



AARHUS
UNIVERSITY

Data Analytics and Machine Learning

Newton methods in n D

Henrik Karstoft

Carl Schultz

Alexandros Iosifidis

landscape of approaches for “solving” polynomial equations/inequalities

- we will learn:
 - some of most popular, useful methods
 - situations they are most effective
 - their limitations

machine learning

optimization

communication,
visualization

today: data overload



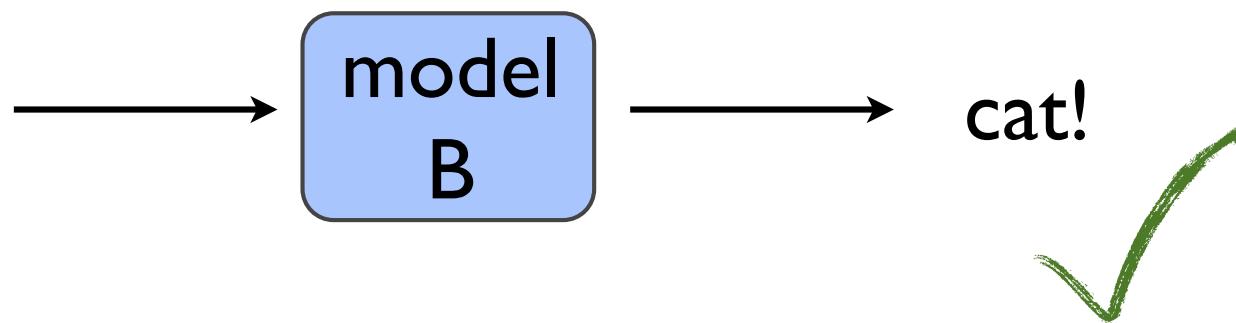
data useless without interpretation

	H	O	P	O	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP				
1	4week	19 week				Growth			4week	19 week						4week	19 week										International equity						
2	Unit Value	Cumulative	4week	Cumulative	19 week			Unit	Unit Value	Cumulative	4week	Cumulative	19 week			Unit	Unit Value	Cumulative	4week	Cumulative	19 week					Unit	Unit Value	Cumulative					
3	%Change	Unit Value	paused	Unit Value	paused			Date	Amount	Value	%Change	Unit Value	paused			Date	Amount	Value	%Change	Unit Value	paused				Date	Amount	Value	%Change					
4	%Change	average	%Chan average	%Chan average	%Chan average			3/31/2006	161,205.64	63.17	%Change	average	%Change	average		3/31/2006	162,078.45	86,4923	%Change	average	%Change	average				Date	Amount	Value	%Change				
5	\$REF!	\$REF!	101,0939	\$REF!	101,0939			12/29/2006	246,555.47	64,802	\$REF!	64,802	\$REF!	64,802		12/29/2006	210,169.10	95,2116	\$REF!	95,2116													
6	0	\$REF!	111.27	9.145	111.27			12/31/2007	93,403.95	75.64	0	\$REF!	75.64	14,328.298	75.64		12/31/2007	392.32	99.68	0	\$REF!	99.68	4,482.7448	99.68									
7	-42.28687	0	74,78495	\$REF!	74,78495			12/31/2008	997.62	45,5532	-39,776.3088	0	53.8758	\$REF!	53.8758		12/31/2008	1,002.93	62,3085	-37,494.482	0	70,95485	\$REF!	70,95485									
8	-23.292442	0	87,1465	\$REF!	87,1465			12/31/2009	307,797.6	62,1984	-17,770.9198	0,000643107	64,636689	\$REF!	64,636689		12/31/2009	67,383.09	79,6642	-20,080.056	0,000251054	84,00541892	\$REF!	84,005419									
9	-19.061472	5,515714244	82,358275	\$REF!	87,1465			4/9/2010	194,338.58	66,0776	-12,6419884	6,237499598	62,509083	\$REF!	64,636689		4/9/2010	192,556.43	86,3398	-13,383.026	8,279495772	81,75790833	\$REF!	84,005419									
10	-19,361552	5,124513049	82,358275	-0.37	87,1465			4/16/2010	194,475.66	66,1762	-12,5116241	6,396,025596	62,509083	0,1492185	64,636689		4/16/2010	245,539.01	86,3663	-13,356.441	8,41210484	81,75790833	0,0364927	84,005419									
11	-18,639346	6,066023027	82,358275	0.896	87,1465			4/23/2010	198,507.8	67,4206	-10,8664742	8,296,733014	62,509083	1,880434	64,636689		4/23/2010	249,415.61	88,3905	-11,25274	10,9541236	81,75790833	2,34273782	84,005419									
12	-20,925137	3,081942761	82,358275	-2.81	87,1465			4/30/2010	193,311.32	65,6577	-13,197177	5,562397505	62,509083	-2,644779	64,636689		4/30/2010	242,950.69	86,0994	-13,624197	8,075178349	81,75790833	-2,592021	84,005419									
13	-27,65966	-4,91926381	87,8914	-7.77	87,1465			5/7/2010	190,419.04	61,6212	-19,0624008	-1,5704685	65,31066	-6,757014	64,636689		5/7/2010	226,651.68	80,2649	-19,477428	0,754293031	85,49218	-6,776747	84,005419									
14	-25,424552	-2,77953194	86,47536	2.251	86,8270875			5/14/2010	195,266.87	62,9662	-16,8876256	1,07431071	64,66839	5,268,69738	64,516843		5/14/2010	232,971.05	82,5028	-17,232343	3,56246656	84,72478	2,7801420	84,232557									
15	-28,65910	-8,25856802	84,36026	-0.41	83,358275			5/21/2010	177,410.73	60,1429	-20,4879693	-3,30412553	63,46715	-4,331889	62,509083		5/21/2010	222,786.10	78,8358	-20,91111	8,1757908	82,21868	-4,444698	8,1757908									
16	-29,955326	-8,63837567	81,8424	-1.53	83,358275			6/4/2010	175,583.19	59,4866	-21,3794288	-4,38824399	61,8752	-1,121613	62,509083		6/4/2010	219,727.16	77,3424	-22,40931	-2,9142381	81,09096	-1,894217	81,757908									
17	-27,595827	-6,08464896	80,2769	2.846	83,358275			6/11/2010	155,405.88	60,6914	-19,7628239	-2,42226438	79,57806	2,0562112	62,509083		6/11/2010	155,643.29	79,3084	-20,434998	-0,44637477	79,65086	2,5419434	81,757908									
18	-27,959827	-6,08464896	80,0779	0	83,358275			6/18/2010	155,405.88	60,6914	-19,7628239	-2,42226438	60,7721	0	62,509083		6/18/2010	155,643.29	79,3084	-20,434998	-0,44637477	79,45956	0	81,757908									
19	-27,899526	-6,06060338	79,572	0.084	83,358275			6/25/2010	153,374.13	59,8476	-20,8783712	-3,7788964	60,16383	-1,39031	62,509083		6/25/2010	153,534.54	78,1681	-21,580959	\$REF!	78,648705	-1,437805	81,757908									
20	-27,899526	-6,06060338	79,74226	0	83,358275			7/2/2010	153,374.13	59,8476	-20,8783712	-3,7788964	60,16383	0	62,509083		7/2/2010	153,534.54	78,1681	-21,580959	-17,077816	78,521025	0	81,757908									
21	-27,330457	-5,2641618	80,326	0.789	83,1640538			7/9/2010	152,708.76	59,5333	-21,2938421	-4,28422136	60,122324	-0,525167	62,280177		7/9/2010	155,475.61	78,0666	-21,580959	-1,3777616	78,738285	-0,129848	81,757908									
22	-24,645277	-1,76362731	81,06362	3.695	83,2147071			7/26/2010	159,220.73	62,0171	-18,0101798	-9,29084536	60,3874	4,1721188	62,261386		7/26/2010	159,349.21	80,9826	-21,632785	-2,00517172	78,4278	3,7352722	81,473962									
23	-23,210824	3,016382356	82,12046	1.904	83,358275			8/2/2010	160,080.00	62,2985	-10,6738145	6,016500788	60,70802	-4,0537458	62,232368		8/8/2010	159,995.83	81,2144	-18,75424	1,655201848	78,84635	3,0227858	81,438864									
24	-26,52174	-4,2101816	82,42702	-4.31	83,2630125			8/15/2010	155,451.96	59,6672	-21,1168694	-4,06894412	60,67274	-4,223367	62,105169		8/15/2010	155,643.29	79,3084	-20,434998	-0,44637477	79,45956	0	81,757908									
25	-26,896648	-4,69843214	82,6502	-0.51	83,1500176			8/22/2010	153,533.50	59,6498	-21,0749603	-4,01794745	60,6452	0,05128	61,960235		8/20/2010	132,67	77,6295	-21,762841	-2,10534244	79,57	-0,458154	81,20931									
26	-27,361643	-5,3042411	82,64263	-0.64	83,02062323			8/20/2010	152,003.68	59,0511	-19,2132055	-5,05948744	60,54656	-5,05948744	60,54656		8/20/2010	268,37	77,2545	-12,1228	-2,55395117	79,460725	-0,483064	81,000259									
27	-24,31446	-1,32325649	82,71688	4.195	83,0837			9/3/2010	161,146.41	61,3642	-18,17067511	-1,35718162	60,4164	3,926,72612	61,775863		9/3/2010	278.36	80,259	-22,497494	-3,02457823	78,5287	3,3898938	80,792161									
28	-24,031816	-3,598385056	82,724211	0.973	83,1640538			9/10/2010	100,352.45	61,52729	-16,8648253	1,02765871	61,52729	-16,8648253	61,52729		9/10/2010	414.09	80,4811	-18,409347	0,746886925	79,282425	2,076219	81,454339									
29	-22,730740	0,7325227403	83,337788	1.713	83,29551035			9/17/2010	105,834.67	62,8239	-16,8648253	1,02765871	61,508652	2,03037872	61,343612		9/17/2010	210,834.67	81,8031	-19,26053	0,1025628468	78,804025	1,6426217	80,127393									
30	-20,932373	3,076887027	84,7052	2.328	82,4610316			9/24/2010	105,705.13	64,5081	-14,769487	3,714105653	61,56746	2,582,951526	61,2090837		9/24/2010	566.79	83,5524	-17,92429	2,685152461	79,949425	2,13965	79,761417									
31	-20,599353	3,510952055	86,21098	0.421	82,409105			10/1/2010	108,462.63	64,3642	-14,907917	3,482746841	62,930908	-2,23073073	61,1471598		10/1/2010	567.17	83,6098	-16,1737192	4,982255473	81,524105	0,06630349	79,619792									
32	-19,080446	5,584842343	93,39088	0.024	82,9519842			10/8/2010	110,529.34	65,5164	-13,38923923	5,335219782	62,7601	1,7901253	61,367821		10/8/2010	711.61	85,0218	-16,322793	4,95179755	82,33616903	0,0900653	79,805744									
33	-18,012852	6,882250945	88,73034	1.229	83,33860368			10/15/2010	112,862.10	64,8931	-11,5638551	7,548635004	64,83214	2,0103059	61,5797463		10/15/2010	718.87	85,8895	-14,70													

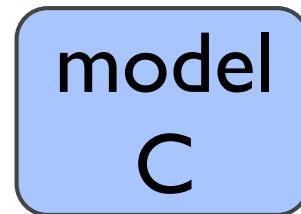
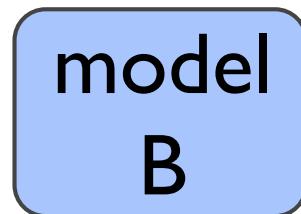
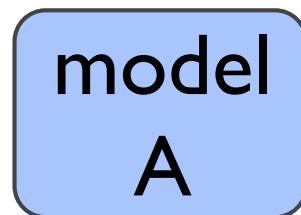
models can “find patterns”



image



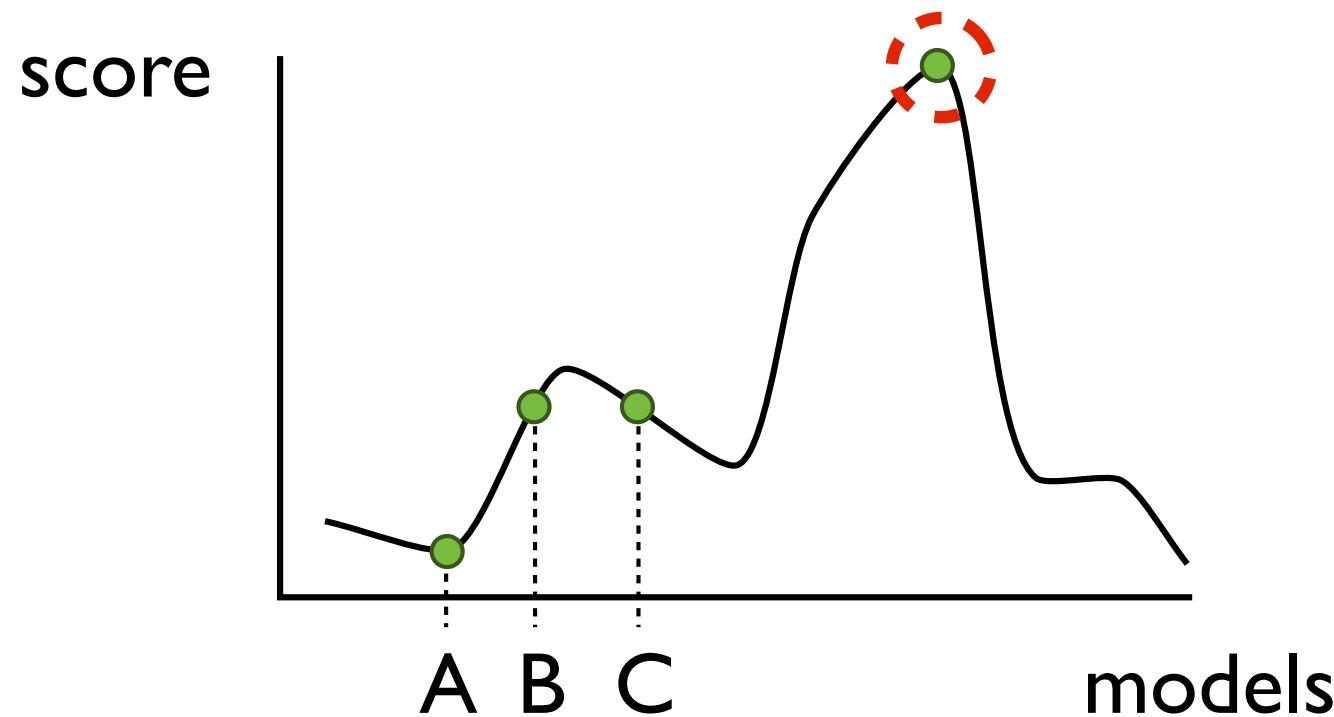
imagine set of **all possible** models



...

let's give each model a **score**

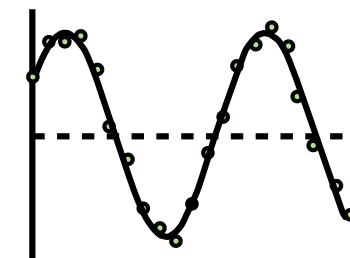
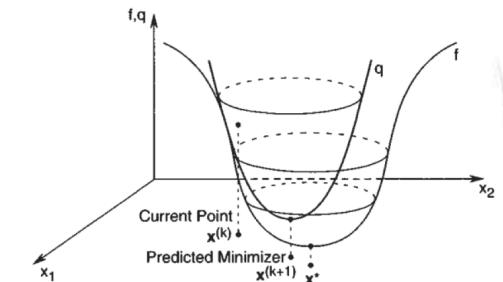
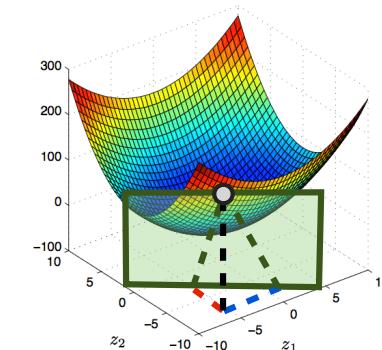
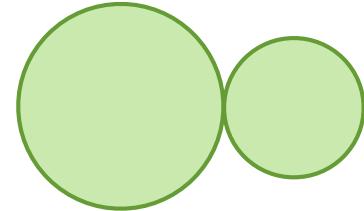
we want **best model** we can find



optimization: find model that maximises score

TODAY'S OUTLINE

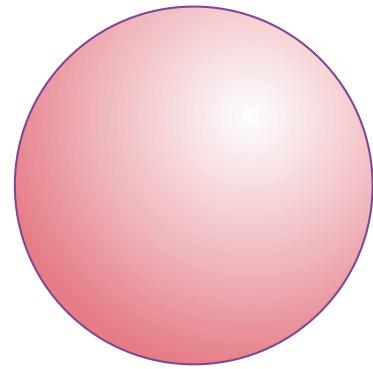
- preliminaries
- Newton from **1D** to **nD**
- *from root finding to optimization*
- *from Jakobian to Hessian*



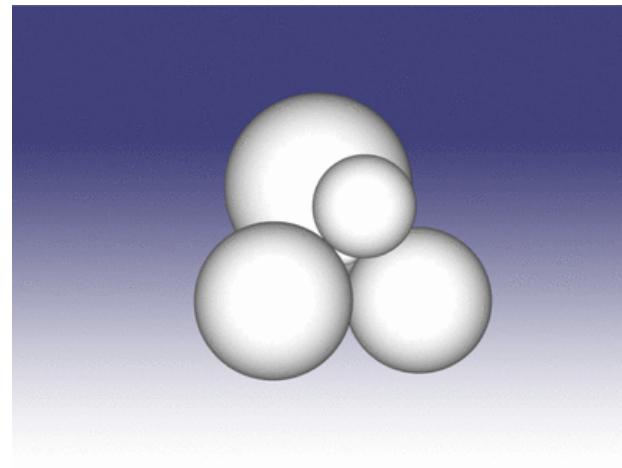
Part I

some preliminary concepts

spatial puzzle solving

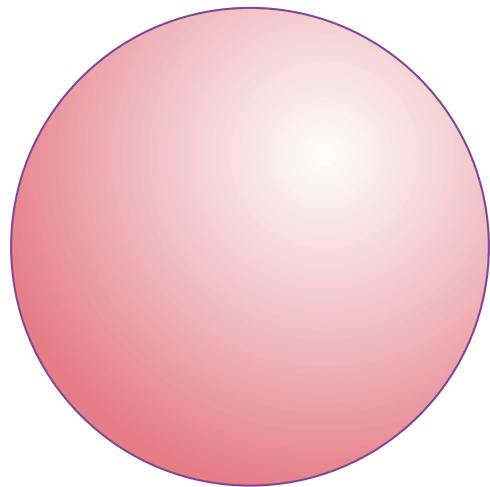


how many spheres (of any size)
can be mutually touching?

