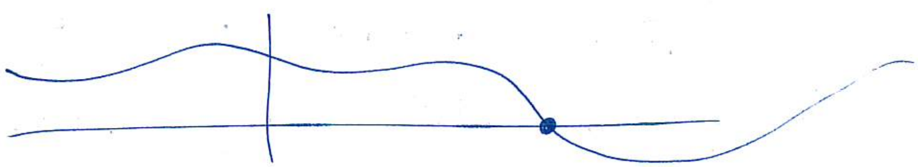
§5.1
Nonlinear Eqs

more difficult than linear!

consider scalar case first.

$$g(x) = y$$
$$g(x) - y = 0$$
$$f(x) = 0$$

find "root" or "zero"
Solution = root or zero

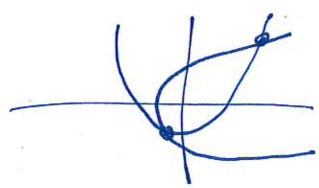


5.2 Existence + Uniqueness

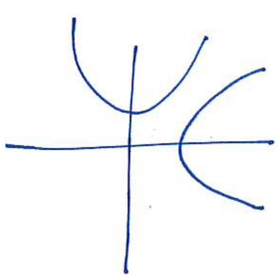
- nonlinear equations can have any # of solution
- each equation is a hypersurface in $\mathbb{R}^n$

Ex. 5.2

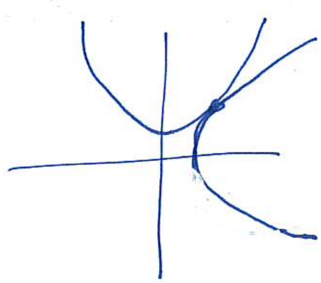
$$x^2 - y + \alpha = 0$$
$$-x + y^2 + \alpha = 0$$



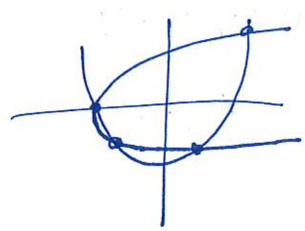
$\alpha = -\frac{1}{2}$
(2 solutions)



$\alpha = \frac{1}{2}$
(0 solutions)



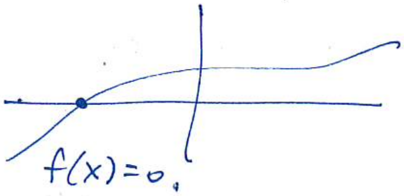
$\alpha = \frac{1}{4}$
(1 solution)



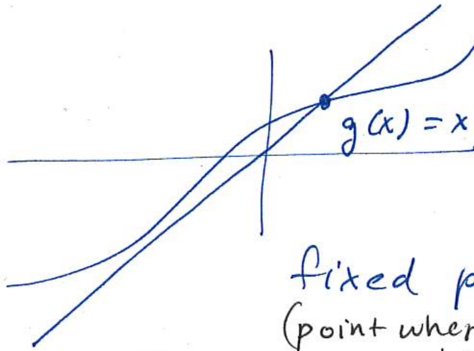
$\alpha = -1$
(4 solutions)

Fixed point iteration

$x = g(x)$ $\quad$ x is a "fixed pt"



root



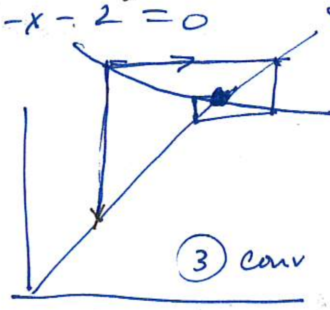
fixed point.
(point where curve intersects diagonal)

many choices of $g(x)$ for $f(x)=0$
— may differ in conv. rates + whether they converge at all!

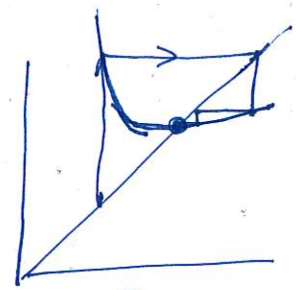
Example 5.8

$f(x) = x^2 - x - 2 = 0$ $\quad$ sol'n: $x^* = 2$, $x^* = -1$

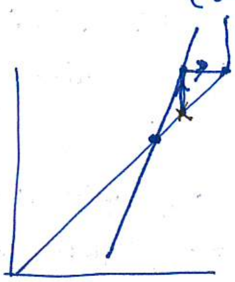
1. $g(x) = x^2 - 2$
2. $g(x) = \sqrt{x+2}$
3. $g(x) = 1 + 2/x$
4. $g(x) = \dfrac{(x^2+2)}{(2x-1)}$



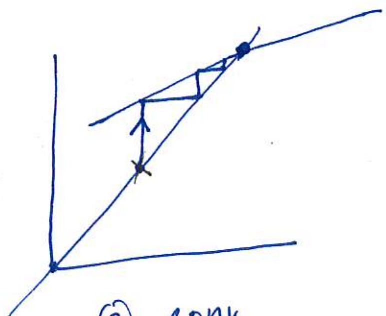
3 conv



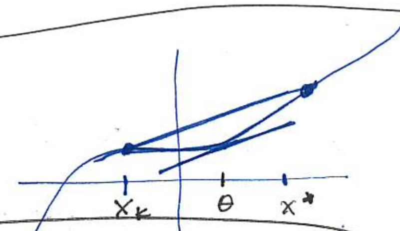
4 conv.



