} y_i^2 \\ &\leq ||\mathbf{x}||_2^2 + 2||\mathbf{x}||_2||\mathbf{y}||_2 + ||\mathbf{y}||_2^2 \end{aligned}$$

which gives norm property:

$$||\mathbf{x}+\mathbf{y}||_2 \leq \left(||\mathbf{x}||_2^2 + 2||\mathbf{x}||_2||\mathbf{y}||_2 + ||\mathbf{y}||_2^2\right)^{1/2} = ||\mathbf{x}||_2 + ||\mathbf{y}||_2$$

Distance between Vectors in $\mathbb{R}^n$

Definition: Distance between Vectors

If $\mathbf{x} = (x_1, x_2, \cdots, x_n)^\top$ and $\mathbf{y} = (y_1, y_2, \cdots, y_n)^\top$ are vectors in $\mathbb{R}^n$, the L_2 and L_∞ distances between $\mathbf{x}$ and $\mathbf{y}$ are defined by

$$||\mathbf{x} - \mathbf{y}||_2 = \left\{ \sum_{i=1}^{n} (x_i - y_i)^2 \right\}^{1/2}$$

and

$$||\mathbf{x} - \mathbf{y}||_\infty = \max_{1 \le i \le n} |x_i - y_i|$$

A Limit of a Sequence of Vectors in $\mathbb{R}^n$

Definition: A Limit of a Sequence of Vectors in $\mathbb{R}^n$

A sequence $\{\mathbf{x}^{(k)}\}_{k=1}^{\infty}$ of vectors in $\mathbb{R}^n$ is said to converge to $\mathbf{x}$ with respect to the norm $||\cdot||$ if, given any $\epsilon > 0$, there exists an integer N_ϵ such that

$$||\mathbf{x}^{(k)} - \mathbf{x}|| < \epsilon, \quad \text{for all } k \geq N_\epsilon$$

Theorem

The sequence of vectors $\{\mathbf{x}^{(k)}\}$ converges to $\mathbf{x}$ in $\mathbb{R}^n$ with respect to the L_∞ norm iff

$$\lim_{k\to\infty} x_i^{(k)} = x_i$$

for each $i = 1, 2, \cdots, n$.

A Limit of a Sequence of Vectors in $\mathbb{R}^n$

Proof (1/2)

Suppose $\{\mathbf{x}^{(k)}\}$ converges to $\mathbf{x}$ with respect to the L_∞ norm. Given any $\epsilon > 0$, there exists an integer N_ϵ such that for all $k \geq N_\epsilon$,

$$\max_{i=1,2,\cdots,n} |x_i^{(k)} - x_i| = ||\mathbf{x}^{(k)} - \mathbf{x}||_\infty < \epsilon$$

This result implies that $|x_i^{(k)} - x_i| < \epsilon$, for each $i = 1, 2, \cdots, n$, so

$$\lim_{k \to \infty} x_i^{(k)} = x_i$$

for each i.

A Limit of a Sequence of Vectors in $\mathbb{R}^n$

Proof (2/2)

Conversely, suppose that $\lim_{k\to\infty} x_i^{(k)} = x_i$, for every $i = 1, 2, \cdots, n$. For a given $\epsilon > 0$, let $N_{i\epsilon}$ for each i represent an integer with the property that

$$|x_i^{(k)} - x_i| < \epsilon$$

whenever $k \geq N_{i\epsilon}$. Define $N_\epsilon = \max_{i=1,2,\cdots,n} N_{i\epsilon}$. If $k \geq N_\epsilon$, then

$$\max_{i=1,2,\cdots,n} |x_i^{(k)} - x_i| = ||\mathbf{x}^{(k)} - \mathbf{x}||_\infty < \epsilon$$

This implies that $\{\mathbf{x}^{(k)}\}$ converges to $\mathbf{x}$ with respect to the L_∞ norm.