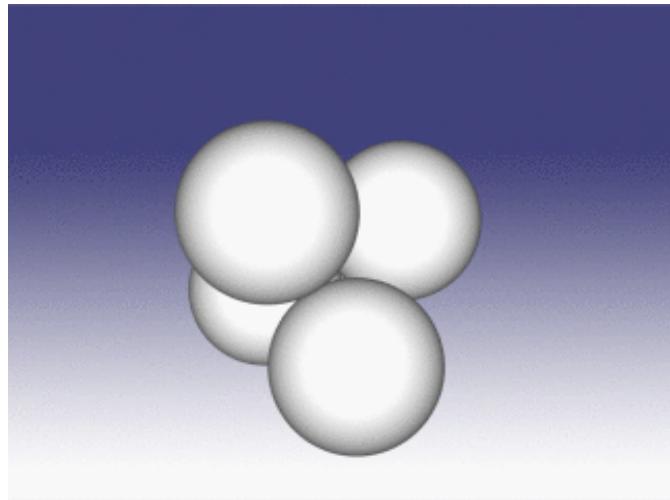


5 spheres

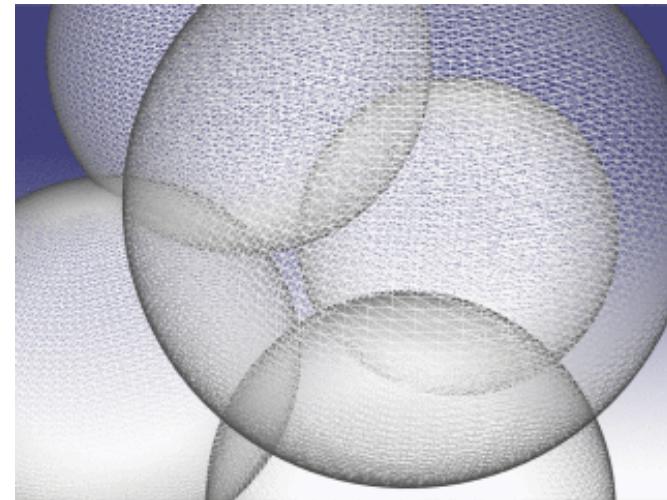
spatial puzzle solving



how many **same-sized** spheres
can be mutually touching?

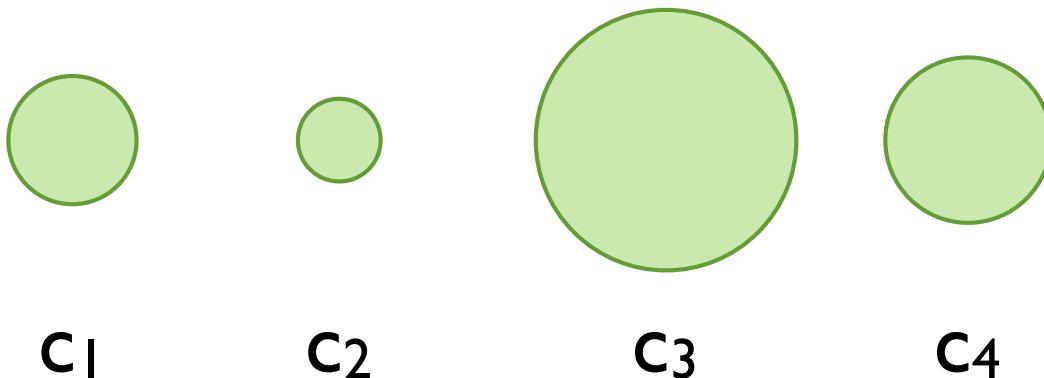


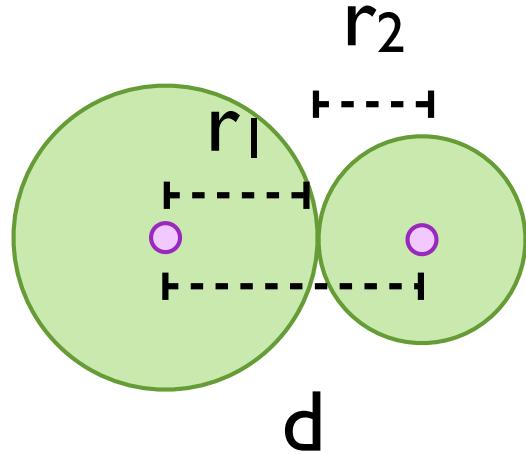
4 spheres



5 spheres

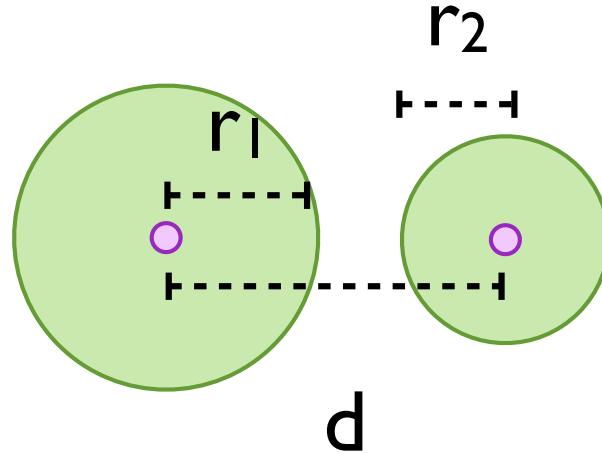
can four circles all touch each other?





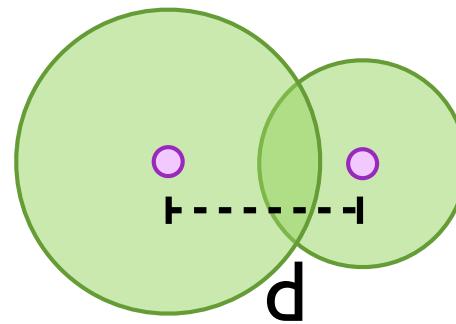
2 circles “touch” IF
(distance between centres) EQUALS (sum of radii)

$$d = (r_1 + r_2)$$



2 circles “touch” IF
(distance between centres) EQUALS (sum of radii)

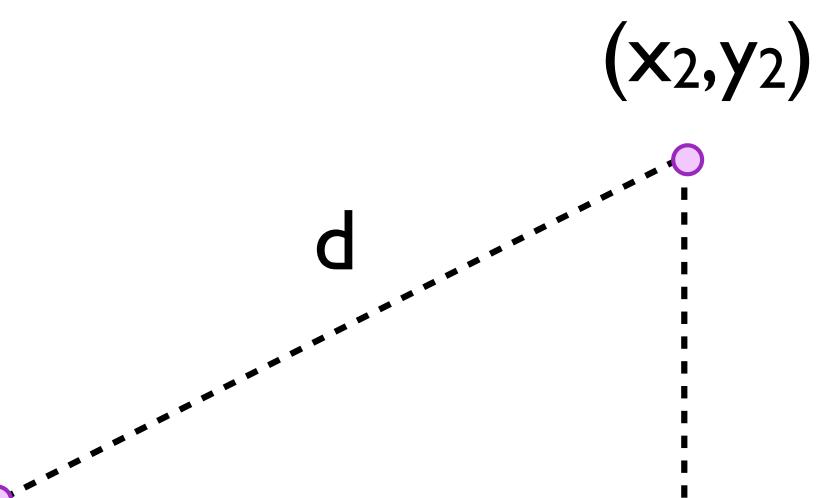
$$d = (r_1 + r_2)$$



$$r_1 \quad r_2$$
$$\text{---} \quad \text{---}$$

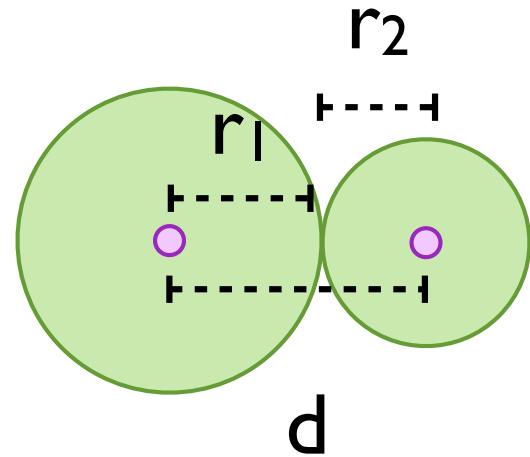
2 circles “touch” IF
(distance between centres) EQUALS (sum of radii)

$$d = (r_1 + r_2)$$

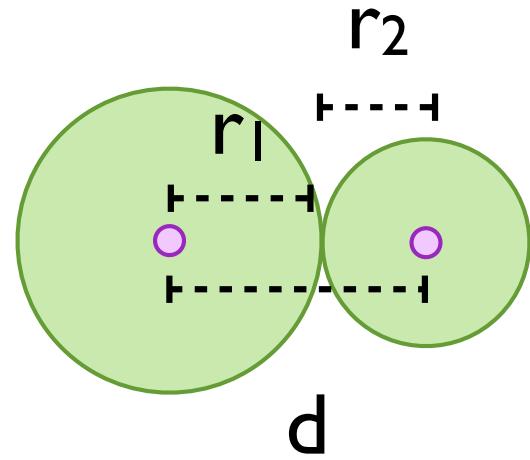


$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

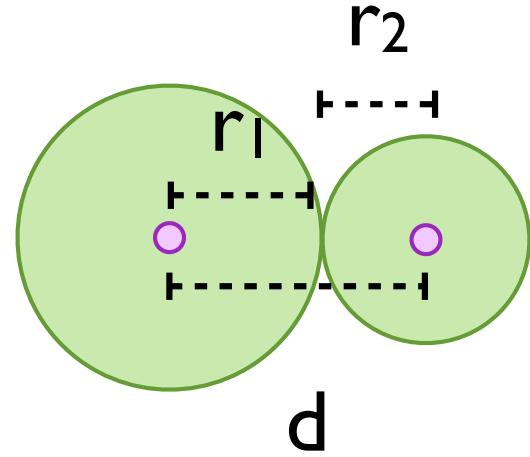
$$d^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2$$



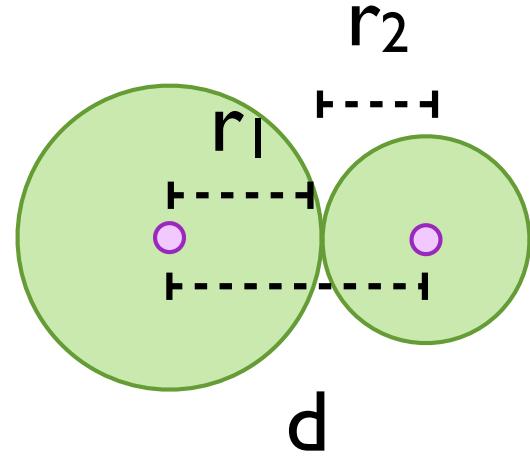
$$d = (r_1 + r_2)$$



$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} = (r_1 + r_2)$$

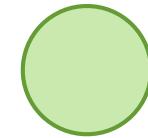
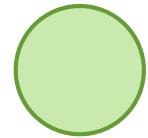
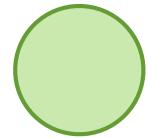


$$(x_1 - x_2)^2 + (y_1 - y_2)^2 = (r_1 + r_2)^2$$



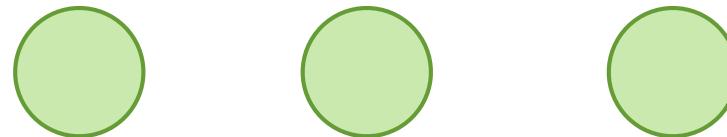
$$(x_1 - x_2)^2 + (y_1 - y_2)^2 - (r_1 + r_2)^2 = 0$$

can three **same-sized** circles all touch each other?



?

can three same-sized circles all touch each other?



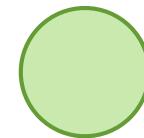
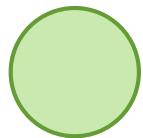
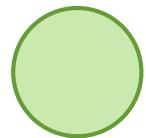
C_1 C_2 C_3

C_1 touches C_2

C_1 touches C_3

C_2 touches C_3

can three same-sized circles all touch each other?

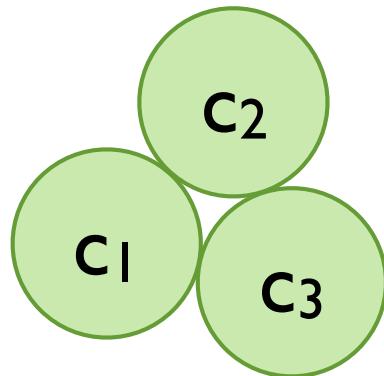


$$(x_1 - x_2)^2 + (y_1 - y_2)^2 - (r+r)^2 = 0$$

$$(x_1 - x_3)^2 + (y_1 - y_3)^2 - (r+r)^2 = 0$$

$$(x_2 - x_3)^2 + (y_2 - y_3)^2 - (r+r)^2 = 0$$

can three same-sized circles all touch each other?

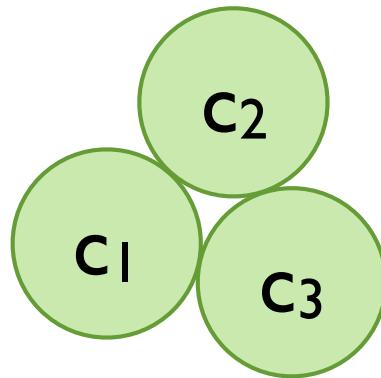


$$(x_1 - x_2)^2 + (y_1 - y_2)^2 - (r+r)^2 = 0$$

$$(x_1 - x_3)^2 + (y_1 - y_3)^2 - (r+r)^2 = 0$$

$$(x_2 - x_3)^2 + (y_2 - y_3)^2 - (r+r)^2 = 0$$

can three same-sized circles all touch each other?



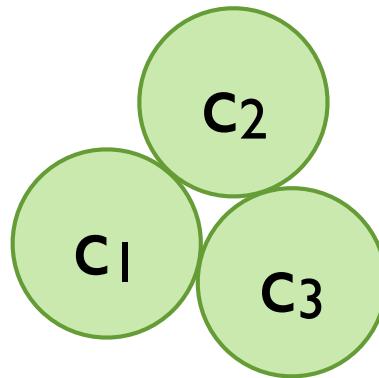
polynomial expression

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 - (r+r)^2 = 0$$

$$(x_1 - x_3)^2 + (y_1 - y_3)^2 - (r+r)^2 = 0$$

$$(x_2 - x_3)^2 + (y_2 - y_3)^2 - (r+r)^2 = 0$$

can three same-sized circles all touch each other?



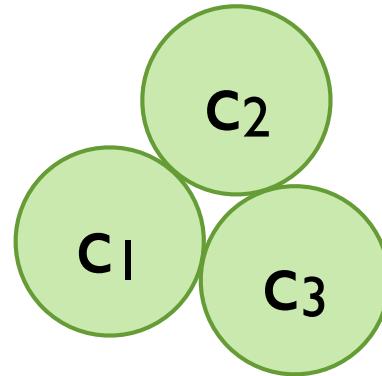
number of variables?

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 - (r+r)^2 = 0$$

$$(x_1 - x_3)^2 + (y_1 - y_3)^2 - (r+r)^2 = 0$$

$$(x_2 - x_3)^2 + (y_2 - y_3)^2 - (r+r)^2 = 0$$

can three same-sized circles all touch each other?



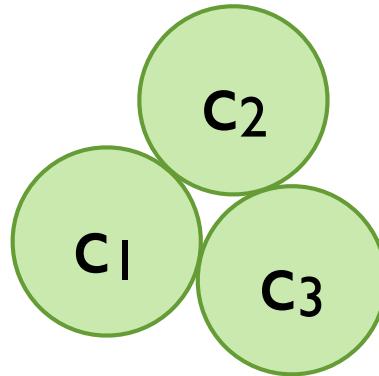
5 variables (“multi-variate” expression)

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 - (r+r)^2 = 0$$

$$(x_1 - x_3)^2 + (y_1 - y_3)^2 - (r+r)^2 = 0$$

$$(x_2 - x_3)^2 + (y_2 - y_3)^2 - (r+r)^2 = 0$$

can three same-sized circles all touch each other?



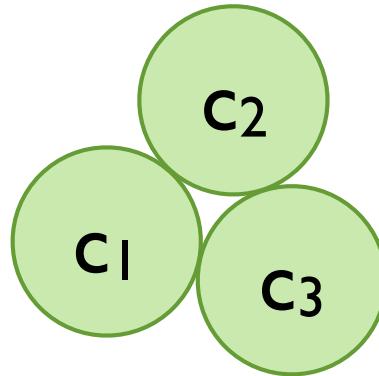
5 “dimensions”

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 - (r+r)^2 = 0$$

$$(x_1 - x_3)^2 + (y_1 - y_3)^2 - (r+r)^2 = 0$$

$$(x_2 - x_3)^2 + (y_2 - y_3)^2 - (r+r)^2 = 0$$

can three same-sized circles all touch each other?



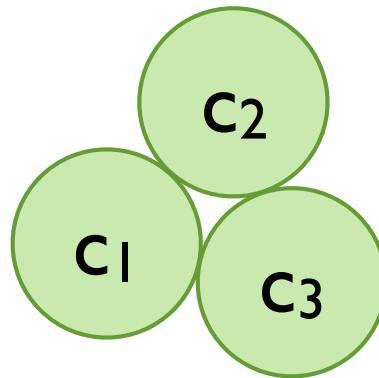
linear?

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 - (r+r)^2 = 0$$

$$(x_1 - x_3)^2 + (y_1 - y_3)^2 - (r+r)^2 = 0$$

$$(x_2 - x_3)^2 + (y_2 - y_3)^2 - (r+r)^2 = 0$$

can three same-sized circles all touch each other?



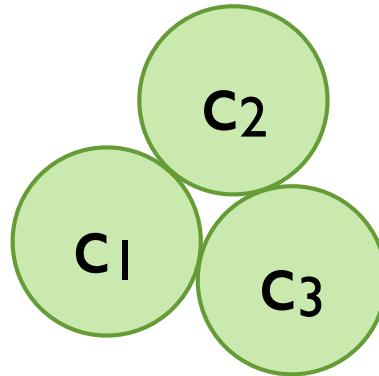
non-linear polynomial expression

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 - (r+r)^2 = 0$$

$$(x_1 - x_3)^2 + (y_1 - y_3)^2 - (r+r)^2 = 0$$

$$(x_2 - x_3)^2 + (y_2 - y_3)^2 - (r+r)^2 = 0$$

can three same-sized circles all touch each other?



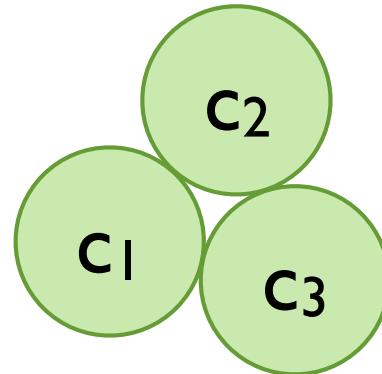
polynomial equation

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 - (r+r)^2 = 0$$

$$(x_1 - x_3)^2 + (y_1 - y_3)^2 - (r+r)^2 = 0$$

$$(x_2 - x_3)^2 + (y_2 - y_3)^2 - (r+r)^2 = 0$$

can three same-sized circles all touch each other?



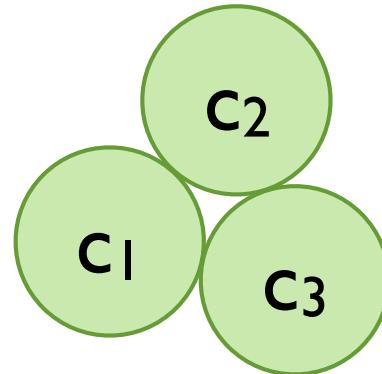
$$(x_1 - x_2)^2 + (y_1 - y_2)^2 - (r+r)^2 = 0$$

$$(x_1 - x_3)^2 + (y_1 - y_3)^2 - (r+r)^2 = 0$$

$$(x_2 - x_3)^2 + (y_2 - y_3)^2 - (r+r)^2 = 0$$

system of polynomial equations

can three same-sized circles all touch each other?