1 diverges



2 conv



locally convergent if $|g'(x^*)| < 1$

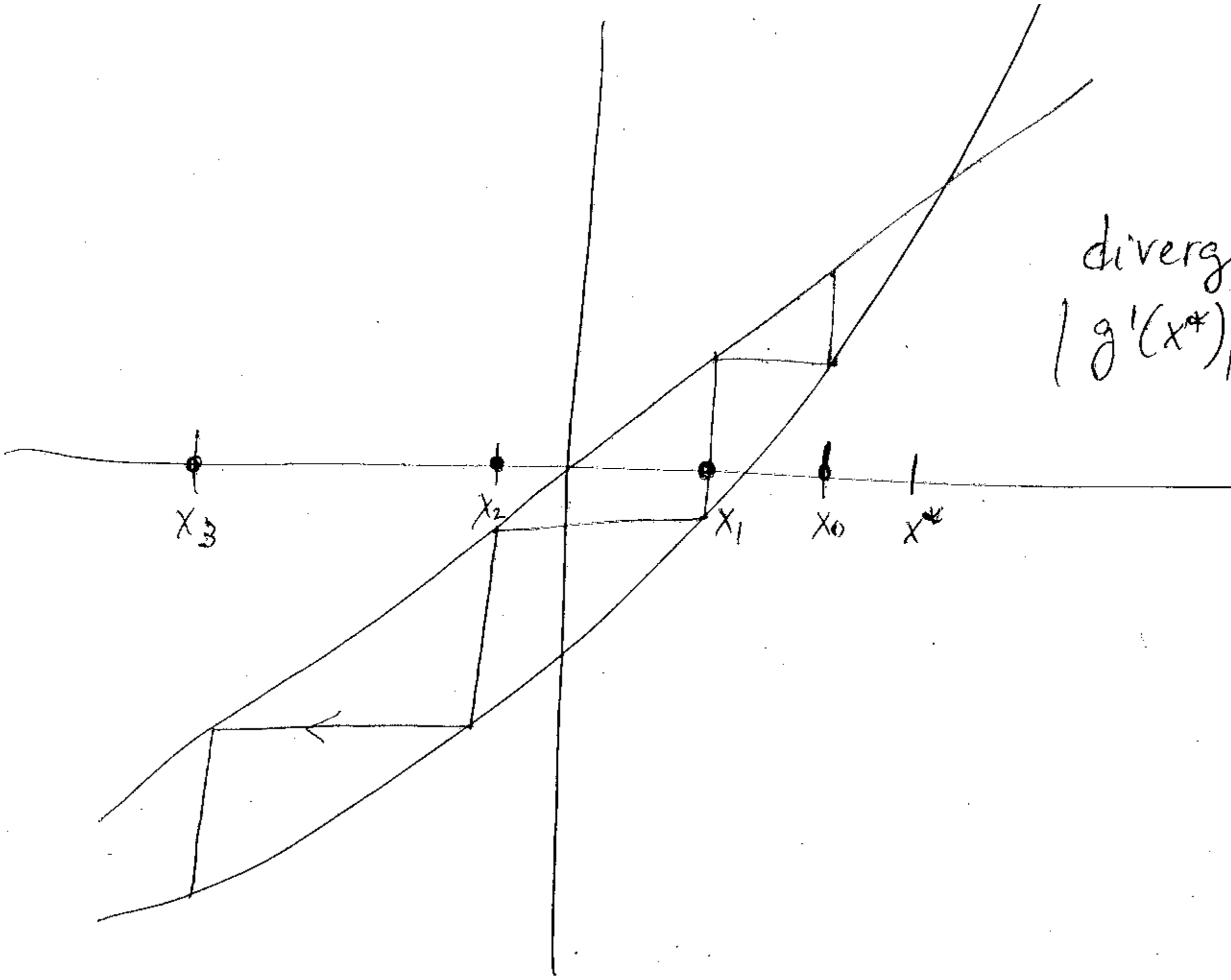
$$e_{k+1} = x_{k+1} - x^* = g(x_k) - g(x^*)$$

By Mean Value Theorem, $\exists$ ~~s.t.~~ θ ~~s.t.~~ between x_k + x^* s.t.

