Chapter 10.1: Fixed Points for Functions of Several Variables

MOTIVATION

A system of nonlinear equations has the form

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

where each function f_i can be thought of as mapping a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)^t$ of the n-dimensional space $\mathbb{R}^n$ into the real line $\mathbb{R}$.

MOTIVATION

This system of n nonlinear equations in n unknowns can also be represented by defining a function $\mathbf{F}$ mapping $\mathbb{R}^n$ into $\mathbb{R}^n$ as

$$\mathbf{F}(x_1, x_2, \dots, x_n) = (f_1(x_1, x_2, \dots, x_n), f_2(x_1, x_2, \dots, x_n), \dots, f_n(x_1, x_2, \dots, x_n))^t.$$

If vector notation is used to represent the variables $x_1, x_2, \dots, x_n$, then system from the previous slide assumes the form

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

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

MOTIVATION

This system of n nonlinear equations in n unknowns can also be represented by defining a function $\mathbf{F}$ mapping $\mathbb{R}^n$ into $\mathbb{R}^n$ as

$$\mathbf{F}(x_1, x_2, \ldots, x_n) = (f_1(x_1, x_2, \ldots, x_n), f_2(x_1, x_2, \ldots, x_n), \ldots, f_n(x_1, x_2, \ldots, x_n))^t.$$

If vector notation is used to represent the variables $x_1, x_2, \ldots, x_n$, then system from the previous slide assumes the form