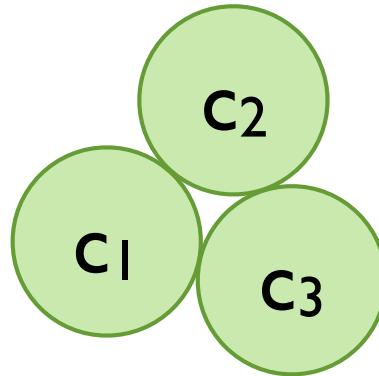
$$(x_1 - x_2)^2 + (y_1 - y_2)^2 - (r+r)^2 = 0$$

$$(x_1 - x_3)^2 + (y_1 - y_3)^2 - (r+r)^2 = 0$$

$$(x_2 - x_3)^2 + (y_2 - y_3)^2 - (r+r)^2 = 0$$

number of variables?

can three same-sized circles all touch each other?



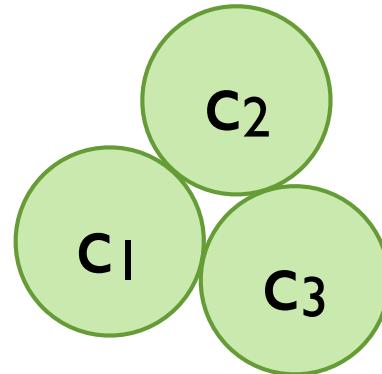
$$(x_1 - x_2)^2 + (y_1 - y_2)^2 - (r+r)^2 = 0$$

$$(x_1 - x_3)^2 + (y_1 - y_3)^2 - (r+r)^2 = 0$$

$$(x_2 - x_3)^2 + (y_2 - y_3)^2 - (r+r)^2 = 0$$

7 variables
(7 “dimensions”)

can three same-sized circles all touch each other?



$$(x_1 - x_2)^2 + (y_1 - y_2)^2 - (r+r)^2 = 0$$

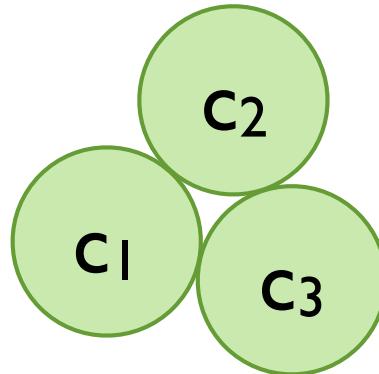
$$(x_1 - x_3)^2 + (y_1 - y_3)^2 - (r+r)^2 = 0$$

$$(x_2 - x_3)^2 + (y_2 - y_3)^2 - (r+r)^2 = 0$$

our task: **solving** a system of equations

what does “solving” mean?

can three same-sized circles all touch each other?



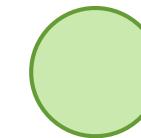
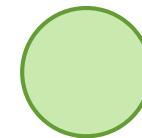
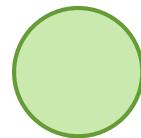
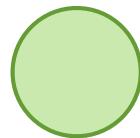
$$(x_1 - x_2)^2 + (y_1 - y_2)^2 - (r+r)^2 = 0$$

$$(x_1 - x_3)^2 + (y_1 - y_3)^2 - (r+r)^2 = 0$$

$$(x_2 - x_3)^2 + (y_2 - y_3)^2 - (r+r)^2 = 0$$

our task: finding roots of system of equations

can four **same-sized** circles all touch each other?



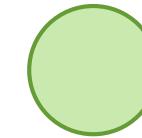
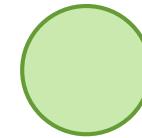
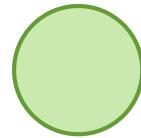
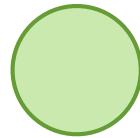
C1

C2

C3

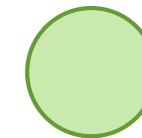
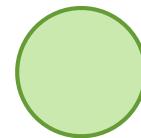
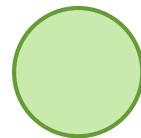
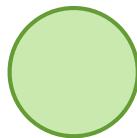
C4

can four **same-sized** circles all touch each other?



no.

can four same-sized circles all touch each other?



C1

C2

C3

C4

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 - (r+r)^2 = 0$$

$$(x_1 - x_3)^2 + (y_1 - y_3)^2 - (r+r)^2 = 0$$

$$(x_1 - x_4)^2 + (y_1 - y_4)^2 - (r+r)^2 = 0$$

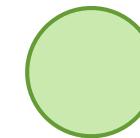
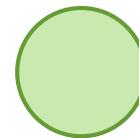
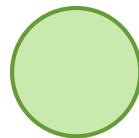
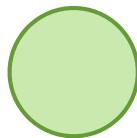
$$(x_2 - x_3)^2 + (y_2 - y_3)^2 - (r+r)^2 = 0$$

$$(x_2 - x_4)^2 + (y_2 - y_4)^2 - (r+r)^2 = 0$$

$$(x_3 - x_4)^2 + (y_3 - y_4)^2 - (r+r)^2 = 0$$

no solution to this system of equations

can four same-sized circles all touch each other?



C1

C2

C3

C4

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 - (r+r)^2 = 0$$

$$(x_1 - x_3)^2 + (y_1 - y_3)^2 - (r+r)^2 = 0$$

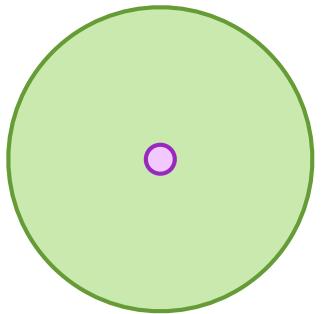
$$(x_1 - x_4)^2 + (y_1 - y_4)^2 - (r+r)^2 = 0$$

$$(x_2 - x_3)^2 + (y_2 - y_3)^2 - (r+r)^2 = 0$$

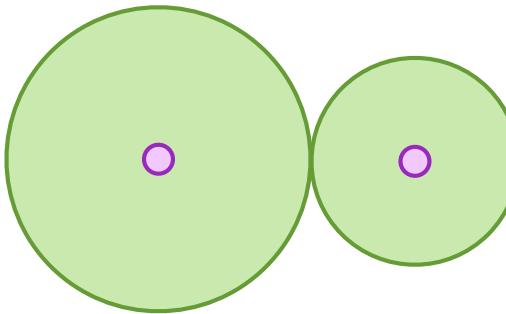
$$(x_2 - x_4)^2 + (y_2 - y_4)^2 - (r+r)^2 = 0$$

$$(x_3 - x_4)^2 + (y_3 - y_4)^2 - (r+r)^2 = 0$$

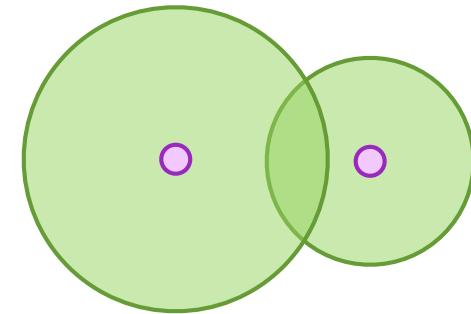
how “close” can we get?



$$d > (r_1 + r_2)$$

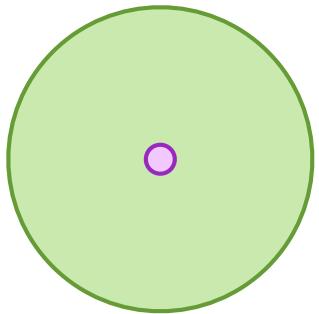


$$d = (r_1 + r_2)$$

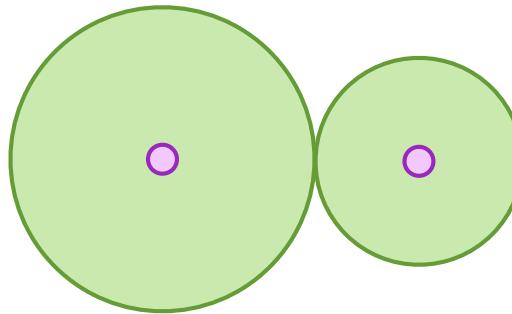


$$d < (r_1 + r_2)$$

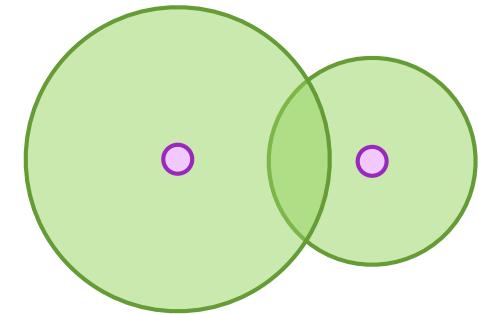
can we measure “cost” of not satisfying constraint?



$$d - (r_1 + r_2) > 0$$

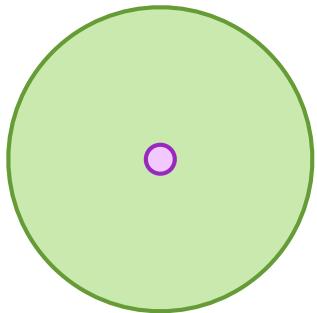


$$d - (r_1 + r_2) = 0$$

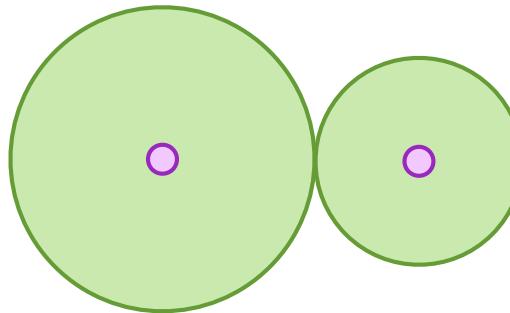


$$d - (r_1 + r_2) < 0$$

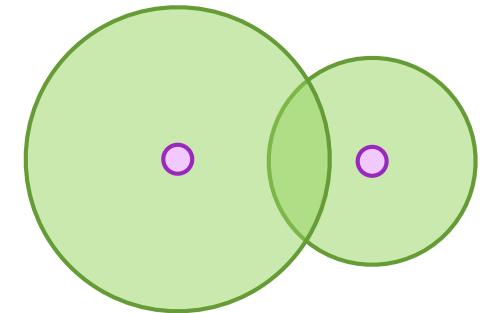
can we measure “cost” of not satisfying constraint?



$$(d - (r_1+r_2))^2 > 0$$



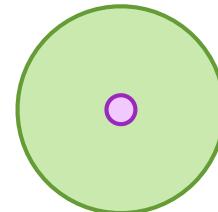
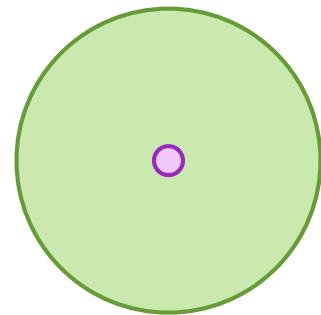
$$(d - (r_1+r_2))^2 = 0$$



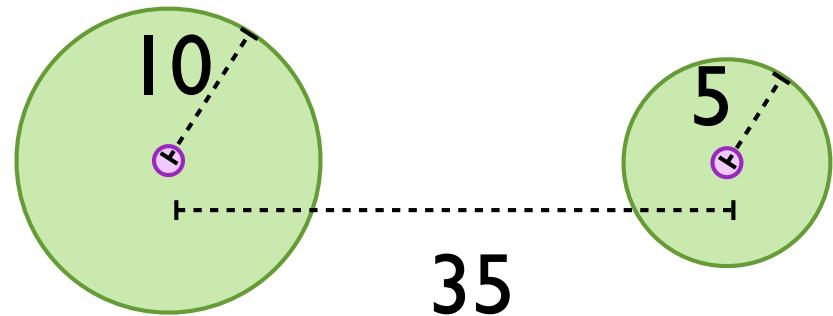
$$(d - (r_1+r_2))^2 > 0$$

idea: measure “how far are we from solution” by taking square of expression?

want to find variable values that **minimise** this “cost”



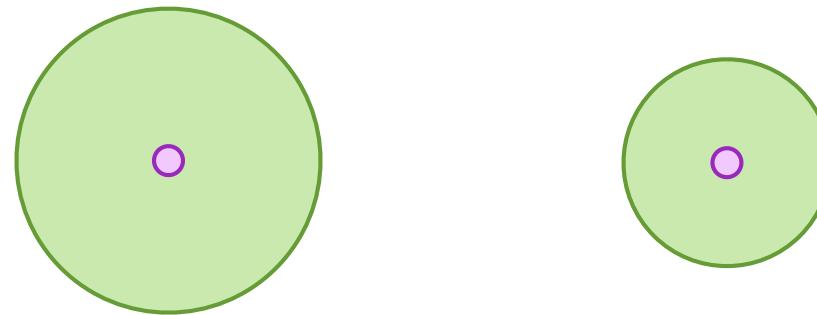
$$(d - (r_1 + r_2))^2$$



$$(d - (r_1 + r_2))^2$$

$$= (35 - (10 + 5))^2$$

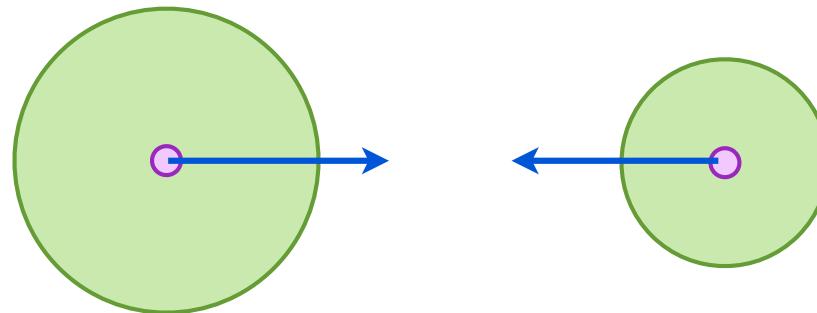
$$= 400$$



$$(d - (r_1 + r_2))^2$$

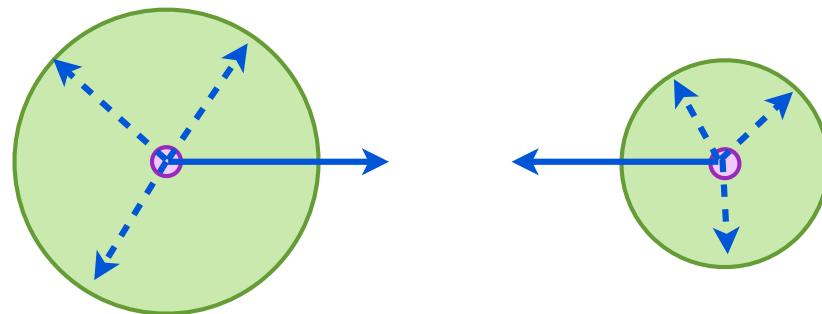
what **direction** should each variable
move to get a better score?

centres should move towards each other...



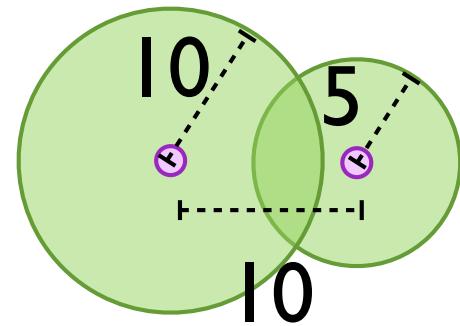
$$(d - (r_1 + r_2))^2$$

radii should increase



$$\frac{(d - (r_1 + r_2))^2}{}$$

the **gradient** of our function can give us
that information

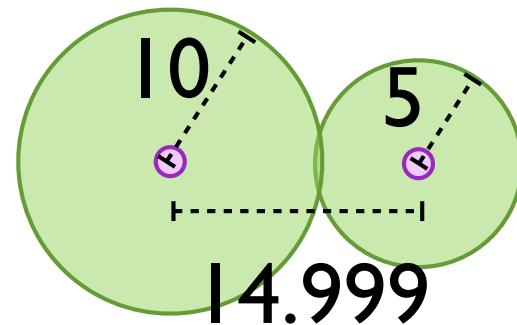


$$(d - (r_1 + r_2))^2$$

$$= (10 - (10+5))^2$$

$$= 25$$

closer!



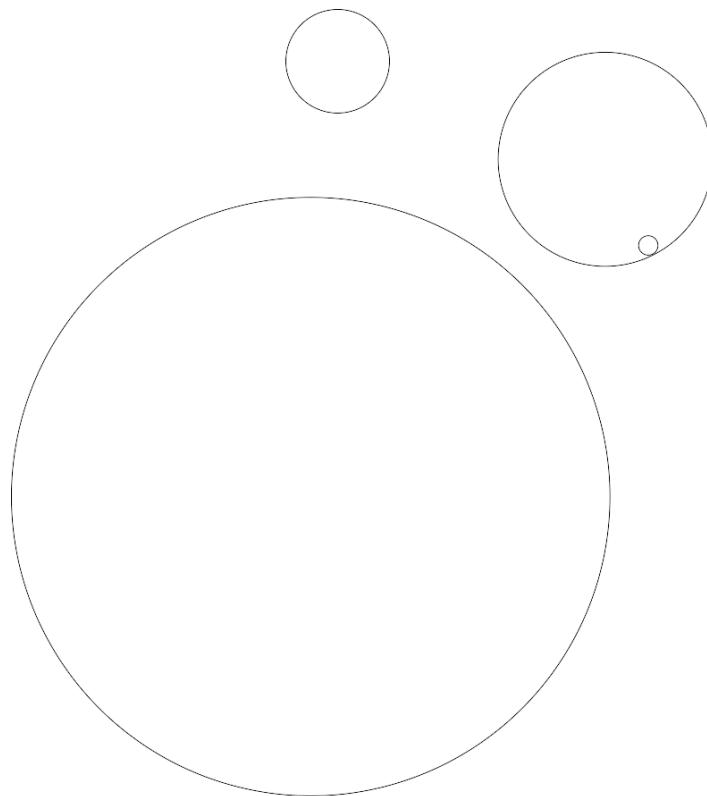
$$(d - (r_1 + r_2))^2$$

$$= (14.999 - (10+5))^2$$

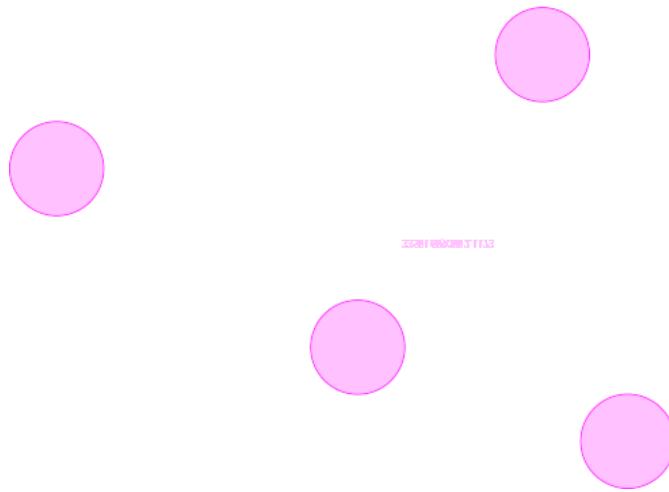
$$= 0.000001$$

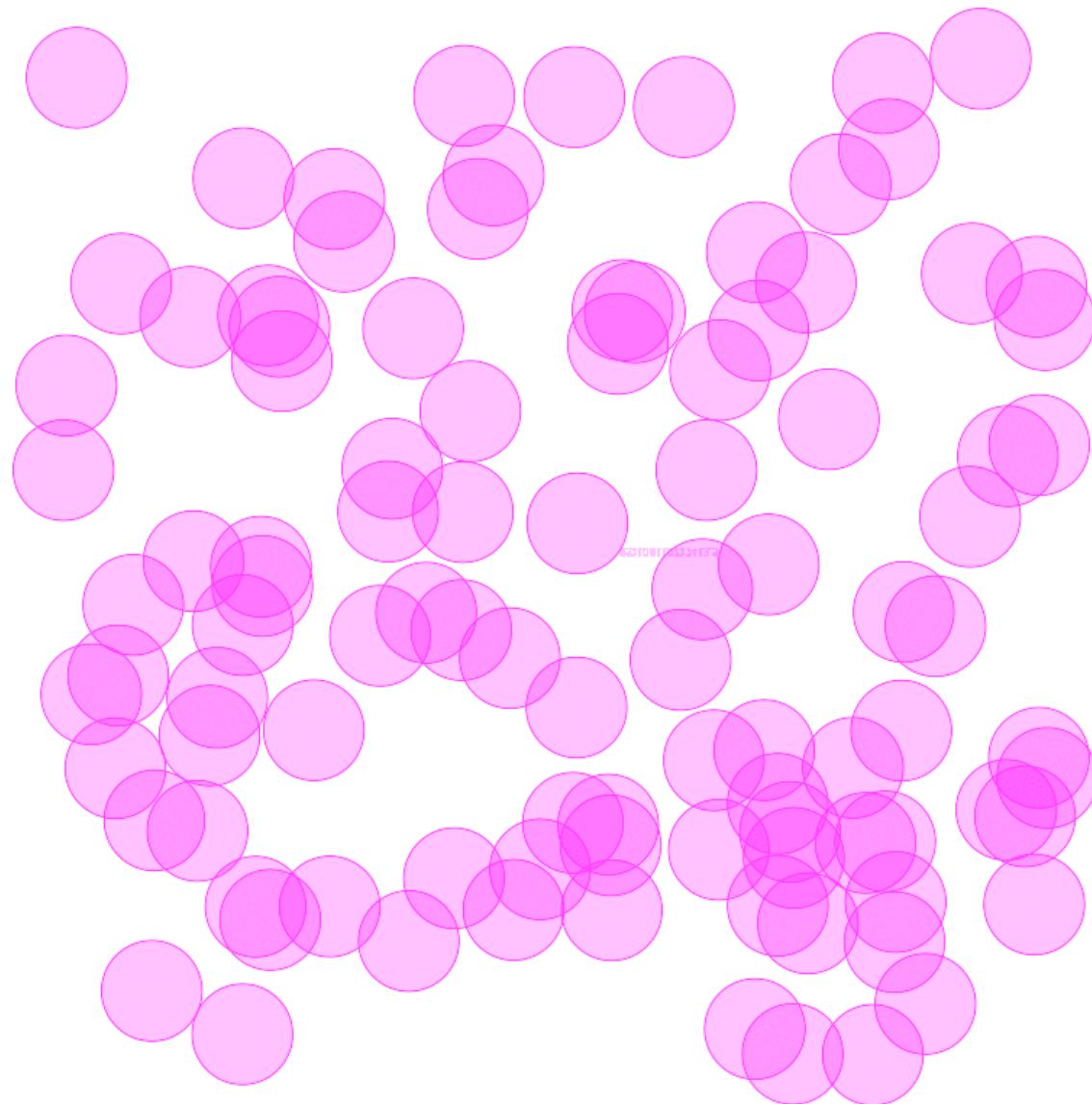
close enough?