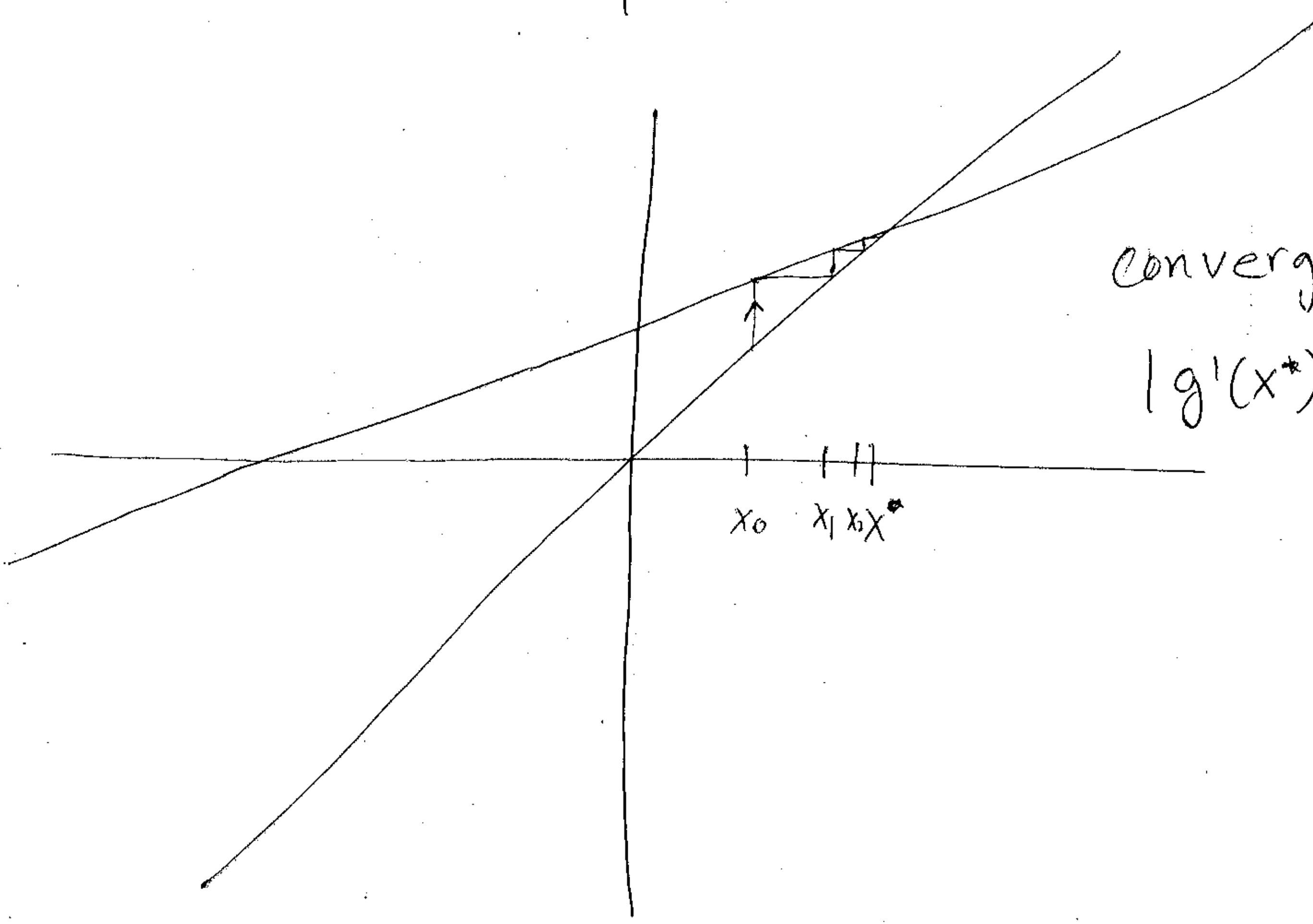
$$g'(\theta) = \frac{g(x_k) - g(x^*)}{x_k - x^*} \Rightarrow e_{k+1} = g'(\theta)\, e_k$$