A Limit of a Sequence of Vectors in $\mathbb{R}^n$

Theorem

For each $\mathbf{x} \in \mathbb{R}^n$, $||\mathbf{x}||_\infty \leq ||\mathbf{x}||_2 \leq \sqrt{n}||\mathbf{x}||_\infty$

A Limit of a Sequence of Vectors in $\mathbb{R}^n$

Theorem

For each $\mathbf{x} \in \mathbb{R}^n$, $||\mathbf{x}||_\infty \leq ||\mathbf{x}||_2 \leq \sqrt{n}||\mathbf{x}||_\infty$

Proof

Let x_j be a coordinate of $\mathbf{x}$ such that $||\mathbf{x}||_\infty = \max_{1\leq i\leq n} |x_i| = |x_j|$.
Then

$$||\mathbf{x}||_\infty^2 = |x_j|^2 \leq \sum_{i=1}^{n} x_i^2 = ||\mathbf{x}||_2^2$$

and

$$||\mathbf{x}||_\infty \leq ||\mathbf{x}||_2$$

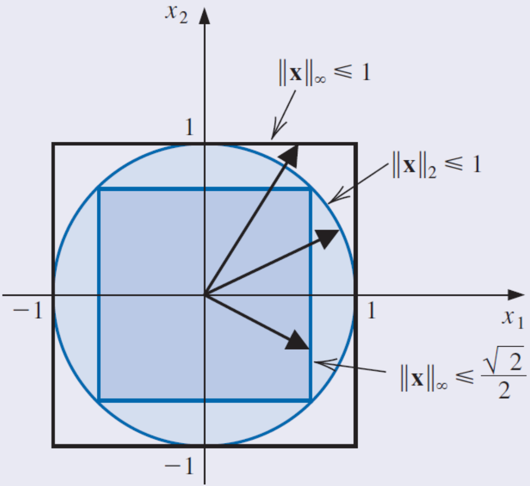
so

$$||\mathbf{x}||_2^2 = \sum_{i=1}^{n} x_i^2 \leq \sum_{i=1}^{n} x_j^2 = n x_j^2 = n||\mathbf{x}||_\infty^2$$

and $||\mathbf{x}||_2 \leq \sqrt{n}||\mathbf{x}||_\infty$.

A Limit of a Sequence of Vectors in $\mathbb{R}^n$

$||\mathbf{x}||_\infty \leq ||\mathbf{x}||_2 \leq \sqrt{n}||\mathbf{x}||_\infty$ Illustration when $n = 2$



A Limit of a Sequence of Vectors in $\mathbb{R}^n$

Theorem

$\| \mathbf{x^{(k)}} - \mathbf{x} \|_\infty < \varepsilon \Rightarrow \| \mathbf{x^{(k)}} - \mathbf{x} \|_\mathbf{2} < \sqrt{\mathbf{n}}\varepsilon$
$\| \mathbf{x^{(k)}} - \mathbf{x} \|_\mathbf{2} < \varepsilon \Rightarrow \| \mathbf{x^{(k)}} - \mathbf{x} \|_\infty < \varepsilon$

Example

Show that

$$\mathbf{x}^{(k)} = (x_1^{(k)}, x_2^{(k)}, x_3^{(k)}, x_4^{(k)})^\top = \left(1, 2 + \frac{1}{k}, \frac{3}{k^2}, e^{-k} sin(k)\right)^\top$$

- converges to $\mathbf{x} = (1, 2, 0, 0)^\top$ with respect to the L_∞ norm.
- It also converges to $\mathbf{x} = (1, 2, 0, 0)^\top$ with respect to the L_2 norm.

A Limit of a Sequence of Vectors in $\mathbb{R}^n$

Solution

Given any $\epsilon > 0$, there exists an integer $N(\epsilon/2)$ with the property that

$$||\mathbf{x}^{(k)} - \mathbf{x}||_\infty < \frac{\epsilon}{2},$$

whenever $k \geq N(\epsilon/2)$. It implies that

$$||\mathbf{x}^{(k)} - \mathbf{x}||_2 \leq \sqrt{4}||\mathbf{x}^{(k)} - \mathbf{x}||_\infty \leq 2\frac{\epsilon}{2} = \epsilon$$

when $k \geq N(\epsilon/2)$. So $\{\mathbf{x}^{(k)}\}$ also converges to $\mathbf{x}$ with respect to the L_2 norm.