**4 circles
touching**



4 circles
touching, same size





100 circles?

SUMMARY Part I. preliminaries

- expression, equation, inequality
- system of equations
- linear vs nonlinear
- 1 dimension (univariate) vs
 n dimensions (multivariate)
- “solving”: root-finding vs optimisation

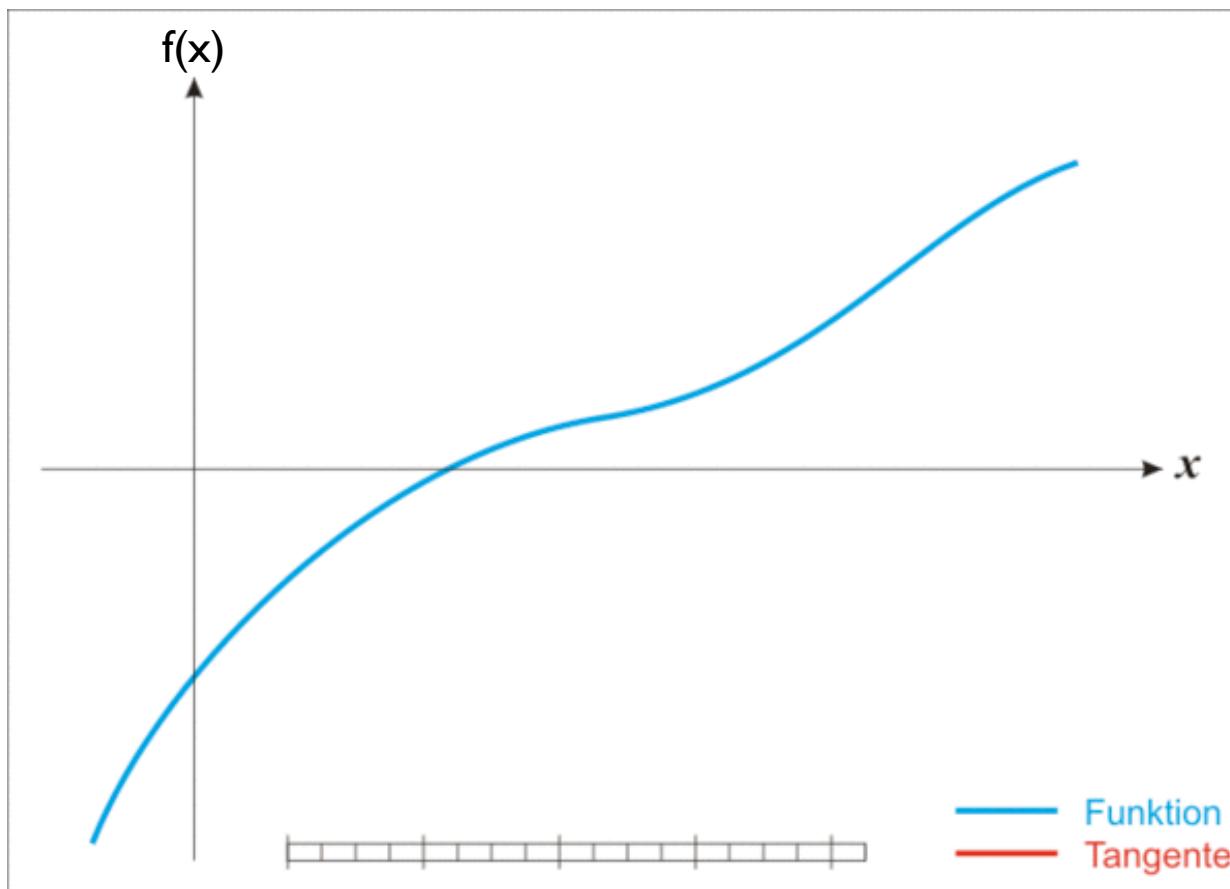
Part 2

iterative methods

guess - check - new guess - check - new guess - check ...

Newton's method for root-finding

- you've seen this for 1 function, for 1 dimension
- first we'll extend this to n dimensions
- then we'll see correspondance to optimisation task



http://en.wikipedia.org/wiki/Newton's_method

Newton's method for root-finding

$$x^{(i+1)} = x^{(i)} + \text{update}$$

Newton's method for root-finding

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

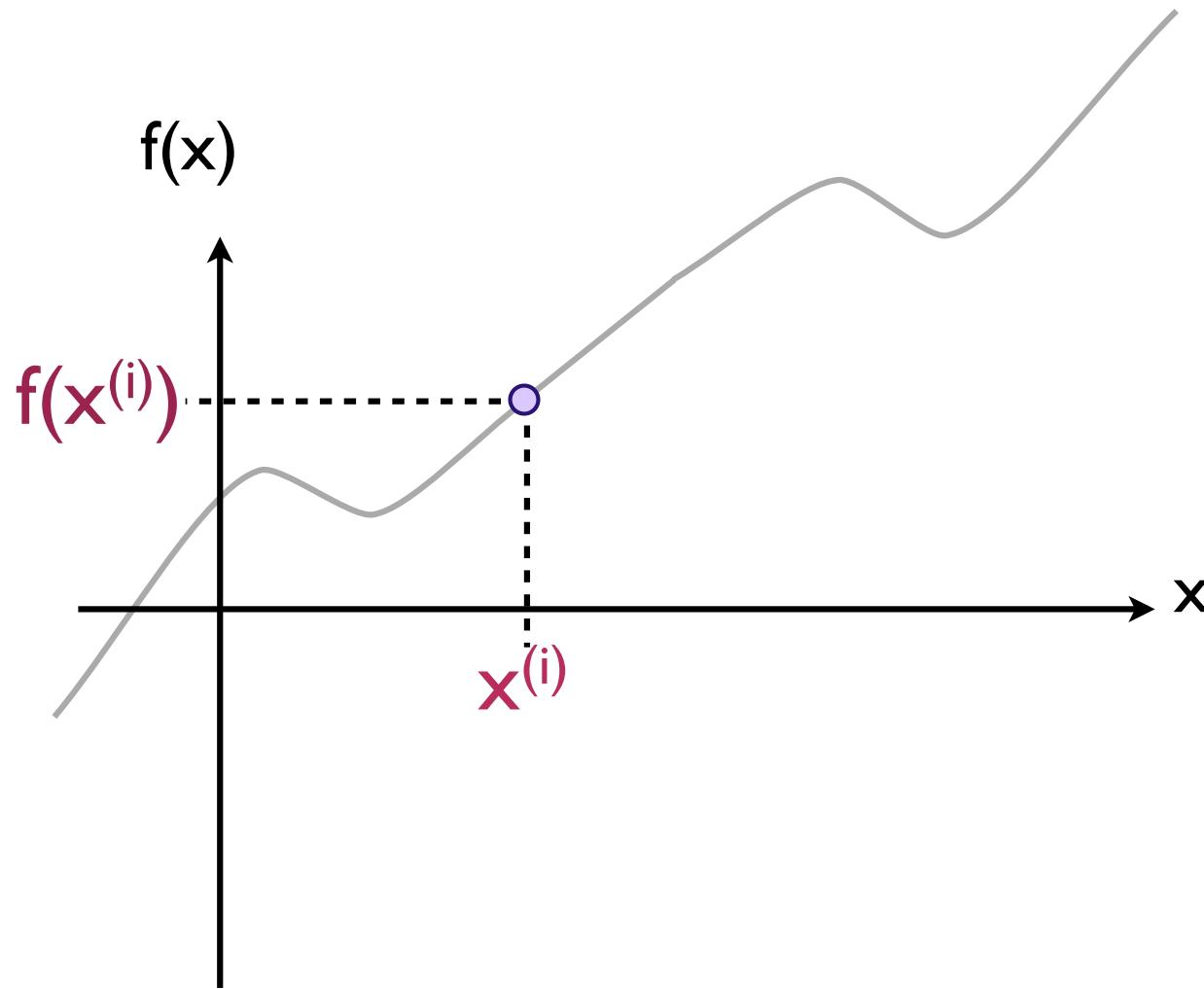
deriving Newton's method

deriving Newton's method

first we'll derive it in the “naive” way

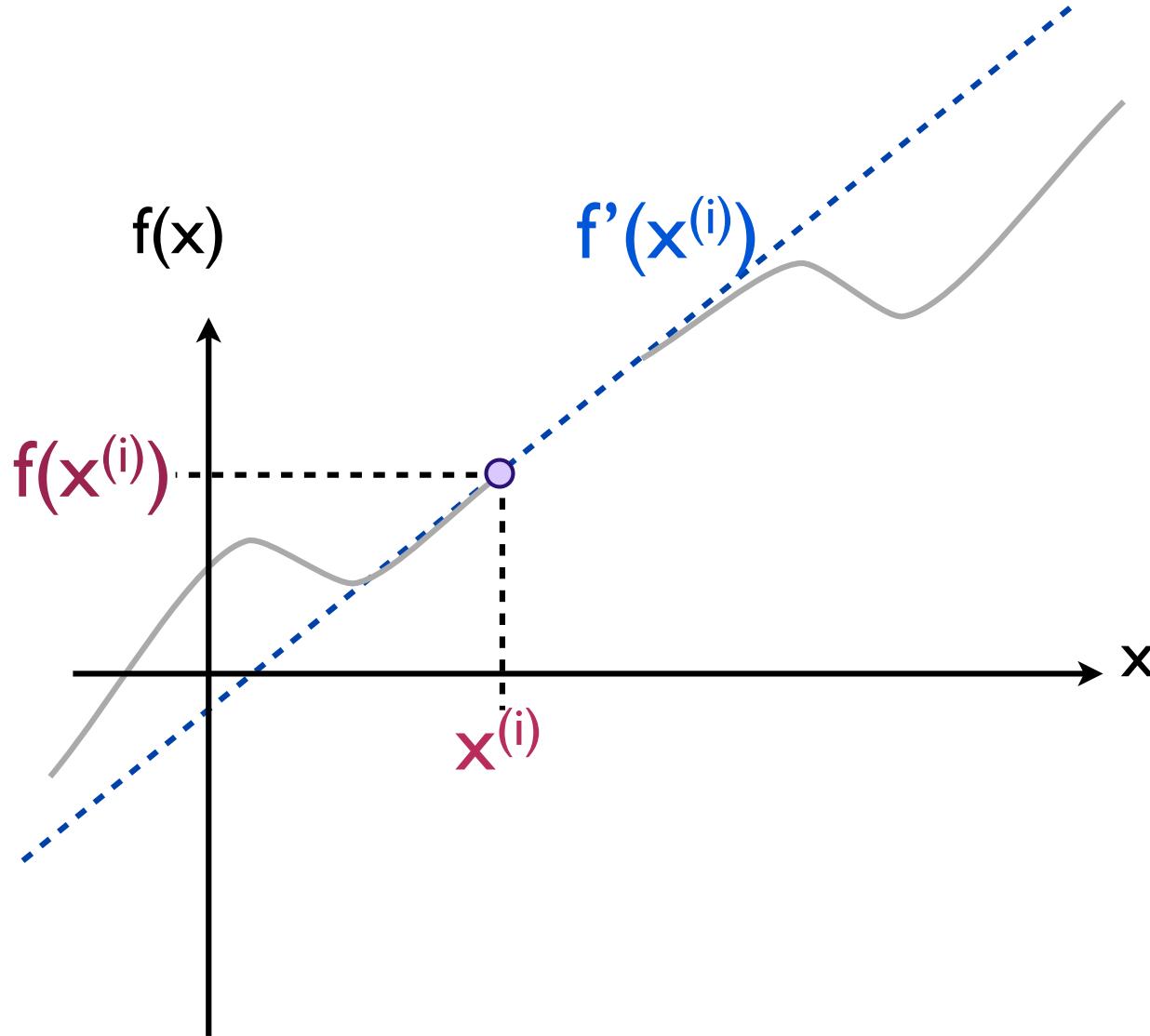
then we'll derive it using Taylor series expansion

say we're at $x^{(i)}$ and we want the next step to find the root, $f(x^{(i+1)})=0$



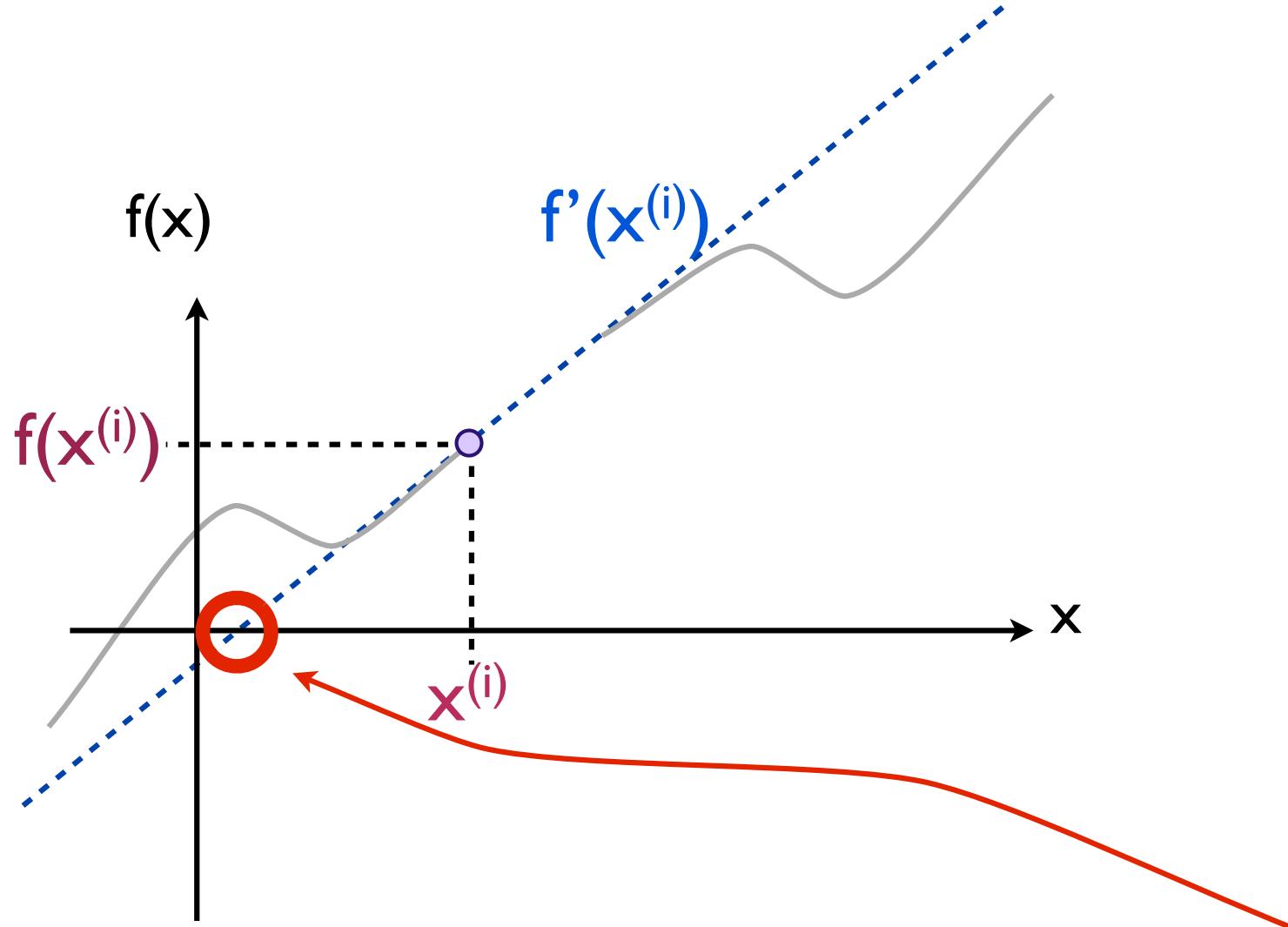
If our function f was linear, then this approximation would be **exact**,