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

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

Chapter 10.1: Fixed Points for Functions of Several Variables

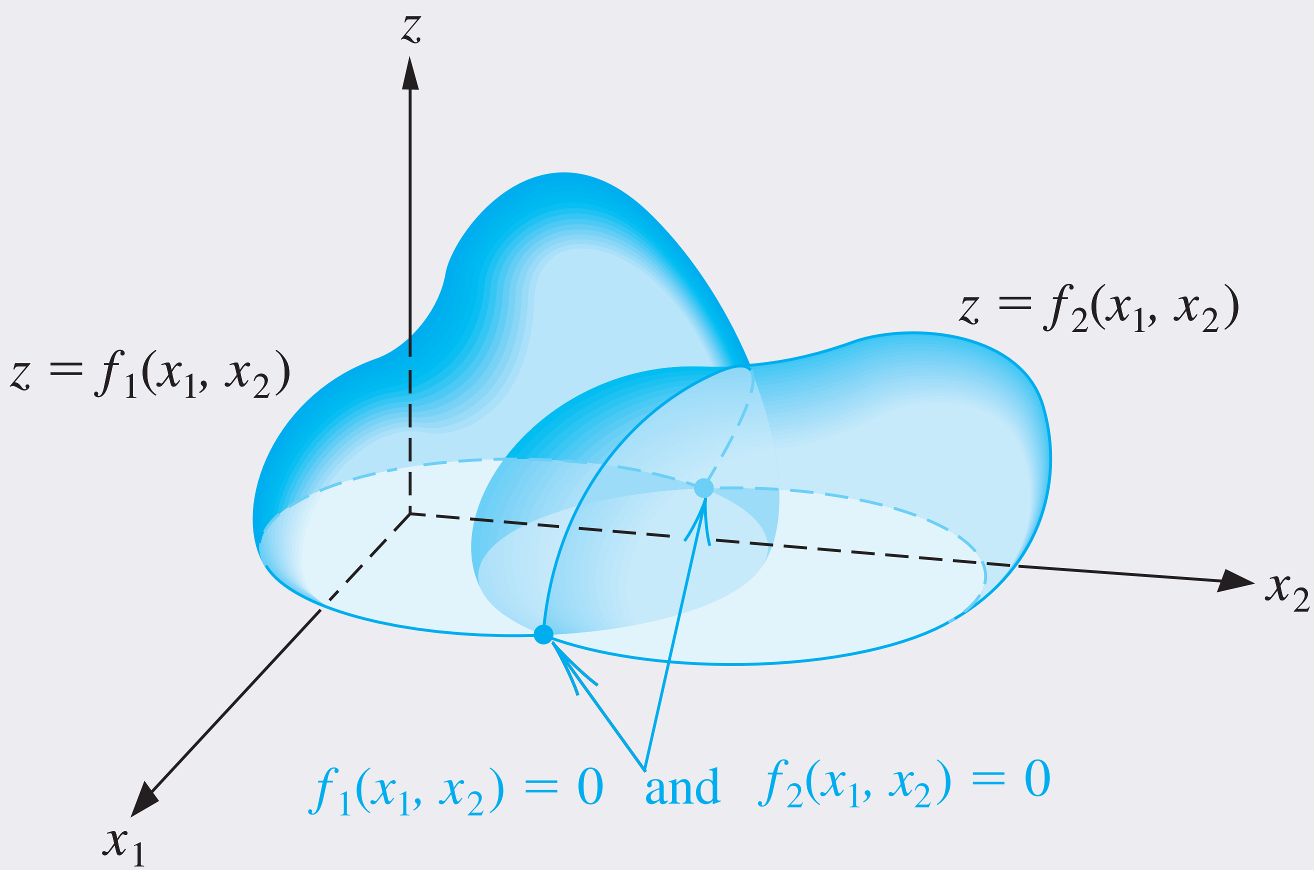

Figure: Figure 10.1

Definition (10.1)

et f be a function defined on a set $D \subset \mathbb{R}^n$ and mapping into $\mathbb{R}$. The function f is said to have the **limit** L at $\mathbf{x}_0$, written

$$\lim_{\mathbf{x} \to \mathbf{x}_0} f(\mathbf{x}) = L,$$

if, given any number $\varepsilon > 0$, a number $\delta > 0$ exists with

$$|f(\mathbf{x}) - L| < \varepsilon,$$

whenever $\mathbf{x} \in D$ and

$$0 < ||\mathbf{x} - \mathbf{x}_0|| < \delta.$$

Definition (10.2)

Let f be a function from a set $D \subset \mathbb{R}^n$ into $\mathbb{R}$. The function f is **continuous** at $\mathbf{x}_0 \in D$ provided $\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x})$ exists and

$$\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = f(\mathbf{x}_0).$$

Moreover, f is **continuous** on a set D if f is continuous at every point of D. This concept is expressed by writing $f \in C(D)$.

Definition (10.3)

Let $\mathbf{F}$ be a function from $D \subset \mathbb{R}^n$ into $\mathbb{R}^n$ of the form

$$\mathbf{F}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_n(\mathbf{x}))^t,$$

where f_i is a mapping from $\mathbb{R}^n$ into $\mathbb{R}$ for each i. We define

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \mathbf{F}(\mathbf{x}) = \mathbf{L} = (L_1, L_2, \ldots, L_n)^t,$$

if and only if $\lim_{\mathbf{x} \to \mathbf{x}_0} f_i(\mathbf{x}) = L_i$, for each $i = 1, 2, \ldots, n$.

Chapter 10.1: Fixed Points for Functions of Several Variables

Theorem (10.4)

Let f be a function from $D \subset \mathbb{R}^n$ into $\mathbb{R}$ and $\mathbf{x}_0 \in D$. Suppose that all the partial derivatives of f exist and constants $\delta > 0$ and $K > 0$ exist so that whenever $\|\mathbf{x} - \mathbf{x}_0\| < \delta$ and $\mathbf{x} \in D$, we have

$$\left| \frac{\partial f(\mathbf{x})}{\partial x_j} \right| \leq K, \quad \text{for each } j = 1, 2, \ldots, n.$$

Then f is continuous at $\mathbf{x}_0$.

Definition (10.5)

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

Theorem (10.6)

Let $D = \{ (x_1, x_2, \ldots, x_n)^t \mid a_i \leq x_i \leq b_i$, 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 $D \subset \mathbb{R}^n$ into $\mathbb{R}^n$ with the property that $\mathbf{G}(\mathbf{x}) \in D$ whenever $\mathbf{x} \in D$. Then $\mathbf{G}$ has a fixed point in D. Suppose also that all the component functions of $\mathbf{G}$ have continuous partial derivatives and a constant $K < 1$ exists with

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

for each $j = 1, 2, \ldots, n$ and each component function g_i. Then the fixed-point sequence $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$ defined by an arbitrarily selected $\mathbf{x}^{(0)}$ in D and generated by $\mathbf{x}^{(k)} = G(\mathbf{x}^{(k-1)}), \quad$ for each $k \geq 1$ converges to the unique fixed point $\mathbf{p} \in D$ and
$\left\| \mathbf{x}^{(k)} - \mathbf{p} \right\|_\infty \leq \frac{K^k}{1-K} \left\| \mathbf{x}^{(1)} - \mathbf{x}^{(0)} \right\|_\infty .$