Outline

Fixed Points for Functions of Several Variables

Nonlinear Systems

A system of nonlinear equations has the form

$$\begin{aligned}
f_1(x_1, x_2, \ldots, x_n) &= 0, \\
f_2(x_1, x_2, \ldots, x_n) &= 0, \\
&\vdots \\
f_n(x_1, x_2, \ldots, x_n) &= 0,
\end{aligned}$$

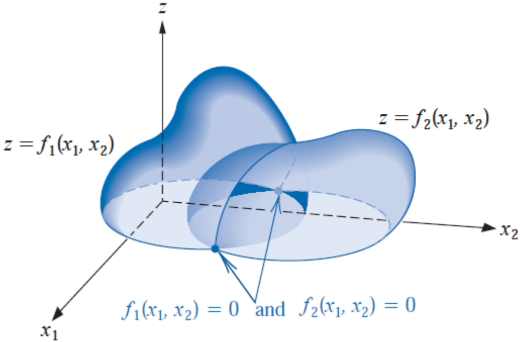
which can be represented by

$$\mathbf{F}(\mathbf{x}) = \mathbf{0}$$

The functions $f_1, f_2, \ldots, f_n$ are called the coordinate functions of **F**.

Fixed Points for Functions of Several Variables

A System of Two Nonlinear Equations



Fixed Points for Functions of Several Variables

Fixed Points in $\mathbb{R}^n$

A function $\mathbf{G}$ from $\mathbf{D} \in \mathbb{R}^n$ into $\mathbb{R}^n$ has a fixed point at $\mathbf{p} \in D$ if $\mathbf{G}(\mathbf{p}) = \mathbf{p}$.

Theorem (Existence of Fixed Points)

Let $\mathbf{D} = \{(x_1, x_2, \ldots, x_n)^\top | a_i \leq x_i \leq b_i, \text{for each } i = 1, 2, \ldots, n\}$ for some collection of constants $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$. Suppose $\mathbf{G}$ is a continuous function from $\mathbf{D} \subset \mathbb{R}^n$ into $\mathbb{R}^n$ with the property that $\mathbf{G}(\mathbf{x}) \in \mathbf{D}$ whenever $\mathbf{x} \in \mathbf{D}$. Then $\mathbf{G}$ has a fixed point in $\mathbf{D}$.

Fixed Points for Functions of Several Variables

Theorem (Fixed Point Theorem)

Let $\mathbf{D} = \{(x_1, x_2, \ldots, x_n)^\top | a_i \leq x_i \leq b_i,$ *for each* $i = 1, 2, \ldots, n\}$ *for some collection of constants* $a_1, a_2, \ldots, a_n$ *and* $b_1, b_2, \ldots, b_n$*. Suppose* $\mathbf{G}$ *is a continuous function from* $\mathbf{D} \in \mathbb{R}^n$ *into* $\mathbb{R}^n$ *with the property that* $\mathbf{G}(\mathbf{x}) \in \mathbf{D}$ *whenever* $\mathbf{x} \in \mathbf{D}$*. Then* $\mathbf{G}$ *has a fixed point in* $\mathbf{D}$*.*

Moreover, suppose that all the component functions of $\mathbf{G}$ *have continuous partial derivatives and a constant* $K < 1$ *exists with*

$$\left| \frac{\partial g_i(\mathbf{x})}{\partial x_j} \right| \leq \frac{K}{n}, \;\; \textit{whenever} \;\; \mathbf{x} \in \mathbf{D},$$

for each $j = 1, 2, \ldots, n$ *and each component function* g_i*. Then the sequence* $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$ *defined by an arbitrarily selected* $\mathbf{x}^{(0)}$ *in* $\mathbf{D}$ *and generated by*

$$\mathbf{x}^{(k)} = \mathbf{G}(\mathbf{x}^{(k-1)}), \;\; \textit{for each} \;\; k \geq 1$$

converges to the unique fixed point $\mathbf{p} \in \mathbf{D}$ *and*

$$||\mathbf{x}^{(k)} - \mathbf{p}||_\infty \leq \frac{K^k}{1-K} ||\mathbf{x}^{(1)} - \mathbf{x}^{(0)}||_\infty$$

Fixed Points for Functions of Several Variables

Example

Find a solution for the nonlinear system

$$
\begin{aligned}
3x_1 - \cos(x_2 x_3) - \frac{1}{2} &= 0, \\
x_1^2 - 81(x_2 + 0.1)^2 + \sin x_3 + 1.06 &= 0, \\
e^{-x_1 x_2} + 20x_3 + \frac{10\pi - 3}{3} &= 0.
\end{aligned}
$$