that's the approximation we're making



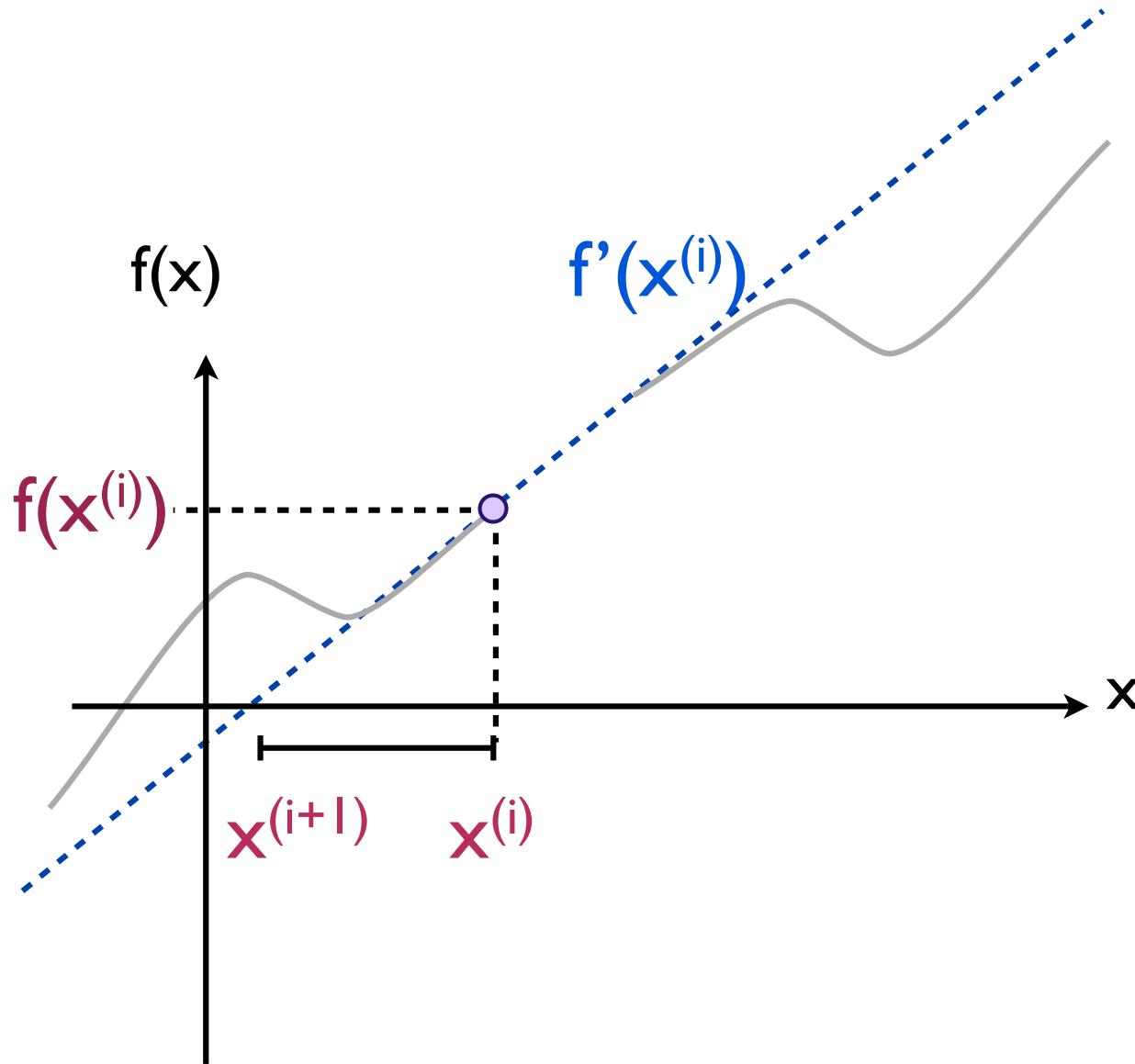
If our function f was linear, then this approximation would be **exact**,

that's the approximation we're making



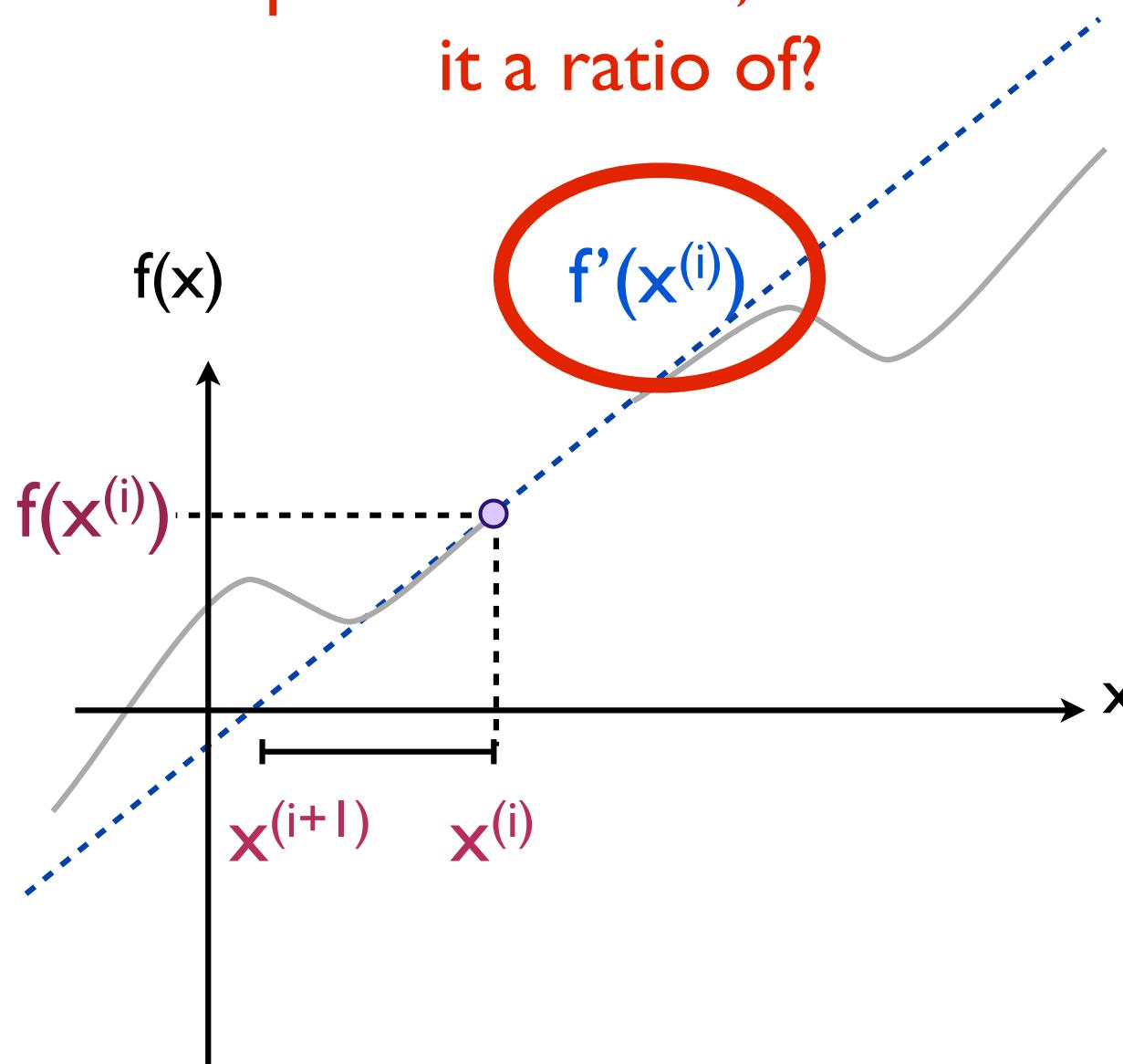
if that was the case, then the root of f would be here

so let's find the change in x



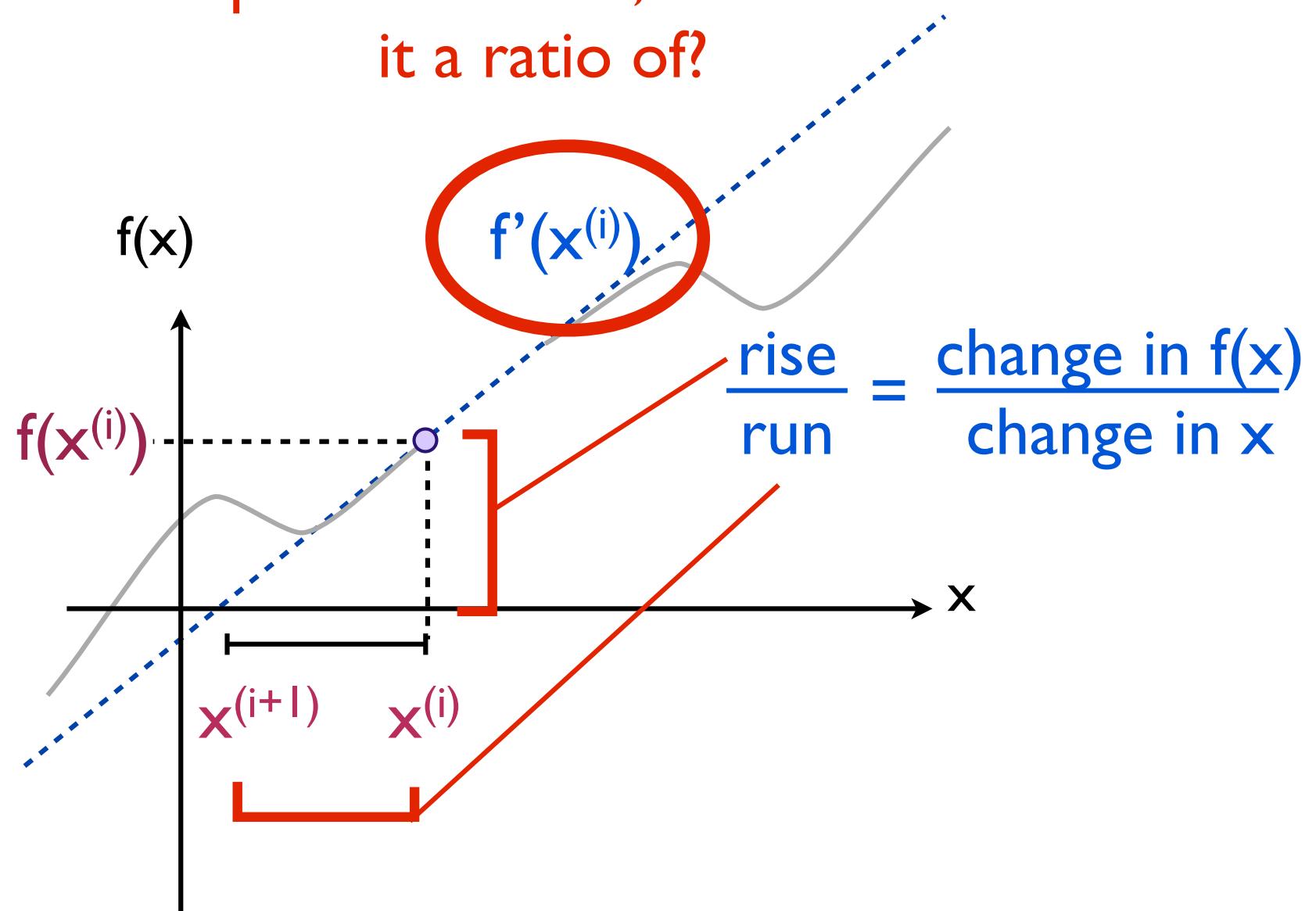
so let's find the change in x

$f'(x^{(i)})$ is the “slope” of this line, i.e. it is a ratio - what is it a ratio of?



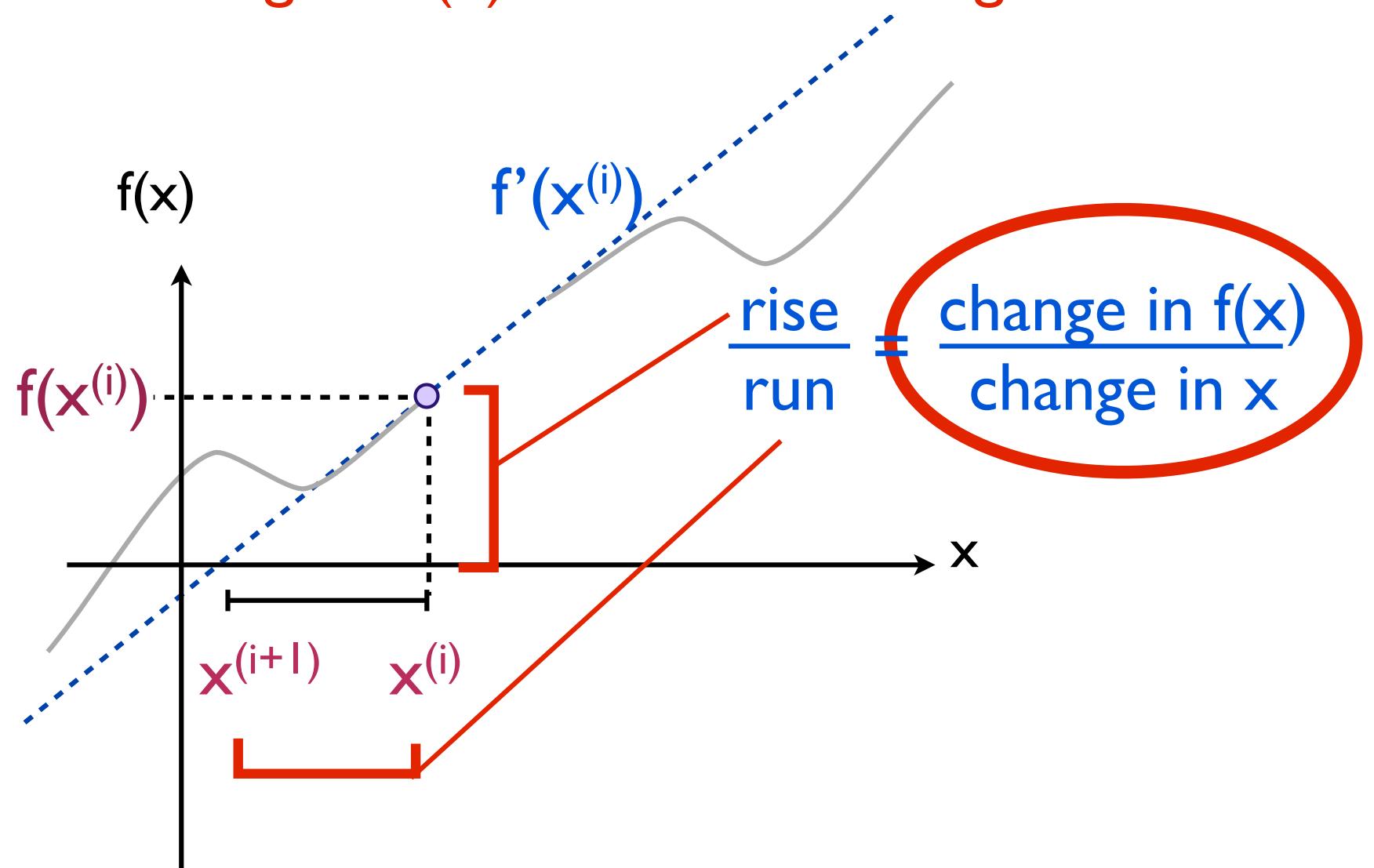
so let's find the change in x

$f'(x^{(i)})$ is the “slope” of this line, i.e. it is a ratio - what is it a ratio of?



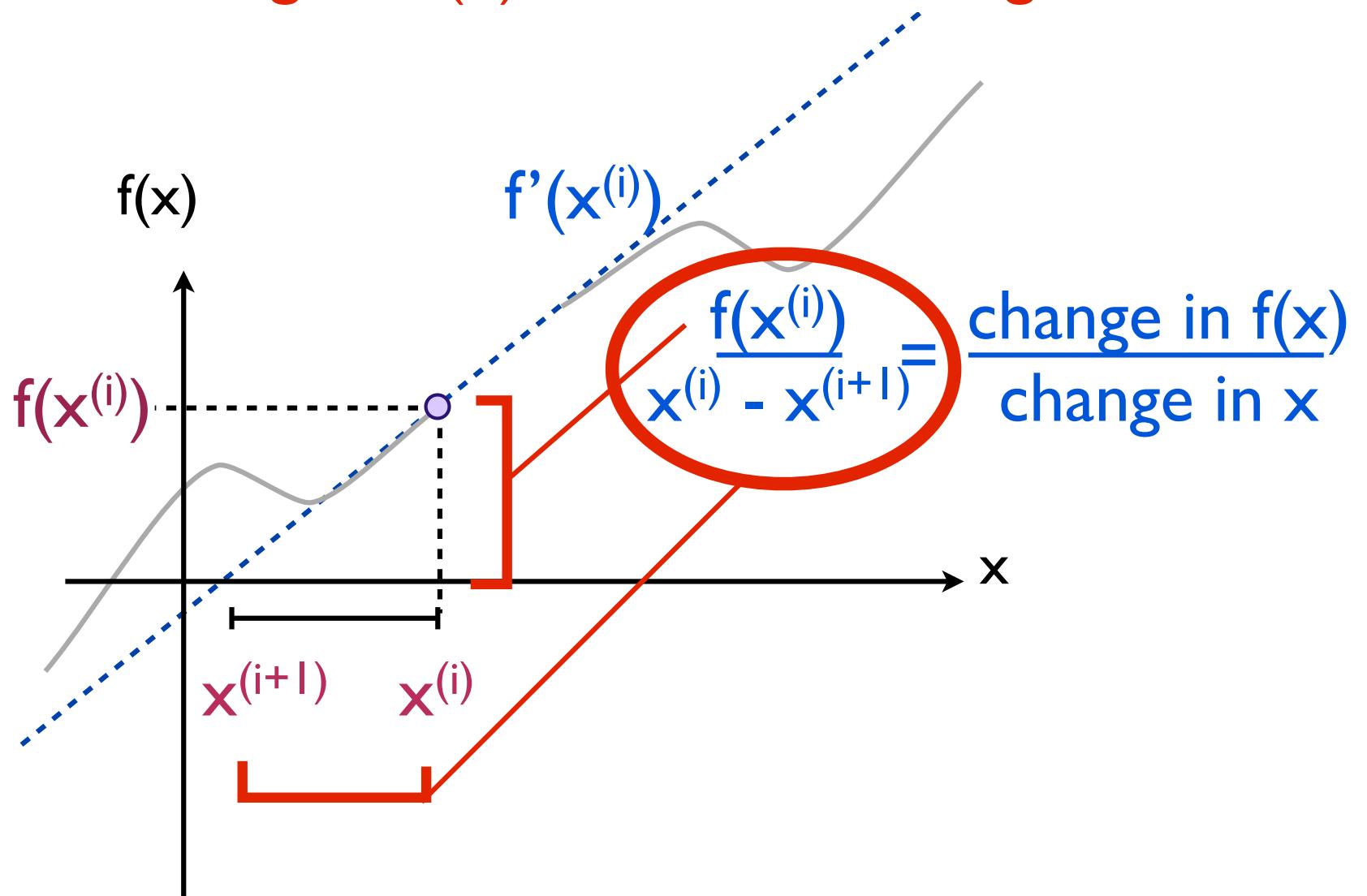
so let's find the change in x

What is “change in $f(x)$ ”? What is “change in x ”?