$$\Rightarrow |e_k| \le c^k |e_0|, \quad c < 1$$

$$\Rightarrow |e_k| \to 0$$

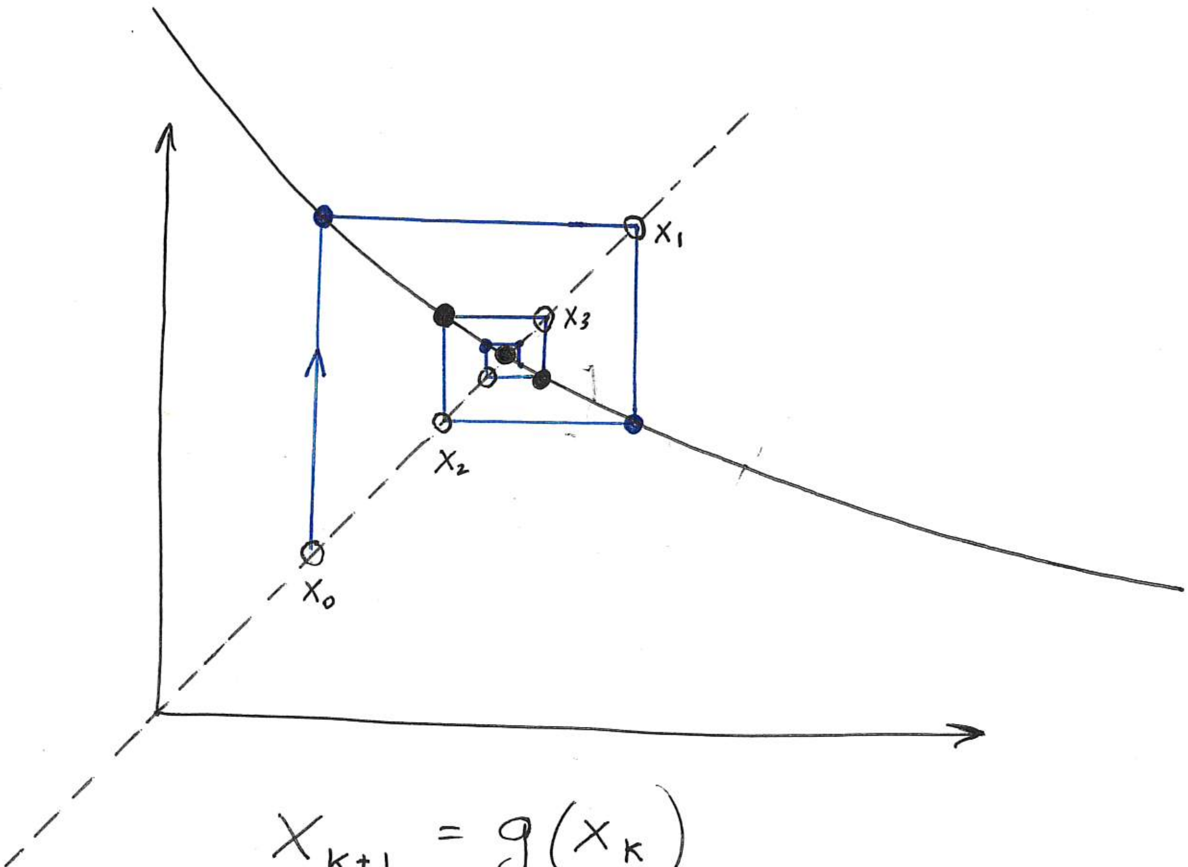


diverges
$|g'(x^*)| > 1$



converges
$|g'(x^*)| < 1$

Fixed Point Iteration



$$x_{k+1} = g(x_k)$$

Convergence rate of fixed point iteration

$$e_{k+1} = x_{k+1} - x^* = g(x_k) - x^* = g(x_k) - g(x^*)$$

$$e_k = x_k - x^*$$

$$\boxed{\frac{e_{k+1}}{e_k} = \frac{g(x_k) - g(x^*)}{x_k - x^*}}$$

$$g(x_k) = g(x^*) + g'(x^*)(x_k - x^*) + \frac{1}{2} g''(x^*)(x_k - x^*)^2 + \ldots$$

$$\Rightarrow g(x_k) - g(x^*) = g'(x^*)(x_k - x^*) + \frac{1}{2} g''(x^*)(x_k - x^*)^2 + \ldots$$

$$\frac{e_{k+1}}{e_k} = g'(x^*) \frac{\cancel{(x_k - x^*)}}{\cancel{(x_k - x^*)}} + \frac{1}{2} g''(x^*) \frac{(x_k - x^*)^{\cancel{2}}}{\cancel{(x_k - x^*)}} + O((x_k - x^*$$

$$\boxed{\frac{e_{k+1}}{e_k} = g'(x^*) + \frac{1}{2} g''(x^*)(x_k - x^*) + O((x_k - x^*)^2)}$$

if $g'(x^*) = 0$,

$$\frac{e_{k+1}}{e_k} = \frac{1}{2} g''(x^*)(x_k - x^*) + \frac{1}{3!} g'''(x^*)(x_k - x^*)^2 + \ldots$$

$$\Rightarrow \frac{e_{k+1}}{e_k^2} = \frac{1}{2} g''(x^*) + \frac{1}{3!} g'''(x^*)(x_k - x^*) + \ldots$$

$$\boxed{\lim_{k \to \infty} \frac{e_{k+1}}{e_k^2} = \frac{1}{2} g''(x^*)}$$

quadratic convergence.

is linear with $C = |g'(x^*)|$

$\boxed{\begin{array}{l} \text{Ideally have } g'(x^*) = 0 \\ \Rightarrow \quad g(x_k) - g(x^*) = \dfrac{g''(\xi_k)}{2}(x_k - x^*)^2 \\ \Rightarrow \quad \dfrac{g(x_k) - g(x^*)}{(x_k - x^*)^2} = \dfrac{g''(\xi_k)}{2} = \dfrac{e_{k+1}}{e_k^2} \\ \qquad \lim\limits_{k\to\infty} \dfrac{\|e_{k+1}\|}{\|e_k\|^2} = \dfrac{g''(x^*)}{2} \quad \text{. quadratic convergence.} \end{array}}$

Example 5.9

1. $g'(x) = 2x$

   $g'(2) = 4 \quad \Rightarrow$ diverges

2. $g'(x) = \frac{1}{2}(x+2)^{-1/2}$

   $g'(2) = \frac{1}{2}(4)^{-1/2} = \frac{1}{4} \Rightarrow$ converges $\quad C = \frac{1}{4}$

   +ive sign $\Rightarrow$ iteration approaches from one side.

3. $g'(x) = -2x^{-2}$

   $g'(2) = -\frac{2}{2^2} = -\frac{1}{2} \Rightarrow$ converges $\quad C = +\frac{1}{2}$

   −ive sign $\Rightarrow$ spiral

4. $g'(x) = \dfrac{2x^2-2x-4}{(2x-1)^2} \Rightarrow$ converges quadratically.