Fixed Points for Functions of Several Variables

Example

Find a solution for the nonlinear system

$$\begin{aligned}
3x_1 - \cos(x_2x_3) - \frac{1}{2} &= 0,\\
x_1^2 - 81(x_2+0.1)^2 + \sin x_3 + 1.06 &= 0,\\
e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3} &= 0.
\end{aligned}$$

$$\begin{aligned}
x_1 &= \frac{1}{3}\cos(x_2x_3) + \frac{1}{6},\\
x_2 &= \frac{1}{9}\sqrt{x_1^2 + \sin x_3 + 1.06} - 0.1,\\
x_3 &= -\frac{1}{20}e^{-x_1x_2} - \frac{10\pi - 3}{60}.
\end{aligned}$$

Example

Solution:

$$g_1(x_1,x_2,x_3) = \frac{1}{3}\cos(x_2x_3) + \frac{1}{6},$$

$$g_2(x_1,x_2,x_3) = \frac{1}{9}\sqrt{x_1^2 + \sin x_3 + 1.06} - 0.1,$$

$$g_3(x_1,x_2,x_3) = -\frac{1}{20}e^{-x_1x_2} - \frac{10\pi - 3}{60}.$$

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$\lVert \mathbf{x}^{(k)} - \mathbf{x}^{(k-1)} \rVert_\infty$
0	0.10000000	0.10000000	−0.10000000	
1	0.49998333	0.00944115	−0.52310127	0.423
2	0.49999593	0.00002557	−0.52336331	9.4×10^{-3}
3	0.50000000	0.00001234	−0.52359814	2.3×10^{-4}
4	0.50000000	0.00000003	−0.52359847	1.2×10^{-5}
5	0.50000000	0.00000002	−0.52359877	3.1×10^{-7}

Fixed Points for Functions of Several Variables

Acceleration Convergence

Use the latest estimates $x_1^{(k)}, \cdots, x_{i-1}^{(k)}$ instead of $x_1^{(k-1)}, \cdots, x_{i-1}^{(k-1)}$ to compute $x_i^{(k)}$.

$$x_1^{(k)} = \frac{1}{3}\cos\left(x_2^{(k-1)} x_3^{(k-1)}\right) + \frac{1}{6},$$

$$x_2^{(k)} = \frac{1}{9}\sqrt{\left(x_1^{(k)}\right)^2 + \sin x_3^{(k-1)} + 1.06} - 0.1,$$

$$x_3^{(k)} = -\frac{1}{20} e^{-x_1^{(k)} x_2^{(k)}} - \frac{10\pi - 3}{60}.$$

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$\left\Vert \mathbf{x}^{(k)} - \mathbf{x}^{(k-1)} \right\Vert_\infty$
0	0.10000000	0.10000000	−0.10000000	
1	0.49998333	0.02222979	−0.52304613	0.423
2	0.49997747	0.00002815	−0.52359807	2.2×10^{-2}
3	0.50000000	0.00000004	−0.52359877	2.8×10^{-5}
4	0.50000000	0.00000000	−0.52359877	3.8×10^{-8}

Outline

Newton's Method for Nonlinear Systems

The fixed-point method for nonlinear equations

$$g(x) = x - \phi(x)f(x).$$

The fixed-point method for nonlinear systems

$$G(\mathbf{x}) = \mathbf{x} - \mathbf{A}(\mathbf{x})^{-\mathbf{1}}\mathbf{F}(\mathbf{x}),$$

where

$$\mathbf{A}(\mathbf{x}) = \begin{bmatrix} a_{11}(\mathbf{x}) & a_{12}(\mathbf{x}) & \dots & a_{1n}(\mathbf{x}) \\ a_{21}(\mathbf{x}) & a_{22}(\mathbf{x}) & \dots & a_{2n}(\mathbf{x}) \\ \dots & \dots & \dots & \dots \\ a_{n1}(\mathbf{x}) & a_{n1}(\mathbf{x}) & \dots & a_{nn}(\mathbf{x}) \end{bmatrix}$$

Newton's Method for Nonlinear Systems

Theorem (Convergence)

Let $\mathbf{p}$ *be a solution of* $\mathbf{G}(\mathbf{x}) = \mathbf{x}$*. Suppose a number* $\delta > 0$ *exists with*