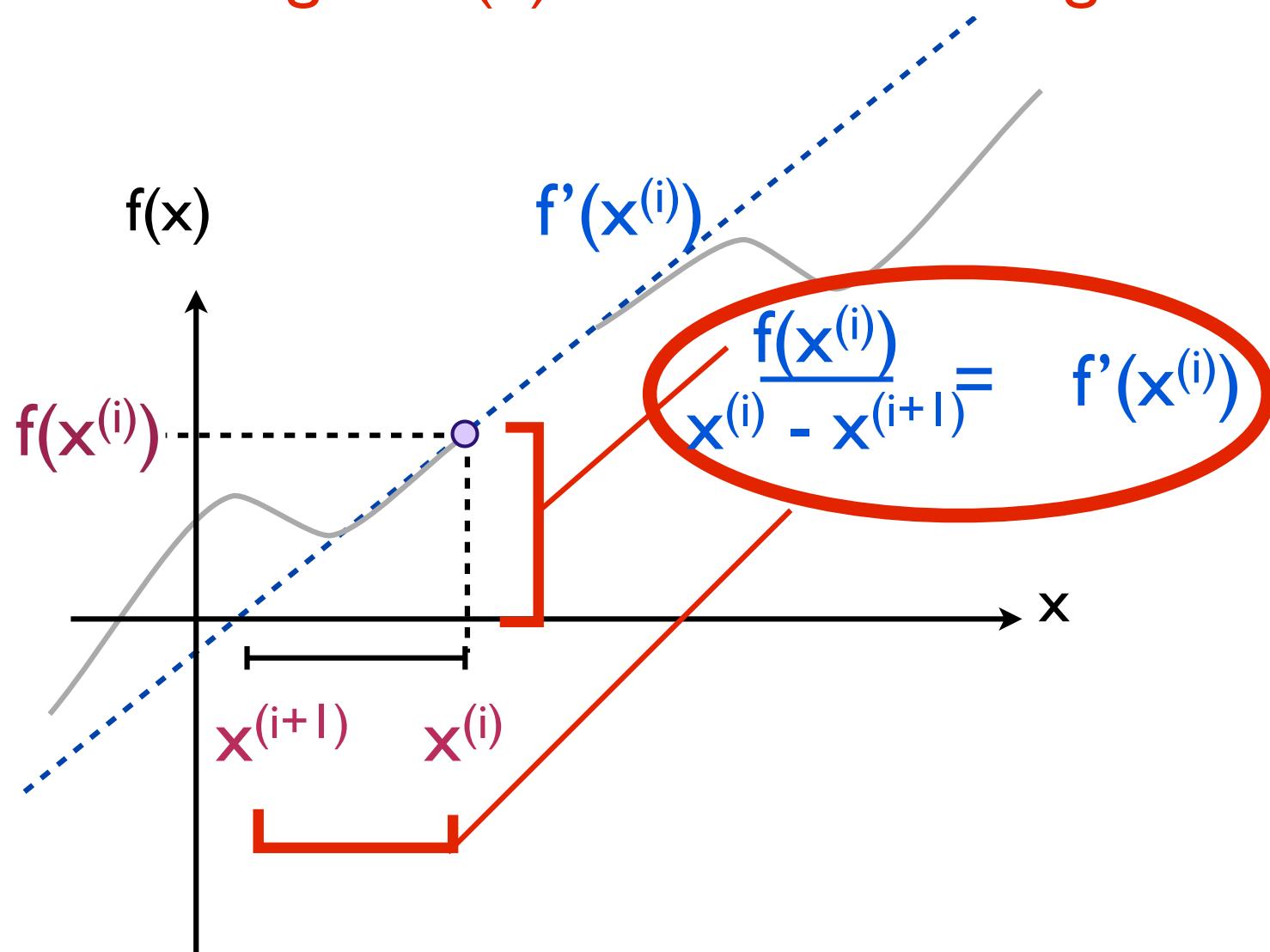
so let's find the change in x

What is “change in $f(x)$ ”? What is “change in x ”?



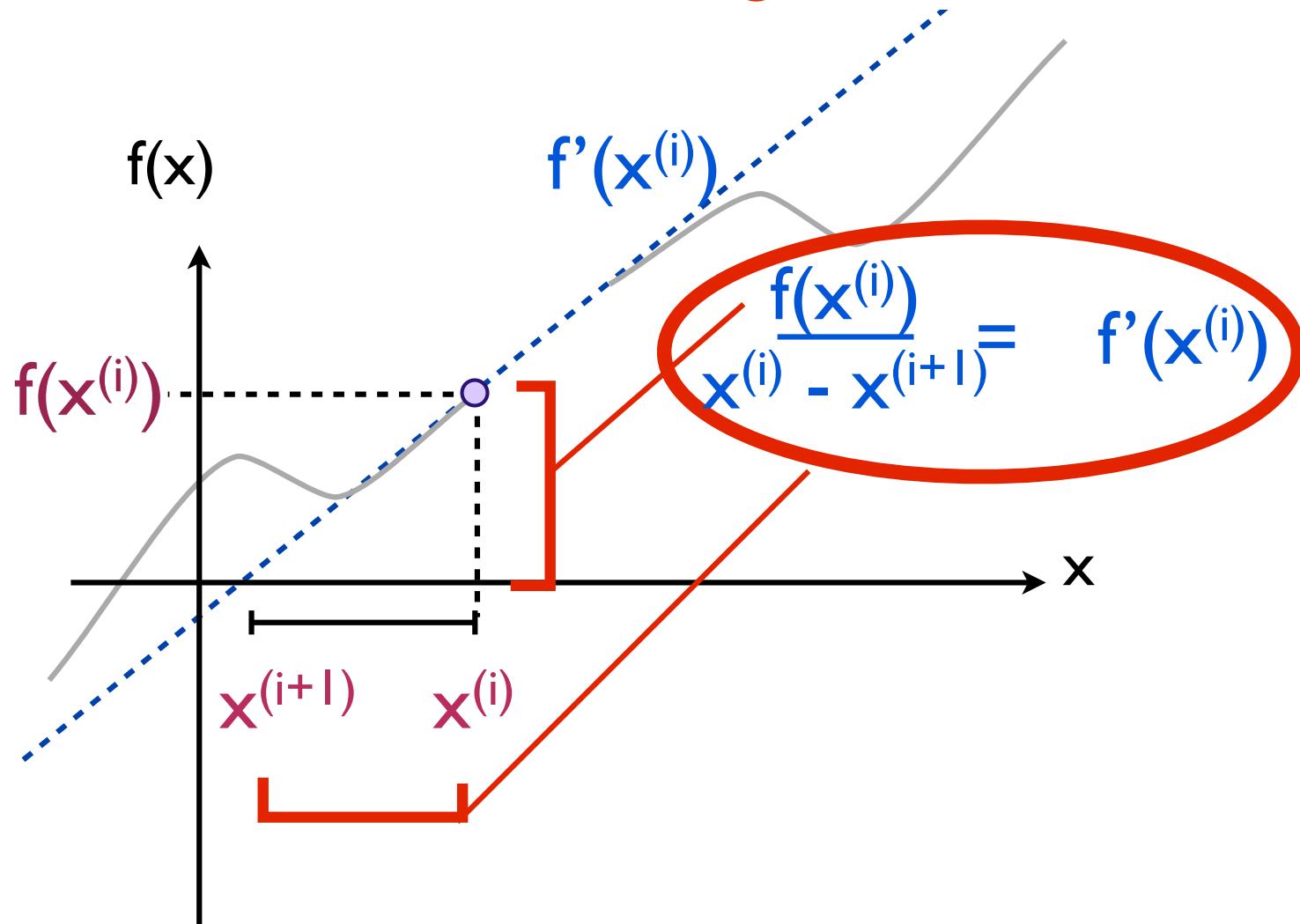
so let's find the change in x

What is “change in $f(x)$ ”? What is “change in x ”?



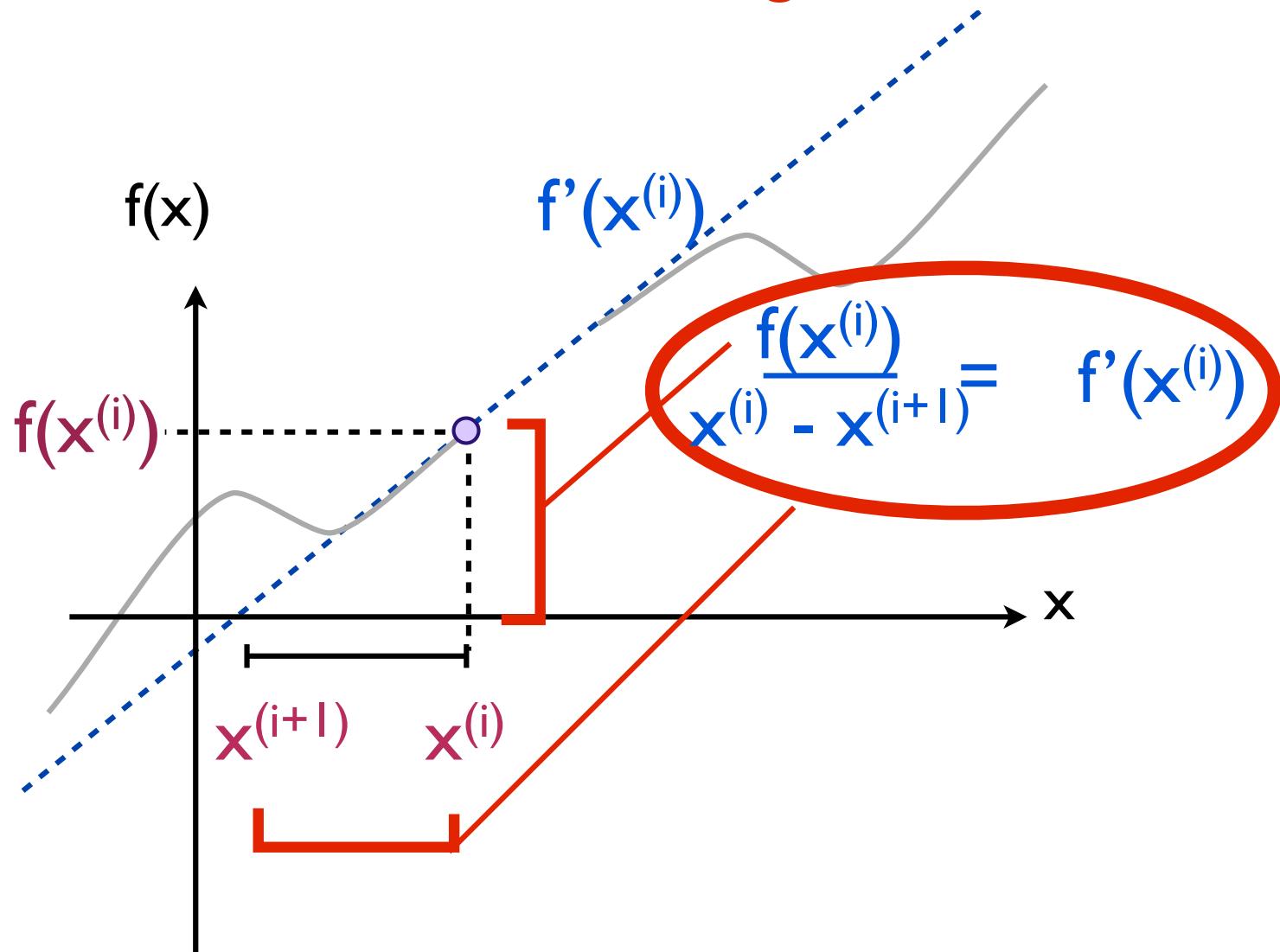
so let's find the change in x

What variable values do we have, and what do we want to solve for again?

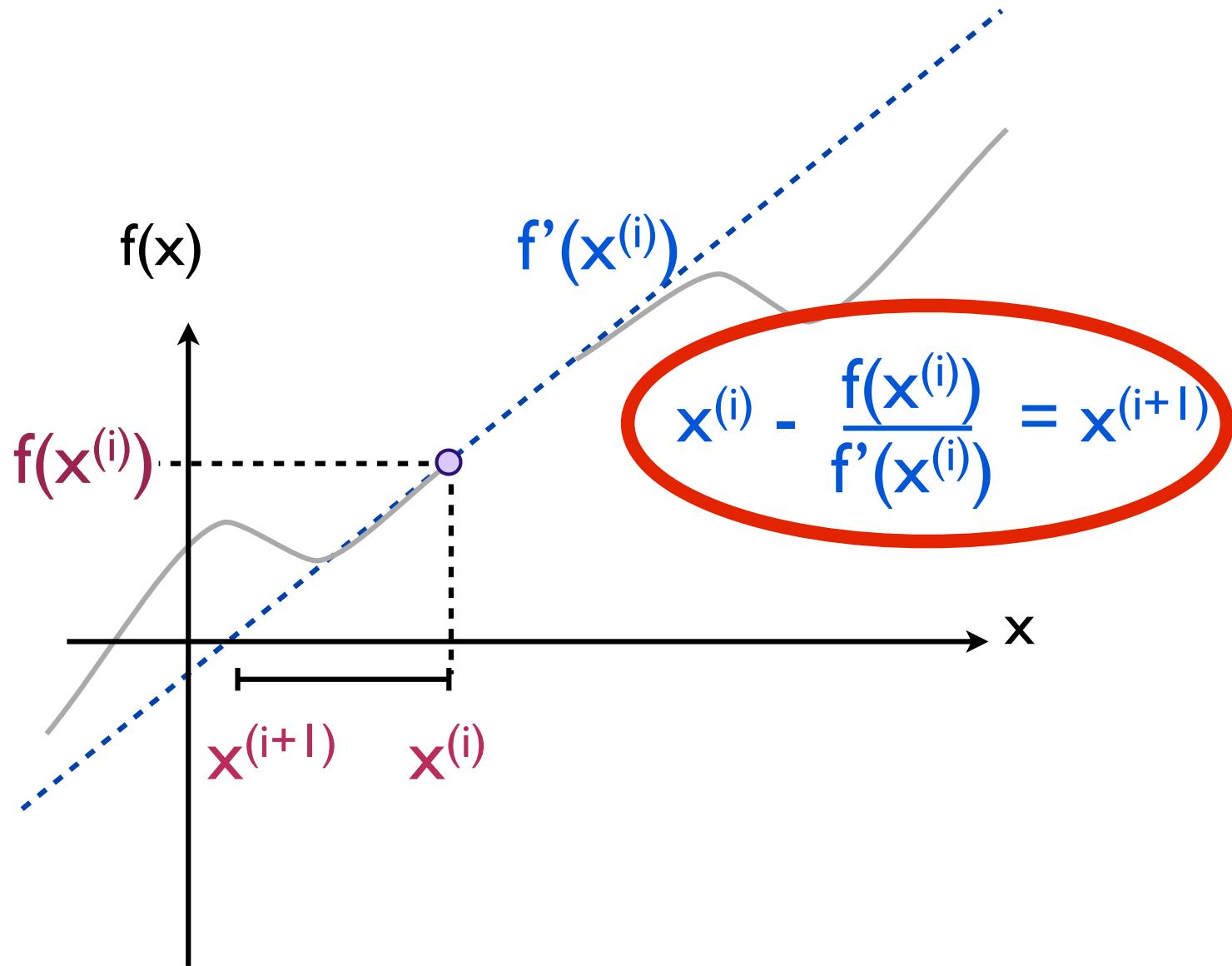


so let's find the change in x

We want $x^{(i+1)}$ - can we rearrange to solve for $x^{(i+1)}$?



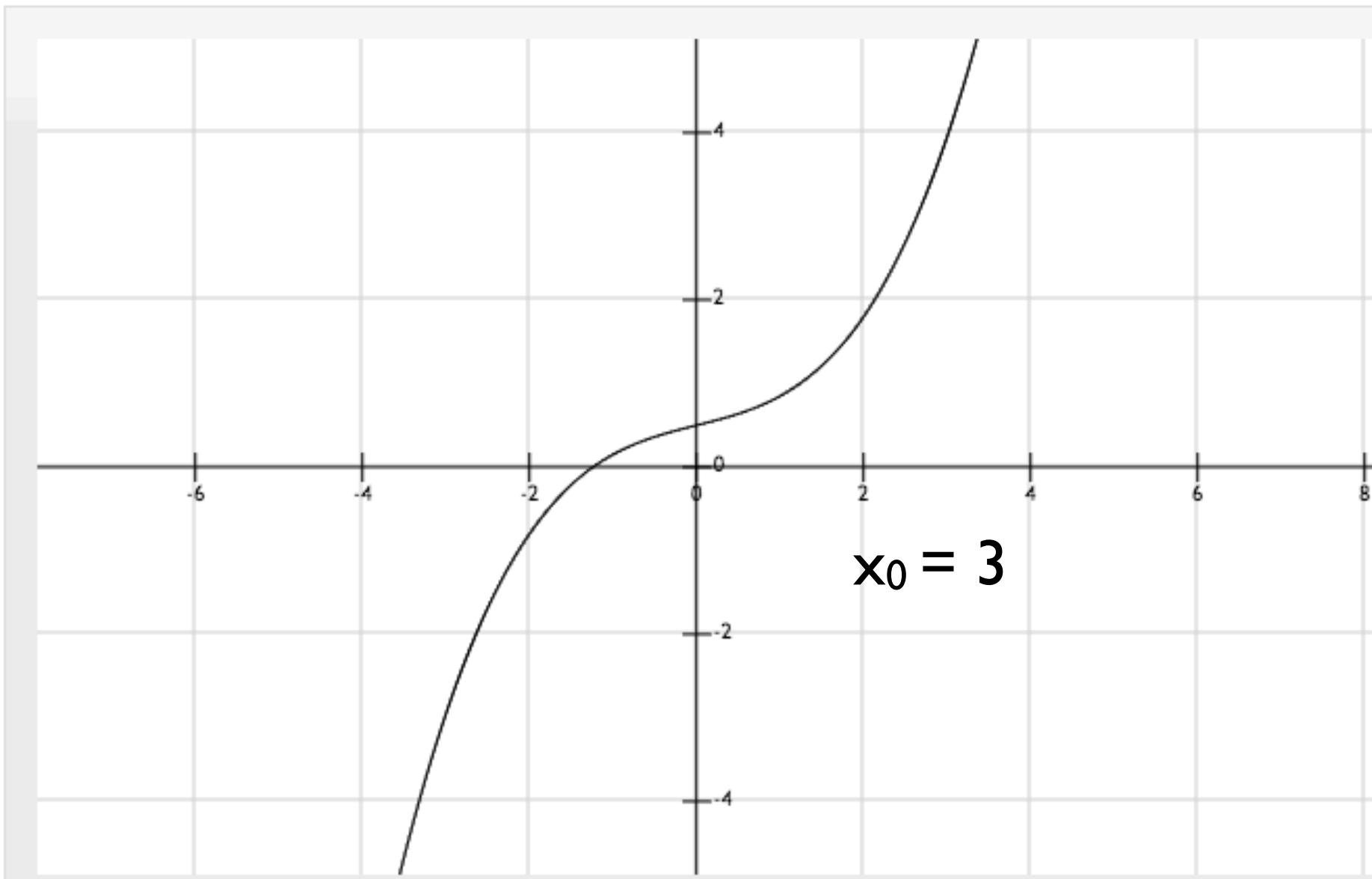
so let's find the change in x



$$f(x) = 0.1x^3 + 0.25x + 0.5$$

$$f'(x) = ?$$

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



$$f(x) = 0.1x^3 + 0.25x + 0.5$$

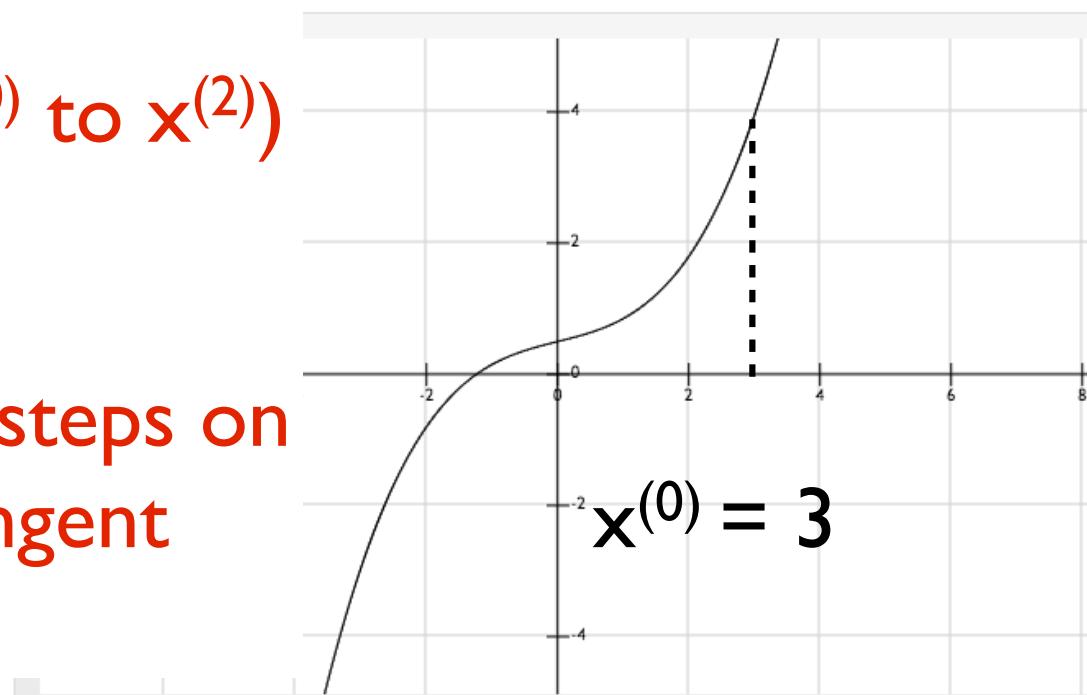
$$f'(x) = 0.3x^2 + 0.25$$

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

Step 1: write pseudo-code
that implements this iterative
“loop” for 10 iterations