   $g'(2) = 0$

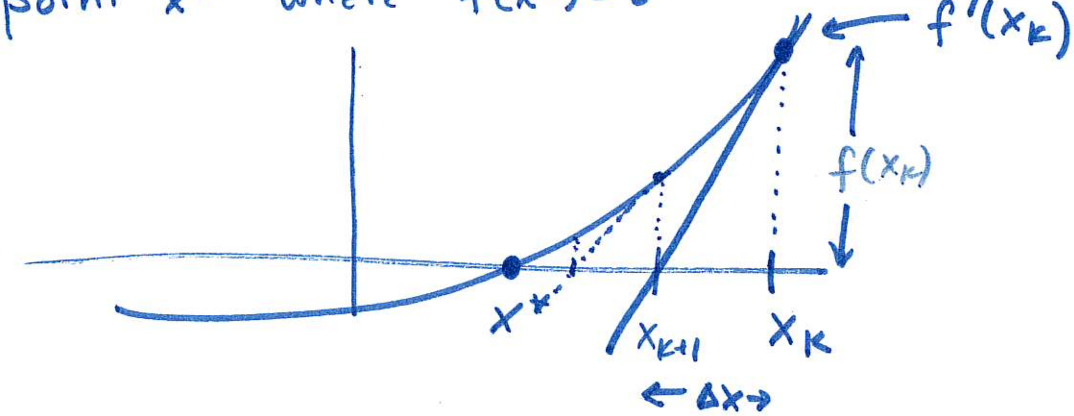
$$g(x_k) = g(x^*) + \frac{(x_k - x^*)}{1} g'(x^*) + \frac{(x_k-x^*)^2}{2} g''(x^*) + \dots$$

Taylor's theorem

$$g(x_k) = g(x^*) + \cancel{(x_k - x^*) g'(x^*)}^{\,0} + \frac{(x_k - x^*)^2}{2} g''(\theta)$$

Newton's Method

First, let's look at Newton's Method for finding a root of scalar $f(x)$, i.e. point x^* where $f(x^*) = 0$



$$\frac{f(x_k)}{\Delta x} = f'(x_k)$$

want Δx which satisfies this

$$\Rightarrow \quad \Delta x = \frac{f(x_k)}{f'(x_k)}$$

$$x_{k+1} = x_k - \Delta x = x_k - \frac{f(x_k)}{f'(x_k)}$$

Newton Step

Newton's method for finding a root

x_0
for $k = 0, 1, 2, \ldots$
 $x_{k+1} = x_k - f(x_k) / f'(x_k)$
end

---

we can also derive this algebraically from Taylor series:

$$f(x+h) \approx f(x) + h f'(x) = 0$$

$$\Rightarrow \quad h = \frac{-f(x)}{f'(x)} \quad \Rightarrow \quad x_{k+1} = x_k + h_k$$

$$f(x+h) = f(x) + h f'(x) + \frac{h^2}{2} f''(x) + \frac{h^3}{3!} f'''(x) + \dots$$

derivations of Newton's Method

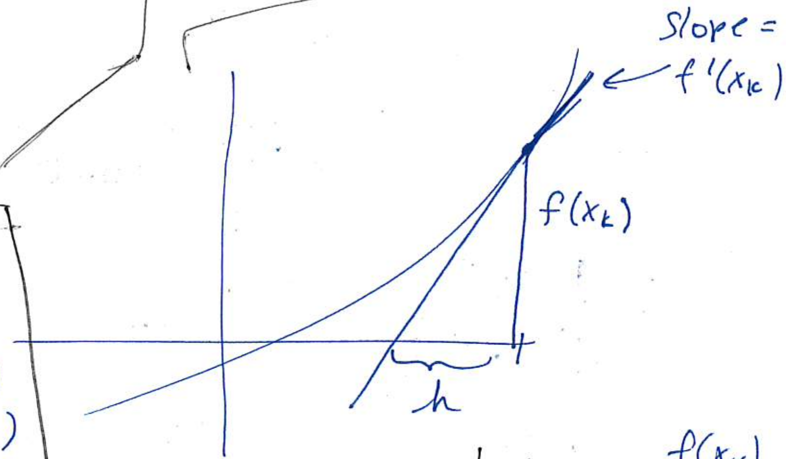
truncate

$$f(x+h) \cong f(x) + h f'(x)$$

$$\Rightarrow \quad h = \frac{f(x+h) - f(x)}{f'(x)}$$

want $f(x+h) = 0$

$$\Rightarrow \quad h = -\frac{f(x)}{f'(x)}$$



$$f'(x_k) = \frac{f(x_k)}{h}$$

Algorithm

x_0 = initial guess
for $k = 1, 2, \dots$
$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$
end

transforms into a fixed pt iteration

$$g(x) = x - \frac{f(x)}{f'(x)}$$

$$= x - f(x)\left(f'(x)\right)^{-1}$$

$$g'(x) = 1 - f'(x)(f'(x))^{-1} + f(x)\,(f'(x))^{-2} f''(x)$$
$$= \cancel{1 - 1} + \frac{f(x) f''(x)}{f'(x)^2}$$

if x^* is a simple root $\Rightarrow$ $g'(x) = 0$ $\Rightarrow$ quadratic convergence

$$\begin{cases} f(x) = 0 \\ f'(x) \,!= 0 \end{cases}$$

if x^* is a multiple root $\Rightarrow$ linear conv
multiplicity m

$$C = 1 - \frac{1}{m}$$

~~m = mul~~

See example 5.11

Newton's Method Convergence Rate

$$\frac{e_{k+1}}{e_k} = \frac{x_{k+1} - x^*}{x_k - x^*} = \frac{g(x_k) - g(x^*)}{x_k - x^*}$$

$$= \frac{\cancel{g(x^*)} + (x_k - x^*) g'(x^*) + (x_k - x^*)^2 g''(\theta) - \cancel{g(x^*)}}{x_k - x^*}$$