MOTIVATION

We will use an approach similar to the one used in the one-dimensional fixed-point method for the n-dimensional case. This involves a matrix

$$A(\mathbf{x}) = \begin{bmatrix} a_{11}(\mathbf{x}) & a_{12}(\mathbf{x}) & \cdots & a_{1n}(\mathbf{x}) \\ a_{21}(\mathbf{x}) & a_{22}(\mathbf{x}) & \cdots & a_{2n}(\mathbf{x}) \\ \vdots & \vdots & & \vdots \\ a_{n1}(\mathbf{x}) & a_{n2}(\mathbf{x}) & \cdots & a_{nn}(\mathbf{x}) \end{bmatrix},$$

where each of the entries $a_{ij}(\mathbf{x})$ is a function from $\mathbb{R}^n$ into $\mathbb{R}$. This requires that $A(\mathbf{x})$ be found so that

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

gives quadratic convergence to the solution of $\mathbf{F}(\mathbf{x}) = \mathbf{0}$, assuming that $A(\mathbf{x})$ is nonsingular at the fixed point $\mathbf{p}$ of $\mathbf{G}$.

Theorem (10.7)

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

(i) $\partial g_i/\partial x_j$ *is continuous on* $N_\delta = \{\, \mathbf{x} \mid \|\mathbf{x} - \mathbf{p}\| < \delta \,\}$, *for each* $i = 1, 2, \ldots, n$ *and* $j = 1, 2, \ldots, n$;

(ii) $\partial^2 g_i(\mathbf{x})/(\partial x_j \partial x_k)$ *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* $i = 1, 2, ..., n$, $j = 1, 2, \ldots, n$, *and* $k = 1, 2, \ldots, n$;

(iii) $\partial g_i(\mathbf{p})/\partial x_k = 0$, *for each* $i = 1, 2, \ldots, n$ *and* $k = 1, 2, \ldots, n$.

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* $\left\|\mathbf{x}^{(0)} - \mathbf{p}\right\| < \hat{\delta}$. *Moreover,*

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

The Jacobian Matrix

Define the matrix $J(\mathbf{x})$ by

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

It is required that

$$A(\mathbf{p})^{-1}J(\mathbf{p}) = I, \text{the identity matrix,} \quad \text{so} A(\mathbf{p}) = J(\mathbf{p}).$$

The Jacobian Matrix

An appropriate choice for $A(\mathbf{x})$ is, consequently, $A(\mathbf{x}) = J(\mathbf{x})$ since this satisfies condition (iii) in Theorem 10.7. The function $\mathbf{G}$ is defined by

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

and the fixed-point iteration procedure evolves from selecting $\mathbf{x}^{(0)}$ and generating, for $k \geq 1$,

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

This is called **Newton's method for nonlinear systems**, and it is generally expected to give quadratic convergence, provided that a sufficiently accurate starting value is known and that $J(\mathbf{p})^{-1}$ exists.

MOTIVATION

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

1. Evaluate g at an initial approximation $\mathbf{x}^{(0)} = \left(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)}\right)^t$.
2. Determine a direction from $\mathbf{x}^{(0)}$ that results in a decrease in the value of 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 $\mathbf{x}^{(1)}$.

MOTIVATION

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

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

A differentiable multivariable function can have a relative minimum at $\mathbf{x}$ only when the gradient at $\mathbf{x}$ is the zero vector. Suppose that $\mathbf{v} = (v_1, v_2, \ldots, v_n)^t$ is a unit vector in $\mathbb{R}^n$; that is,

$$||\mathbf{v}||_2^2 = \sum_{i=1}^{n} v_i^2 = 1.$$

MOTIVATION

The **directional derivative** of g at $\mathbf{x}$ in the direction of $\mathbf{v}$ measures the change in the value of the function g relative to the change in the variable in the direction of $\mathbf{v}$. It is defined by

$$D_{\mathbf{v}}g(\mathbf{x}) = \lim_{h \to 0} \frac{1}{h}[g(\mathbf{x} + h\mathbf{v}) - g(\mathbf{x})] = \mathbf{v}^t \cdot \nabla g(\mathbf{x}).$$

When g is differentiable, the direction that produces the maximum value for the directional derivative occurs when $\mathbf{v}$ is chosen to be parallel to $\nabla g(\mathbf{x})$, provided that $\nabla g(\mathbf{x}) \neq \mathbf{0}$. As a consequence, the direction of greatest decrease in the value of g at $\mathbf{x}$ is the direction given by $-\nabla g(\mathbf{x})$.