Step 2: do 2 steps ($x^{(0)}$ to $x^{(2)}$)
by hand

Step 3: draw the two steps on
this graph (including tangent
lines)



$$f(x) = 0.1x^3 + 0.25x + 0.5$$

$$f'(x) = 0.3x^2 + 0.25$$

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

double xi = 3

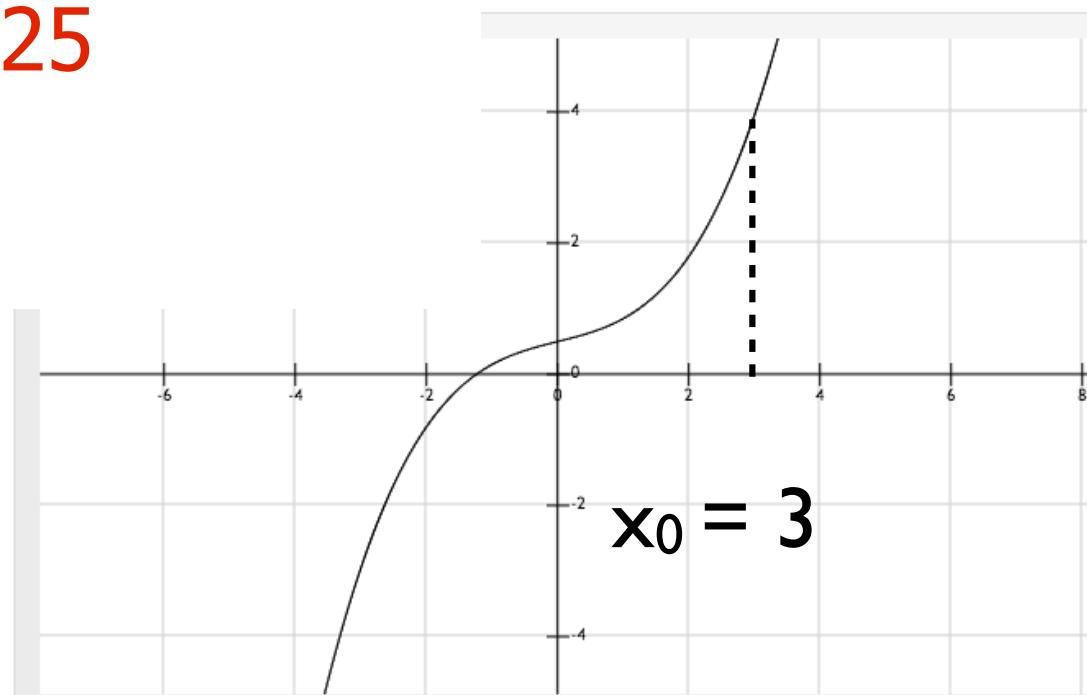
for i = 1 to 10 {

 fx = 0.1*xi^3 + 0.25*xi + 0.5

 f'x = 0.3*xi^2 + 0.25

 xi = xi - (fx / f'x)

}

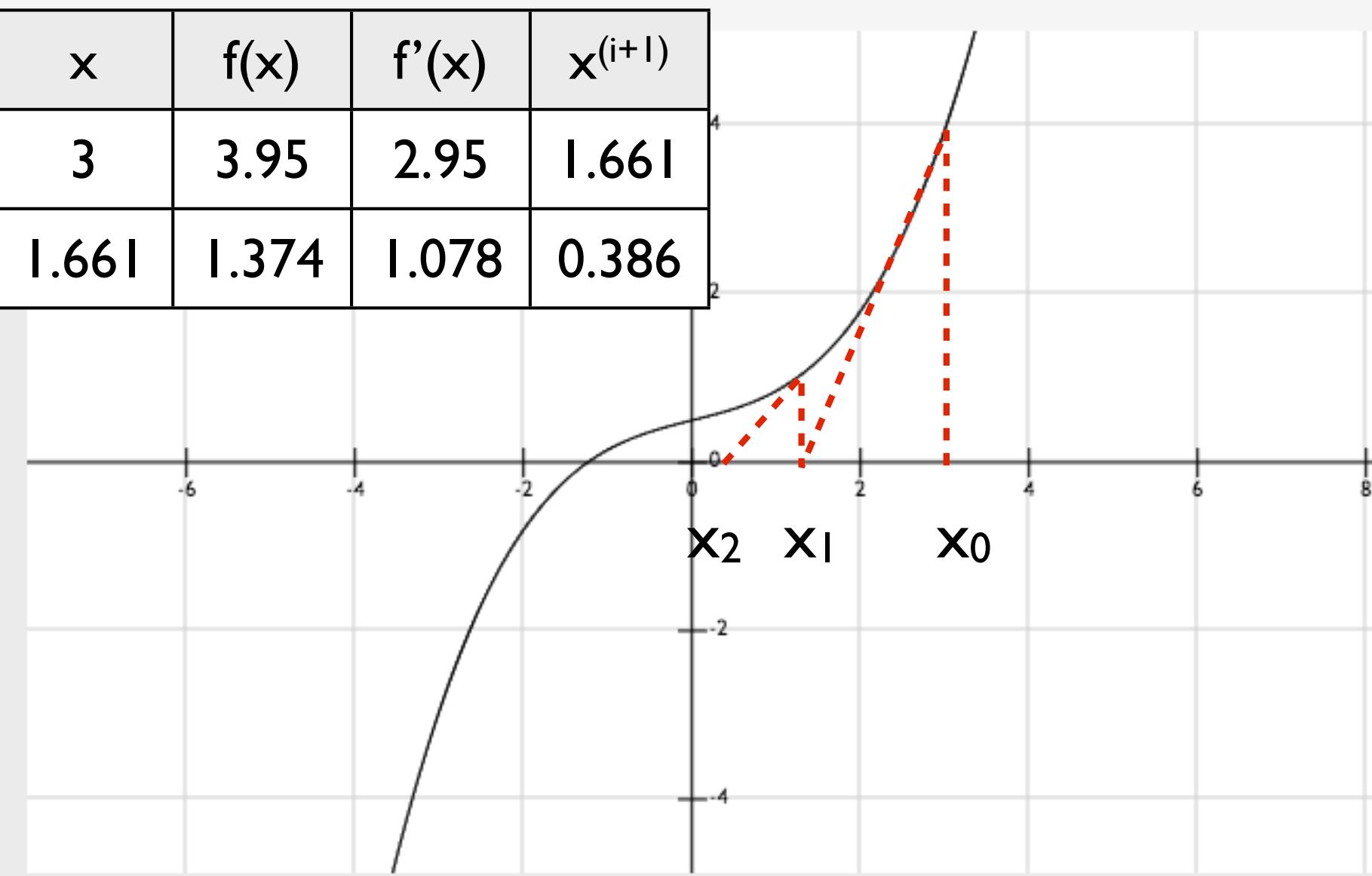


$$f(x) = 0.1x^3 + 0.25x + 0.5$$

$$f'(x) = 0.3x^2 + 0.25$$

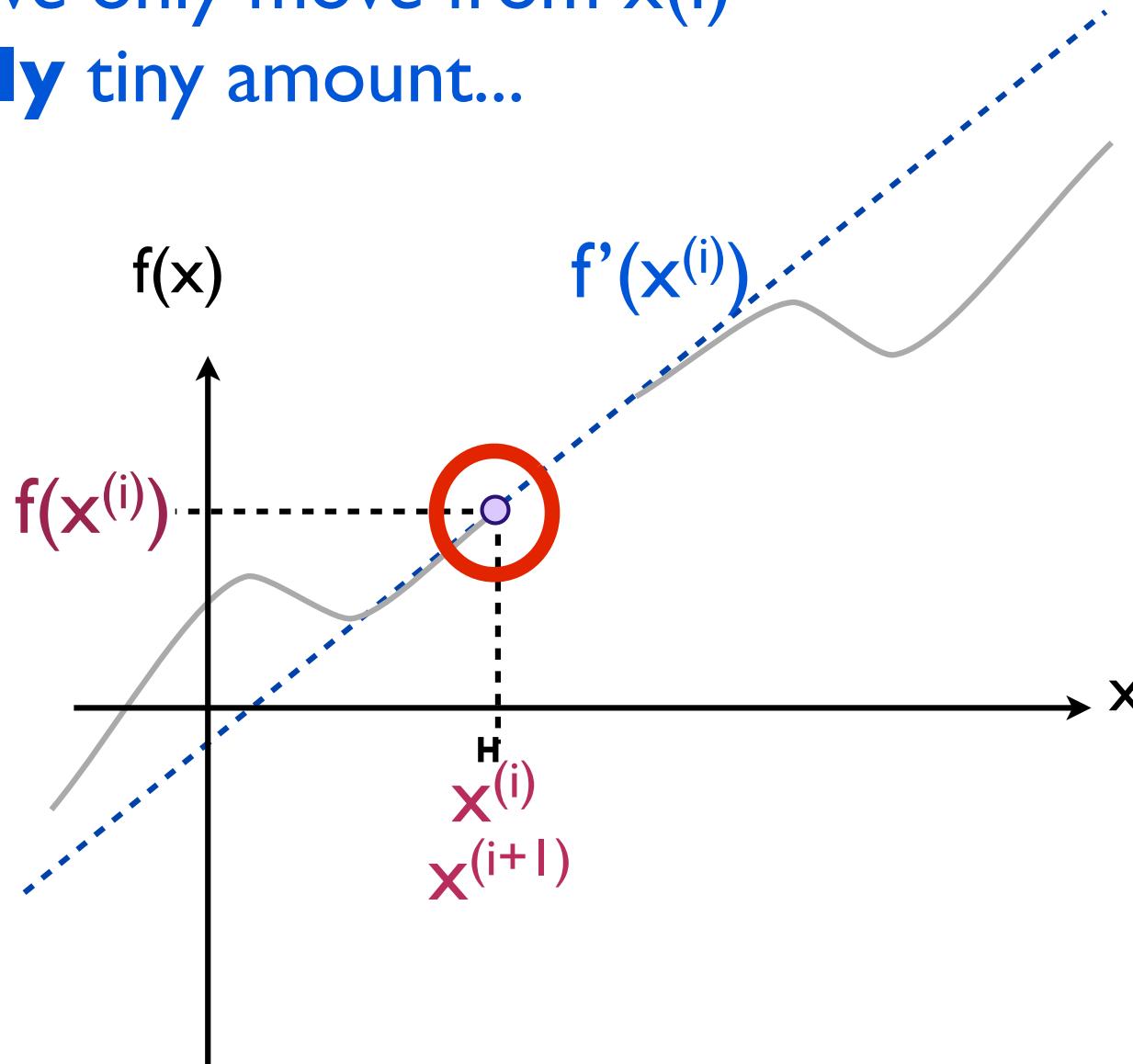
$$x^{(i+1)} = x^{(i)} + \frac{-f(x^{(i)})}{f'(x^{(i)})}$$

x	f(x)	f'(x)	$x^{(i+1)}$
3	3.95	2.95	1.661
1.661	1.374	1.078	0.386

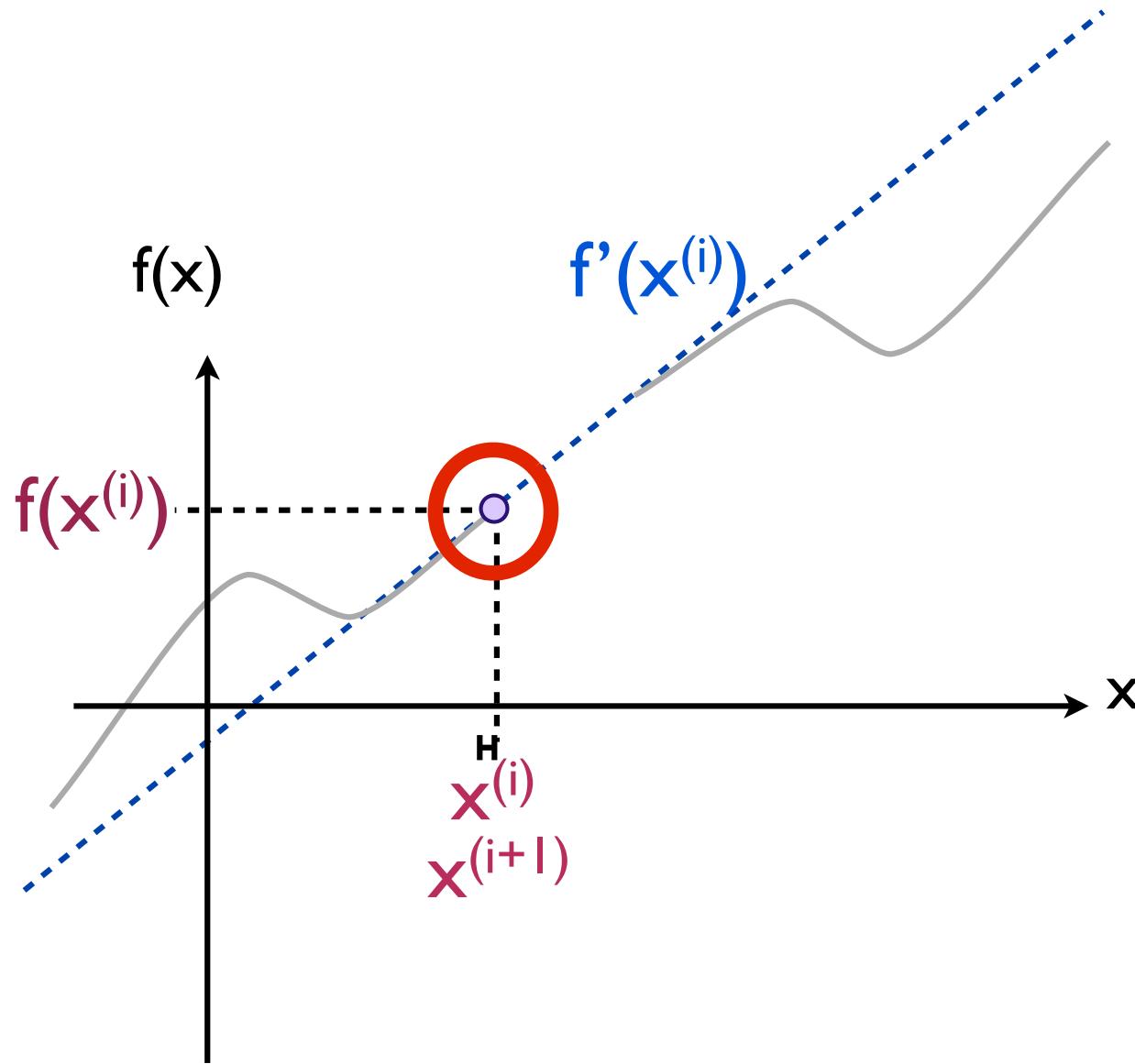


another way to think about it...

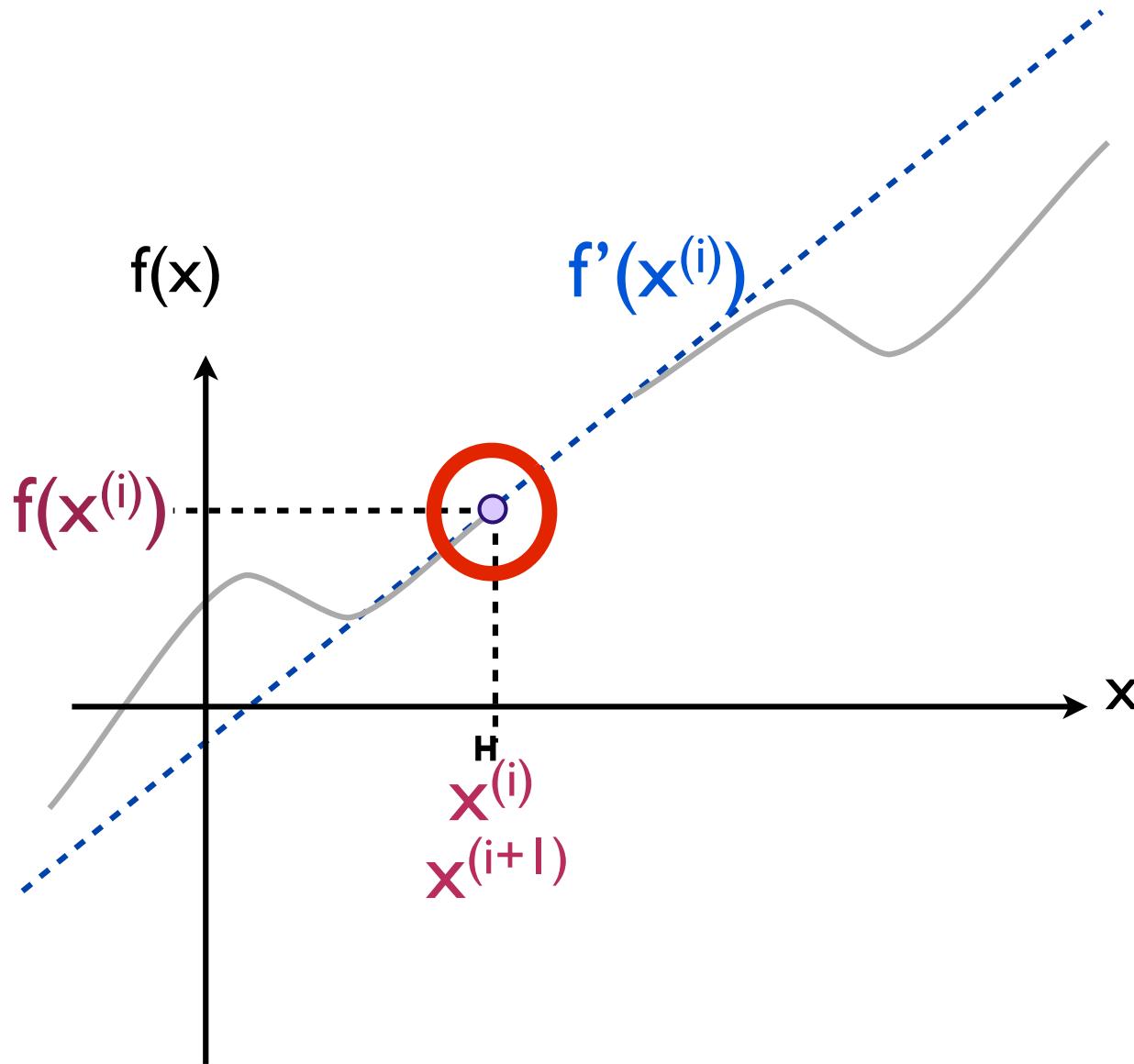
suppose we only move from $x^{(i)}$
by a **really** tiny amount...