1. $\partial g_i/\partial x_j$ *is continuous on* $N_\delta = \{\mathbf{x} | ||\mathbf{x} - \mathbf{p}|| < \delta\}$*, for each* $i = 1, 2, \cdots, n$ *and* $j = 1, 2, \cdots, n$*;*
2. $\partial^2 g_i(\mathbf{x})/(\partial x_j \partial x_k)$ *is continuous, and* $|\partial^2 g_i(\mathbf{x})/(\partial x_j \partial x_k)| \leq M$ *for some constant* M*, whenever* $\mathbf{x} \in N_\delta$*, for each* $i = 1, 2, \cdots, n$*,* $j = 1, 2, \cdots, n$*, and* $k = 1, 2, \cdots, n$*;*
3. $\partial g_i(\mathbf{p})/\partial x_k = 0$*, for each* $i = 1, 2, \cdots, n$ *and* $k = 1, 2, \cdots, n$*.*

Then a number $\hat{\delta} \leq \delta$ *exists such that the sequence generated by* $\mathbf{x}^{(k)} = \mathbf{G}(\mathbf{x}^{(k-1)})$ *converges quadratically to* $\mathbf{p}$ *for any choice of* $\mathbf{x}^{(0)}$*, provided that* $||\mathbf{x}^{(0)} - \mathbf{p}|| < \hat{\delta}$*. Moreover,*

$$||\mathbf{x}^{(k)} - \mathbf{p}||_\infty \leq \frac{n^2 M}{2} ||\mathbf{x}^{(k-1)} - \mathbf{p}||_\infty^2, \textit{for each } k \geq 1$$

Newton's Method for Nonlinear Systems

Construction of the Matrix

Let $b_{ij}(\mathbf{x})$ denote the entry of $\mathbf{A}(\mathbf{x})^{-1}$, we have

$$g_i(\mathbf{x}) = x_i - \sum_{j=1}^{n} b_{ij}(\mathbf{x}) f_j(\mathbf{x})$$

$$\frac{\partial g_i}{\partial x_k}(\mathbf{x}) = \begin{cases} 1 - \sum\limits_{j=1}^{n} \left(b_{ij}(\mathbf{x}) \frac{\partial f_i}{\partial x_k}(\mathbf{x}) + \frac{\partial b_{ij}}{\partial x_k}(\mathbf{x}) f_j(\mathbf{x}) \right), \textit{if } i = k, \\ -\sum\limits_{j=1}^{n} \left(b_{ij}(\mathbf{x}) \frac{\partial f_i}{\partial x_k}(\mathbf{x}) + \frac{\partial b_{ij}}{\partial x_k}(\mathbf{x}) f_j(\mathbf{x}) \right), \textit{if } i \neq k. \end{cases}$$

This means that for $i = k$, $\sum\limits_{j=1}^{n} b_{ij}(\mathbf{p}) \frac{\partial f_j}{\partial x_i}(\mathbf{p}) = 1$

and for $i \neq k$, $\sum\limits_{j=1}^{n} b_{ij}(\mathbf{p}) \frac{\partial f_j}{\partial x_i}(\mathbf{p}) = 0$

Newton's Method for Nonlinear Systems

$$J(\mathbf{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \frac{\partial f_1}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{x}) & \frac{\partial f_2}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial f_2}{\partial x_n}(\mathbf{x}) \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1}(\mathbf{x}) & \frac{\partial f_n}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial f_n}{\partial x_n}(\mathbf{x}) \end{bmatrix}.$$

The Jacobian matrix

Then, $\mathbf{A}(\mathbf{p})^{-1}\mathbf{J}(\mathbf{p}) = \mathbf{I}$. Therefore, $\mathbf{A}(\mathbf{p}) = \mathbf{J}(\mathbf{p})$.