Algorithm 10.3 STEEPEST DESCENT

To approximate a solution **p** to the minimization problem

$$g(\mathbf{p}) = \min_{\mathbf{x} \in \mathbb{R}^n} g(\mathbf{x})$$

given an initial approximation **x**:

INPUT number n of variables; initial approximation $\mathbf{x} = (x_1, \ldots, x_n)^t$
TOL; maximum number of iterations N.
OUTPUT approximate solution $\mathbf{x} = (x_1, \ldots, x_n)^t$ or message of failure.
Step 1 Set $k = 1$.
Step 2 While ($k \leq N$) do Steps 3–15.
Step 3 Set $g_1 = g(x_1, \ldots, x_n)$; $\left(\textit{Note: } g_1 = g\left(\mathbf{x}^{(k)}\right).\right)$
$\mathbf{z} = \nabla g(x_1, \ldots, x_n)$; $\left(\textit{Note: } \mathbf{z} = \nabla g\left(\mathbf{x}^{(k)}\right).\right)$
$z_0 = ||\mathbf{z}||_2$.

Algorithm 10.3 STEEPEST DESCENT

Step 4 If $z_0 = 0$ then OUTPUT ('Zero gradient');
OUTPUT $(x_1, \ldots, x_n, g_1)$;
(*The procedure completed, may have a minimum.*)
STOP.

Step 5 Set $\mathbf{z} = \mathbf{z}/z_0$; (*Make* $\mathbf{z}$ *a unit vector.*)
$\alpha_1 = 0$;
$\alpha_3 = 1$;
$g_3 = g(\mathbf{x} - \alpha_3 \mathbf{z})$.

Step 6 While $(g_3 \geq g_1)$ do Steps 7 and 8.

Step 7 Set $\alpha_3 = \alpha_3/2$;
$g_3 = g(\mathbf{x} - \alpha_3 \mathbf{z})$.

Step 8 If $\alpha_3 < TOL/2$ then
OUTPUT ('No likely improvement');
OUTPUT $(x_1, \ldots, x_n, g_1)$; STOP
(*Procedure completed, may have a minimum.*)

Algorithm 10.3 STEEPEST DESCENT

Step 9 Set $\alpha_2 = \alpha_3/2$;

$g_2 = g(\mathbf{x} - \alpha_2 \mathbf{z})$.

Step 10 Set $h_1 = (g_2 - g_1)/\alpha_2$;

$h_2 = (g_3 - g_2)/(\alpha_3 - \alpha_2)$;

$h_3 = (h_2 - h_1)/\alpha_3$.

(*Note: Newton's forward divided-difference formula* used to find *the quadratic* $P(\alpha) = g_1 + h_1\alpha + h_3\alpha(\alpha - \alpha_2)$ *that interpolates* $h(\alpha)$ *at* $\alpha = 0, \alpha = \alpha_2, \alpha = \alpha_3$.)

Step 11 Set $\alpha_0 = 0.5(\alpha_2 - h_1/h_3)$; (*Critical point of P at* α_0.)

$g_0 = g(\mathbf{x} - \alpha_0 \mathbf{z})$.

Step 12 Find α from $\{\alpha_0, \alpha_3\}$ so $g = g(\mathbf{x} - \alpha \mathbf{z}) = \min\{g_0, g_3\}$.

Step 13 Set $\mathbf{x} = \mathbf{x} - \alpha \mathbf{z}$.

Algorithm 10.3 STEEPEST DESCENT

Step 14 If $|g - g_1| < TOL$ then
OUTPUT $(x_1, \ldots, x_n, g)$;
(*The procedure was successful.*)
STOP.
Step 15 Set $k = k + 1$.
Step 16 OUTPUT ('Maximum iterations exceeded');
(*The procedure was unsuccessful.*)
STOP.