$$\Rightarrow \boxed{\frac{e_{k+1}}{e_k} = g'(x^*) + (x_k - x^*) g''(\theta)}$$

$$g(x) = x - f(x)[f'(x)]^{-1} = x - \frac{f(x)}{f'(x)}$$

$$g'(x) = 1 - \left[\frac{f'(x)^2 - f(x) f''(x)}{f'(x)^2}\right]$$

If $f(x^*) = 0$ with multiplicity 1, i.e., $f'(x^*) \neq 0$, then

$$g'(x^*) = 1 - [1 - 0] = 0 \Rightarrow \text{\underline{quadratic conv.}}$$

If $f(x^*) = 0$ with multiplicity m, i.e., $f(x) \sim x^m$ <u>near x^*</u>

$$g'(x) = 1 - \left[\frac{(m x^{m-1})^2 - x^m m(m-1) x^{m-2}}{(m x^{m-1})^2}\right]$$

$$= 1 - \left[\frac{m^2 \cancel{x^{2m-2}} - \cancel{x^{2m-2}}(m)(m-1)}{m^2 \cancel{x^{2m-2}}}\right] = 1 - \left[\frac{\cancel{m^2} - \cancel{m^2} + m}{m^2}\right]$$

$$= 1 - \frac{1}{m}$$

Multi-dimensional fixed point iteration

$$\vec{g} : \mathbb{R}^n \to \mathbb{R}^n$$

$$\vec{x} = \vec{g}(\vec{x})$$

$$\vec{x}_{k+1} = \vec{g}(\vec{x}_k)$$

convergence if started close to solution and

$$\rho(J(x^*)) < 1$$

- the smaller ρ, the faster the convergence

$J(x^*) = \vec{0} \Rightarrow$ convergence rate at least quadratic

Newton's Method

x_0 = initial guess
for $k = 0, 1, 2, \ldots$
    solve $J(x_k)\vec{s}_k = -\vec{f}(\vec{x}_k)$ Newton step
    $\vec{x}_{k+1} = \vec{x}_k + \vec{s}_k$ update solution
end

Optimization

- objective (function)
- constrained vs. unconstrained
  - feasible choices (or points)

$f: \mathbb{R}^n \to \mathbb{R}$, $S \subseteq \mathbb{R}^n$

find x^* in S x^* "minimizer"

"minimum"

(max f is min of $-f$ $\to$ consider only minimization)

f objective function (linear or nonlinear usually differentiable).

S constraints
inequalities, or equalities.

$x \in S \Rightarrow x$ "feasible"

$S = \mathbb{R}^n \Rightarrow$ "unconstrained"

$$\min_x f(x)$$

subj. to $g(x) = 0$, and

$h(x) \leq 0$.

CLASSIFICATION

f, g, h linear or affine $\to$ linear programming
any of f, g, h nonlinear $\Rightarrow$ nonlinear programming

Any

is a global min.

strictly convex $\Rightarrow$ unique

§6.2.2 Unconstrained Optimality Conditions

Scalar case:

$f'(x) = 0$ $\quad$ $f''(x) > 0$ $\quad$ min

$\qquad\qquad\quad$ $f''(x) < 0$ $\quad$ max

$\qquad\qquad\quad$ $f''(x) = 0$ $\quad$ ~~inflection pt~~ / inconclusive $\quad$ x^3 (inflection)

x^4 (min)

$-x^4$ (max)

Vector case:

$f(x)$, $x \in \mathbb{R}^n$

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

gradient of f.

∇f points uphill

$-\nabla f$ points downhill

$$\boxed{\cancel{f(x+s) = f(x) + \nabla f(x+\alpha s)^T s \quad \text{for some } \alpha \in (0,1)} \\ \cancel{\text{choose } s = -\nabla f}}$$

First order necessary condition

Taylor's theorem

$$f(x+s) = f(x) + \nabla f(x)^T s + \quad \alpha \in (0,1)$$
$$+ \frac{1}{2} s^T (H) s + \ldots$$

let $s = -\alpha \nabla f(x)$

stationary pt.
equilibrium pt.

$$f(x - \alpha\nabla f) = f(x) - \alpha \nabla f^T \nabla f + \frac{\alpha^2}{2} \nabla f^T H \nabla f + \ldots$$

$< f(x)$ for some $\alpha \in (0,1)$.

unless $\nabla f = 0$ $\rightarrow$ $= f(x) + \frac{\alpha^2}{2} \nabla f^T H \nabla f$ $\quad s^T H s$

system of nonlinear equations.

x is a "critical point" ← necessary, but not sufficient

"stationary point"

– x may be min, max, or neither (saddle pt.).

$f : \mathbb{R}^n \to \mathbb{R}$ twice differentiable

Hessian matrix of f

$H_f : \mathbb{R}^n \to \mathbb{R}^{n \times n}$

$$H_f(x) = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \ddots & & \\ \vdots & & & \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & & \frac{\partial^2 f(x)}{\partial x_n^2} \end{pmatrix}$$

if 2^{nd} partial derivs of f continuous, then H_f symmetric

Let x^* be a critical pt. of f. + that $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable.

Taylor's theorem, $s \in \mathbb{R}^n$

$$f(x^* + s) = f(x) + \nabla f(x^*)^T s + \frac{1}{2} s^T H_f(x^* + \alpha s) s,$$
$$\alpha \in (0, 1)$$

$H_f(x^*) > 0$ second-order sufficient condition

$\nabla f(x^*) = 0$, + $H_f(x^*)$ is

CLASSIFICATION

- pos. def $\Rightarrow$ x^* is a min of f
- neg. def $\Rightarrow$ x^* is a max of f
- indef $\Rightarrow$ x^* is a saddle pt. of f
- singular $\Rightarrow$ various pathological cases.