Newton's Method for Nonlinear Systems

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

Newton's Method for Nonlinear Systems

$$\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} - \mathbf{J}(\mathbf{x}^{(k-1)})^{-1}\mathbf{F}(\mathbf{x}^{(k-1)}),$$

where $\mathbf{J}(\mathbf{x})$ is the Jacobian matrix

$$J(\mathbf{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \frac{\partial f_1}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{x}) & \frac{\partial f_2}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial f_2}{\partial x_n}(\mathbf{x}) \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1}(\mathbf{x}) & \frac{\partial f_n}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial f_n}{\partial x_n}(\mathbf{x}) \end{bmatrix}.$$

Newton's Method for Nonlinear Systems

$$\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} - \mathbf{J}(\mathbf{x}^{(k-1)})^{-1}\mathbf{F}(\mathbf{x}^{(k-1)}),$$

Remarks

- A weakness in Newton's method arises from the need to compute and invert the matrix $\mathbf{J}(\mathbf{x})$ at each step.
- In practice, explicit computation of $\mathbf{J}(\mathbf{x})^{-1}$ is avoided by performing the operation in a two-step manner.
  1. A vector $\mathbf{y}$ is found that satisfies $\mathbf{J}(\mathbf{x}^{(k-1)})\mathbf{y} = -\mathbf{F}(\mathbf{x}^{(k-1)})$
  2. The new approximation, $\mathbf{x}^{(k)}$, is obtained by adding $\mathbf{y}$ to $\mathbf{x}^{(k-1)}$.

Newton's Method for Nonlinear Systems

Algorithm

INPUT number n of equations and unknowns; initial approximation $\mathbf{x} = (x_1, \ldots, x_n)^t$, tolerance TOL; maximum number of iterations N.
OUTPUT approximate solution $\mathbf{x} = (x_1, \ldots, x_n)^t$ or a message that the number of iterations was exceeded.

Step 1 Set $k = 1$.

Step 2 While ($k \leq N$) do Steps 3–7.

Step 3 Calculate $\mathbf{F}(\mathbf{x})$ and $J(\mathbf{x})$, where $J(\mathbf{x})_{i,j} = (\partial f_i(\mathbf{x})/\partial x_j)$ for $1 \leq i, j \leq n$.

Step 4 Solve the $n \times n$ linear system $J(\mathbf{x})\mathbf{y} = -\mathbf{F}(\mathbf{x})$.

Step 5 Set $\mathbf{x} = \mathbf{x} + \mathbf{y}$.

Step 6 If $||\mathbf{y}|| < TOL$ then OUTPUT ($\mathbf{x}$);
(*The procedure was successful.*)
STOP.

Step 7 Set $k = k + 1$.

Step 8 OUTPUT ('Maximum number of iterations exceeded');
(*The procedure was unsuccessful.*)
STOP.

Newton's Method for Nonlinear Systems

Example

Find a solution for the nonlinear system

$$
\begin{aligned}
3x_1 - \cos(x_2 x_3) - \frac{1}{2} &= 0,\\
x_1^2 - 81(x_2 + 0.1)^2 + \sin x_3 + 1.06 &= 0,\\
e^{-x_1 x_2} + 20x_3 + \frac{10\pi - 3}{3} &= 0.
\end{aligned}
$$

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|_\infty$
0	0.1000000000	0.1000000000	−0.1000000000	
1	0.4998696728	0.0194668485	−0.5215204718	0.4215204718
2	0.5000142403	0.0015885914	−0.5235569638	1.788×10^{-2}
3	0.5000000113	0.0000124448	−0.5235984500	1.576×10^{-3}
4	0.5000000000	8.516×10^{-10}	−0.5235987755	1.244×10^{-5}
5	0.5000000000	-1.375×10^{-11}	−0.5235987756	8.654×10^{-10}

Newton's Method for Nonlinear Systems

Remarks

- The advantage of the Newton's method for solving systems of nonlinear equations is its speed of convergence once a sufficiently accurate approximation is known.
- A weakness of this method is that an accurate initial approximation to the solution is needed to ensure convergence.

Outline

Gradient Descent Techniques

Gradient Descent

- A.K.A Steepest Descent method, converges only linearly to the solution, but it will usually converge even for poor initial approximations.
- The method of Steepest Descent determines a local minimum for a multivariable function.