Note: $H_f(x^*) > 0$ then f is convex in some nbhd of x^*.

Test for positive definiteness

1. try to compute Cholesky factorization
2. LDL^T

} simple & cheap

3. eigenvalues — expensive!

Example. 6.5 Classifying Critical Pts.

$$f(x) = 2x_1^3 + 3x_1^2 + 12x_1x_2 + 3x_2^2 - 6x_2 + 6$$

$$\nabla f(x) = \begin{pmatrix} 6x_1^2 + 6x_1 + 12x_2 \\ 12x_1 + 6x_2 - 6 \end{pmatrix} = 0$$

Solving $\nabla f(x) = 0$, get $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ -3 \end{pmatrix}$ critical points

$$H_f(x) = \begin{pmatrix} 12x_1 + 6 & 12 \\ 12 & 6 \end{pmatrix} \quad \text{symmetric ✓}$$

saddle $H_f\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right) = \begin{pmatrix} 12+6 & 12 \\ 12 & 6 \end{pmatrix} = \begin{pmatrix} 18 & 12 \\ 12 & 6 \end{pmatrix}$ not p.def ✗, $\lambda \cong 25.4, -1.4$

local min $H_f\left(\begin{pmatrix} 2 \\ -3 \end{pmatrix}\right) = \begin{pmatrix} 30 & 12 \\ 12 & 6 \end{pmatrix}$ pos def ✓, $\lambda \cong 35.0, 1.0$

Minimum Problems: Convexity and Newton's Method (Strang VI.1)

$$F(\vec{x}) = F(x_1, x_2, \ldots, x_n)$$

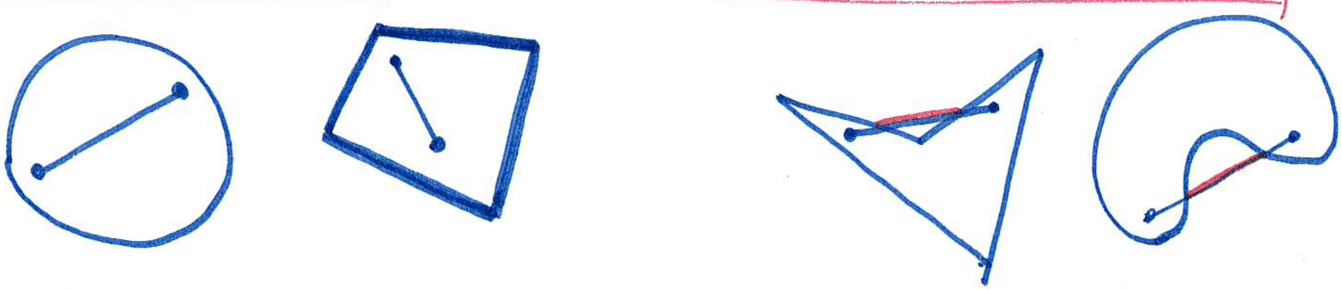
$$\min_x F(x)$$ ← objective function, or cost function

subj. to $G(x) = 0$ ← equality constraints

$H(x) \le 0$ ← inequality constraints

Convexity (sets)

$\boxed{K \text{ convex set}}$ $\boxed{K \text{ non-convex set}}$

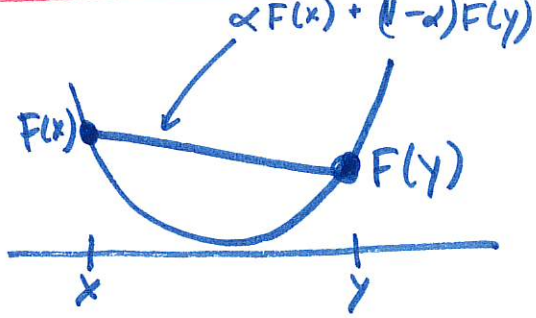


K convex $\quad x, y \in K \Rightarrow \alpha x + (1-\alpha) y \in K \quad \forall\, 0 \le \alpha \le 1$

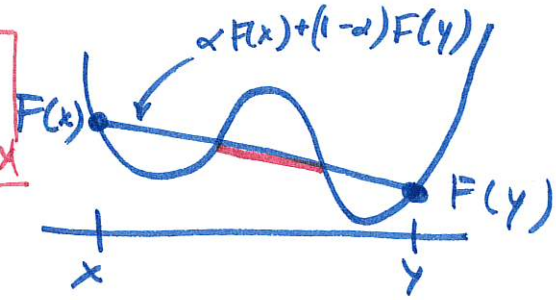
Convexity (function)

F convex function $\quad F(\alpha x + (1-\alpha) y) \le \alpha F(x) + (1-\alpha) F(y)$