Gradient Descent Techniques

Gradient Descent Method

A system of nonlinear equations has the form

$$\begin{aligned} f_1(x_1, x_2, \ldots, x_n) &= 0, \\ f_2(x_1, x_2, \ldots, x_n) &= 0, \\ &\vdots \\ f_n(x_1, x_2, \ldots, x_n) &= 0. \end{aligned}$$

Then the following function has the minimal value of 0:

$$g(x_1, x_2, \cdots, x_n) = \sum_{i=1}^{n} f_i^2(x_1, x_2, \cdots, x_n)$$

Gradient Descent Techniques

Gradient Descent Method

The method of Steepest Descent for finding a local minimum for an arbitrary function $\mathbf{g}$ from $\mathbb{R}^n$ into $\mathbb{R}$ can be intuitively described as follows:

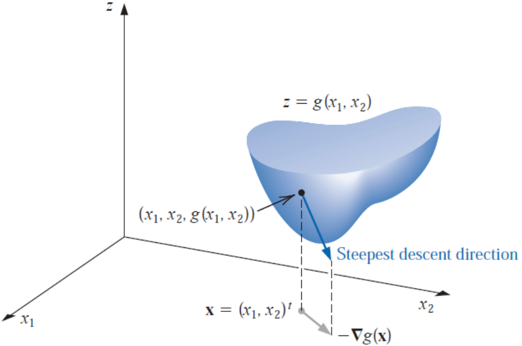
1. Evaluate $\mathbf{g}$ at an initial approximation $\mathbf{x}^{(0)}$;
2. Determine a direction from $\mathbf{x}^{(0)}$ that results in a decrease in the value of $\mathbf{g}$;
3. Move an appropriate amount in this direction and call the new value $\mathbf{x}^{(1)}$;
4. Repeat steps 1 through 3 with $\mathbf{x}^{(0)}$ replaced by $\mathbf{x}^{(1)}$.

Gradient Descent Techniques

The Gradient of a Function

For $g : \mathbb{R}^n \to \mathbb{R}$, the gradient of g at $\mathbf{x} = (x_1, x_2, \cdots, x_n)^\top$ is denoted $\nabla g(\mathbf{x})$ and defined by

$$\nabla g(\mathbf{x}) = \left(\frac{\partial g}{\partial x_1}(\mathbf{x}), \frac{\partial g}{\partial x_2}(\mathbf{x}), \cdots, \frac{\partial g}{\partial x_n}(\mathbf{x}) \right)^\top .$$



Gradient Descent Techniques

Gradient Descent Method

$$\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} - \alpha \nabla \mathbf{g}(\mathbf{x}^{(k-1)}),$$

where α is the step size.

Gradient Descent Techniques

Example

Using gradient descent method to find a solution for the nonlinear system

$$
\begin{aligned}
3x_1 - \cos(x_2x_3) - \frac{1}{2} &= 0, \\
x_1^2 - 81(x_2 + 0.1)^2 + \sin x_3 + 1.06 &= 0, \\
e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3} &= 0.
\end{aligned}
$$

Gradient Descent Techniques

Example

Using gradient descent method to find a solution for the nonlinear system

$$
\begin{aligned}
3x_1 - \cos(x_2x_3) - \frac{1}{2} &= 0,\\
x_1^2 - 81(x_2+0.1)^2 + \sin x_3 + 1.06 &= 0,\\
e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3} &= 0.
\end{aligned}
$$

Solution Let $g(x_1,x_2,x_3) = [f_1(x_1,x_2,x_3)]^2 + [f_2(x_1,x_2,x_3)]^2 + [f_3(x_1,x_2,x_3)]^2$.

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$g(x_1^{(k)}, x_2^{(k)}, x_3^{(k)})$
2	0.137860	−0.205453	−0.522059	1.27406
3	0.266959	0.00551102	−0.558494	1.06813
4	0.272734	−0.00811751	−0.522006	0.468309
5	0.308689	−0.0204026	−0.533112	0.381087
6	0.314308	−0.0147046	−0.520923	0.318837
7	0.324267	−0.00852549	−0.528431	0.287024

Gradient Descent vs. Newton's Method

Gradient Descent vs. Newton's Method

- Newton's method is to find a root $\mathbf{F}(\mathbf{x}) = 0$

$$\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} - \mathbf{J}(\mathbf{x}^{(k-1)})^{-1}\mathbf{F}(\mathbf{x}^{(k-1)}),$$

- Gradient descent is to find a local minimum $g(\mathbf{x}) = ||\mathbf{F}(\mathbf{x})||_2^2$

$$\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} - \alpha\nabla g(\mathbf{x}^{(k-1)}).$$

Assignments

- Reading Assignment: Chap 10
- Homework 2.