$\boxed{F \text{ convex}}$



$\boxed{F \text{ non-convex}}$



Gradient

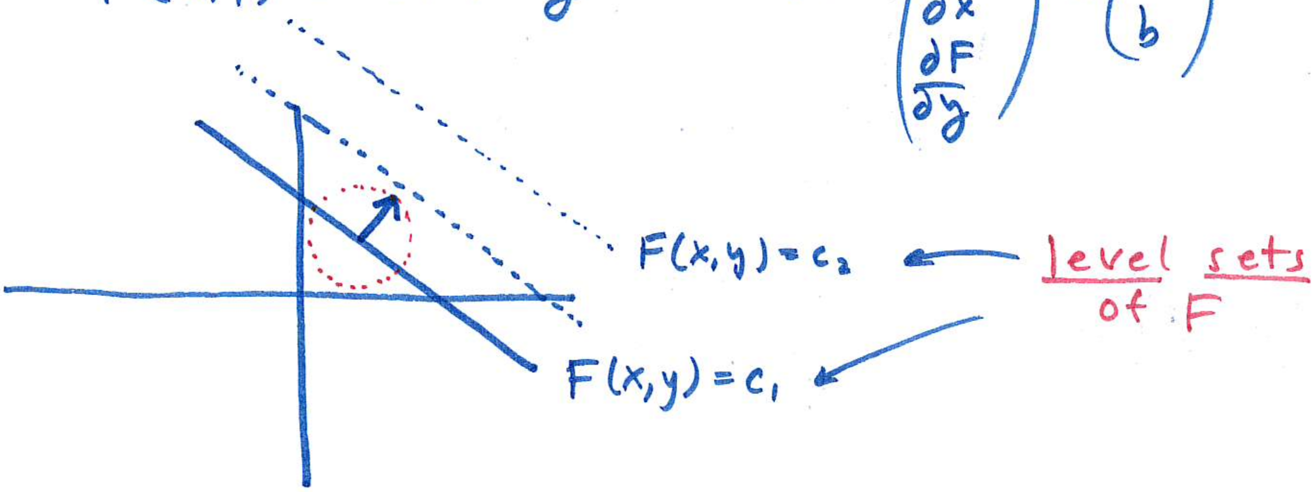
$$F(x_1, x_2, \ldots, x_n)$$

Def'n

$$\nabla F(\vec{x}) = \nabla F(x_1, x_2, \ldots, x_n) = \begin{pmatrix} \frac{\partial F}{\partial x_1}(\vec{x}) \\ \frac{\partial F}{\partial x_2}(\vec{x}) \\ \vdots \\ \frac{\partial F}{\partial x_n}(\vec{x}) \end{pmatrix}$$

Example

$$F(x,y) = ax + by \qquad \nabla F = \begin{pmatrix} \frac{\partial F}{\partial x} \\ \frac{\partial F}{\partial y} \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$$



$$F(x+s) = F(x) + \nabla F(x)^T s + O(\|s\|^2)$$

$\nabla F(x)^T s = 0$ for s tangent to level set

i.e. $\nabla F(x)$ is orthogonal to level set

$\nabla F(x)$ is direction of steepest ascent at x.

$-\nabla F(x)$ " " " " descent at x.

The second derivative matrix

$f(x)$ function of single variable is convex if

$$f''(x) \geq 0$$

example

$f(x)$ convex

For $F(x_1, x_2, \dots, x_n)$, we have a matrix of second derivatives

Hessian

$$H(x) = \begin{pmatrix} \frac{\partial^2 F}{\partial x_1^2} & \frac{\partial^2 F}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 F}{\partial x_1 \partial x_n} \\ \vdots & & \ddots & \\ \frac{\partial^2 F}{\partial x_n \partial x_1} & \frac{\partial^2 F}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 F}{\partial x_n^2} \end{pmatrix}$$

$$H_{ij}(x) = \frac{\partial^2 F}{\partial x_i \partial x_j}$$

$F(x)$ convex $\Leftrightarrow$ $H(x)$ positive semi-definite at all x.

$F(x)$ strictly convex $\Leftrightarrow$ $H(x)$ positive definite at all x.

Example $F(x) = \frac{1}{2} x^T S x$

S s.p.d. $\Rightarrow$ F strictly convex

Newton's Method for finding a minimum

$$\min f(x)$$

the minimum is a stationary point of $f(x)$, i.e., a point where

$$f'(x) = 0$$

N.M. on $f'(x)$ gives

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

We can also derive this from quadratic approx. to f with its Taylor series

$$f(x+h) \approx f(x) + h f'(x) + \frac{h^2}{2} f''(x) = \phi(h)$$

$\phi(h)$ is <u>quadratic</u> function in h.

minimize $\phi(h)$

$$\phi'(h) = f'(x) + h f''(x) = 0$$

$$\Rightarrow h = \frac{-f'(x)}{f''(x)}$$

$$\Rightarrow x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

Newton's Method for function of several variables

$$F(x+s) \cong F(x) + \nabla F(x)^T s + \frac{1}{2} s^T H_f(x) s = \phi(s)$$

Find s where

$$\nabla \phi(s) = 0$$

$$\Rightarrow \quad \nabla F(x)^T + s^T H_f(x) = 0$$

$$\Rightarrow \quad H_f(x) s = -\nabla F(x)$$

$$\boxed{s = -H_f(x)^{-1} \nabla F(x)}$$

Newton's Method for minimizing $F(x)$

x_0 = initial guess

for $k = 0, 1, 2, \dots$

  Solve $H_f(x_k)\, s_k = -\nabla F(x_k)$

  $x_{k+1} = x_k + s_k$

end

Newton's Method

- is a <u>second order</u> method.
  (uses <u>second</u> derivatives of F)

- exhibits <u>quadratic convergence</u>

$$\|x_{k+1} - x^*\| \le C\,\|x_k - x^*\|^2$$

  near the solution x^*

- is done with <u>safeguarding</u> in practice