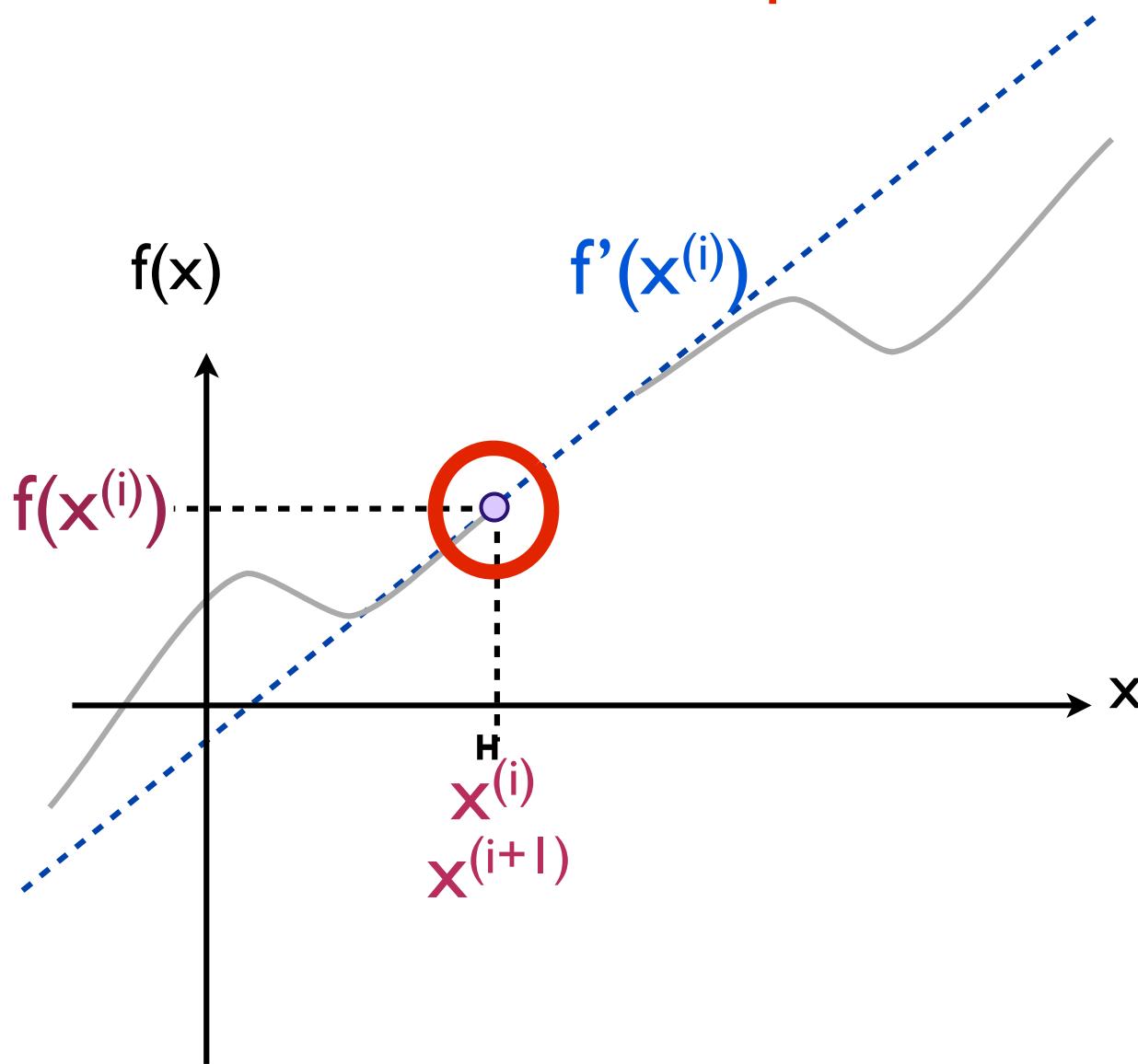
$(x^{(i+1)} - x^{(i)})$ is **almost** zero



$f'(x^{(i)})(x^{(i+1)} - x^{(i)})$ is also **almost** zero

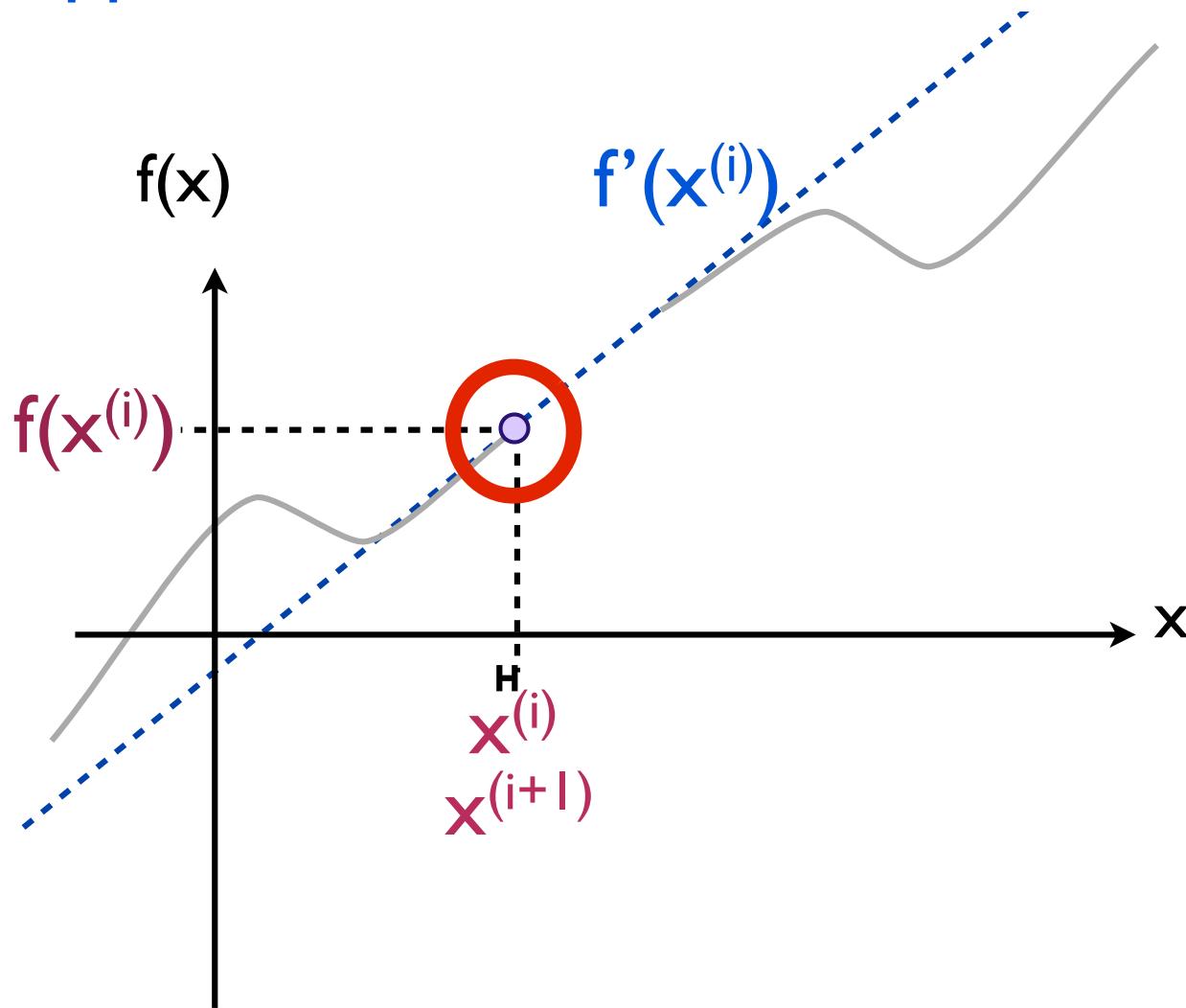


$\frac{f'(x^{(i)})(x^{(i+1)} - x^{(i)}) + f(x^{(i)})}{}$
is also **almost** equal to what?



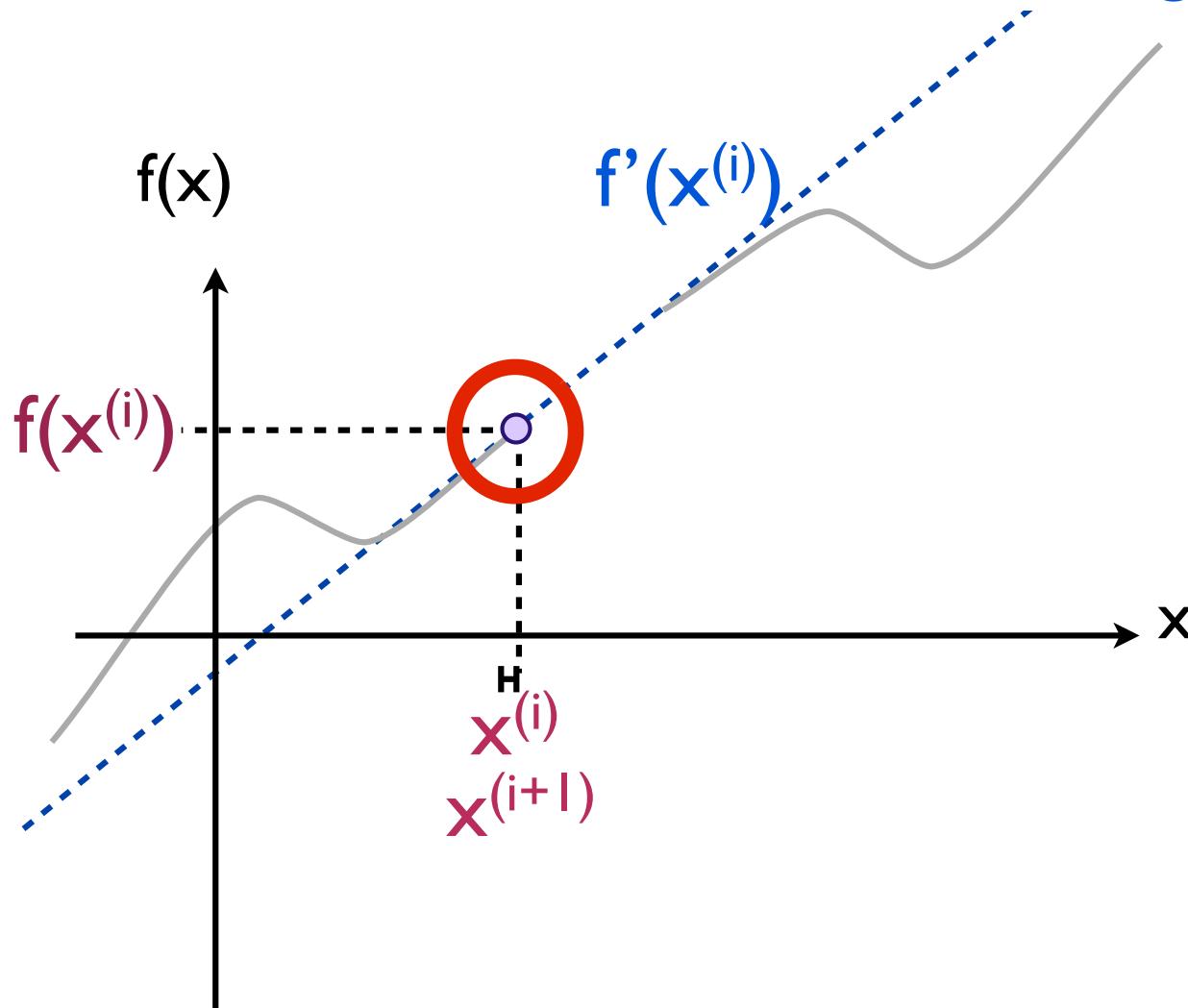
$$\frac{f'(x^{(i)})(x^{(i+1)} - x^{(i)}) + f(x^{(i)})}{}$$

so, for **tiny** change in x , this is a decent approximation of our function at $f(x^{(i+1)})$



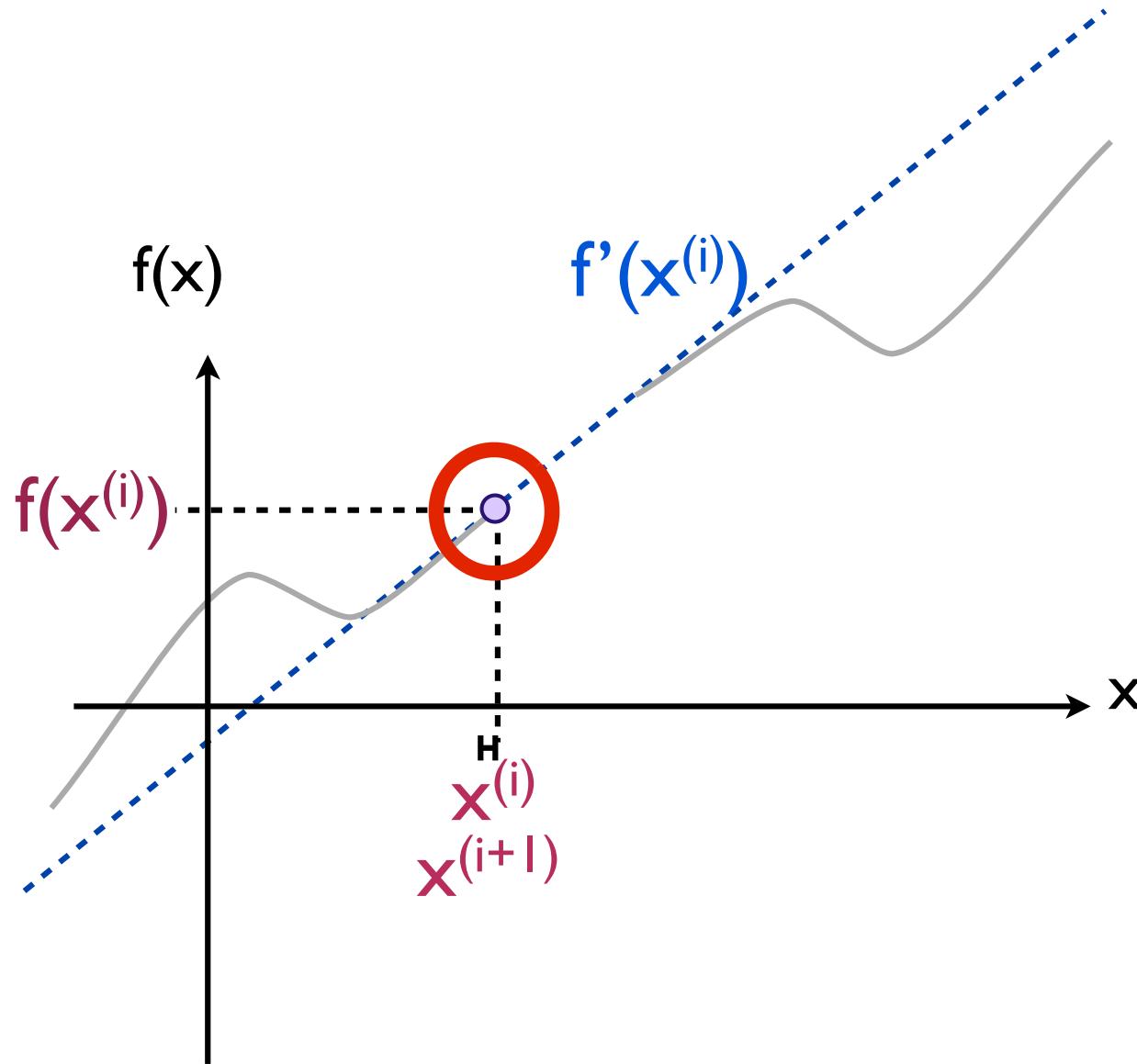
$$\frac{f'(x^{(i)})(x^{(i+1)} - x^{(i)})}{f(x)}$$

...and if our function was linear, it would perfectly describe our function for all change in x



suppose our function was linear: $y = mx + c$

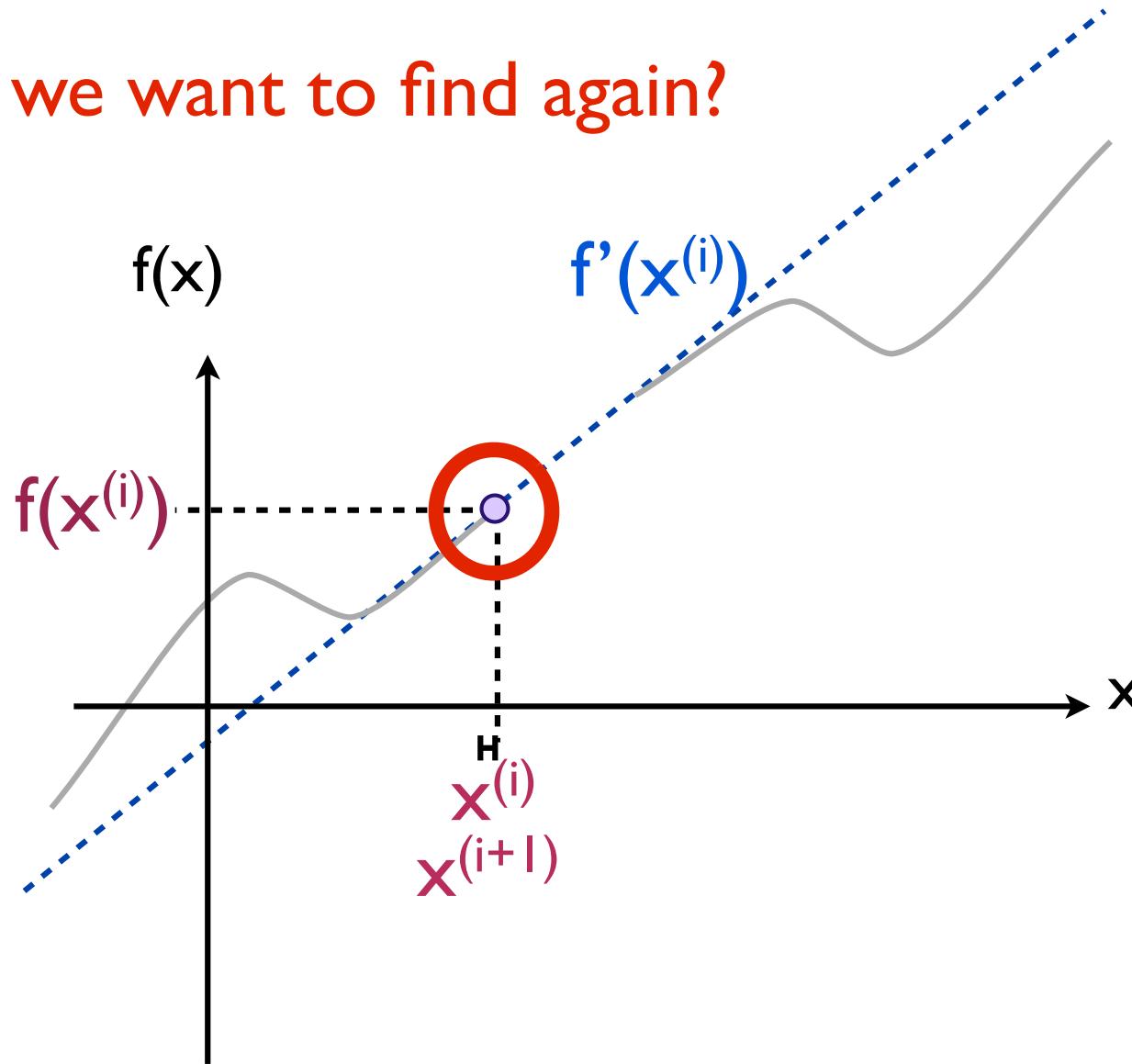
$$f(x^{(i+1)}) = f'(x^{(i)})(x^{(i+1)} - x^{(i)}) + f(x^{(i)})$$



suppose our function was linear: $y = mx + c$

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

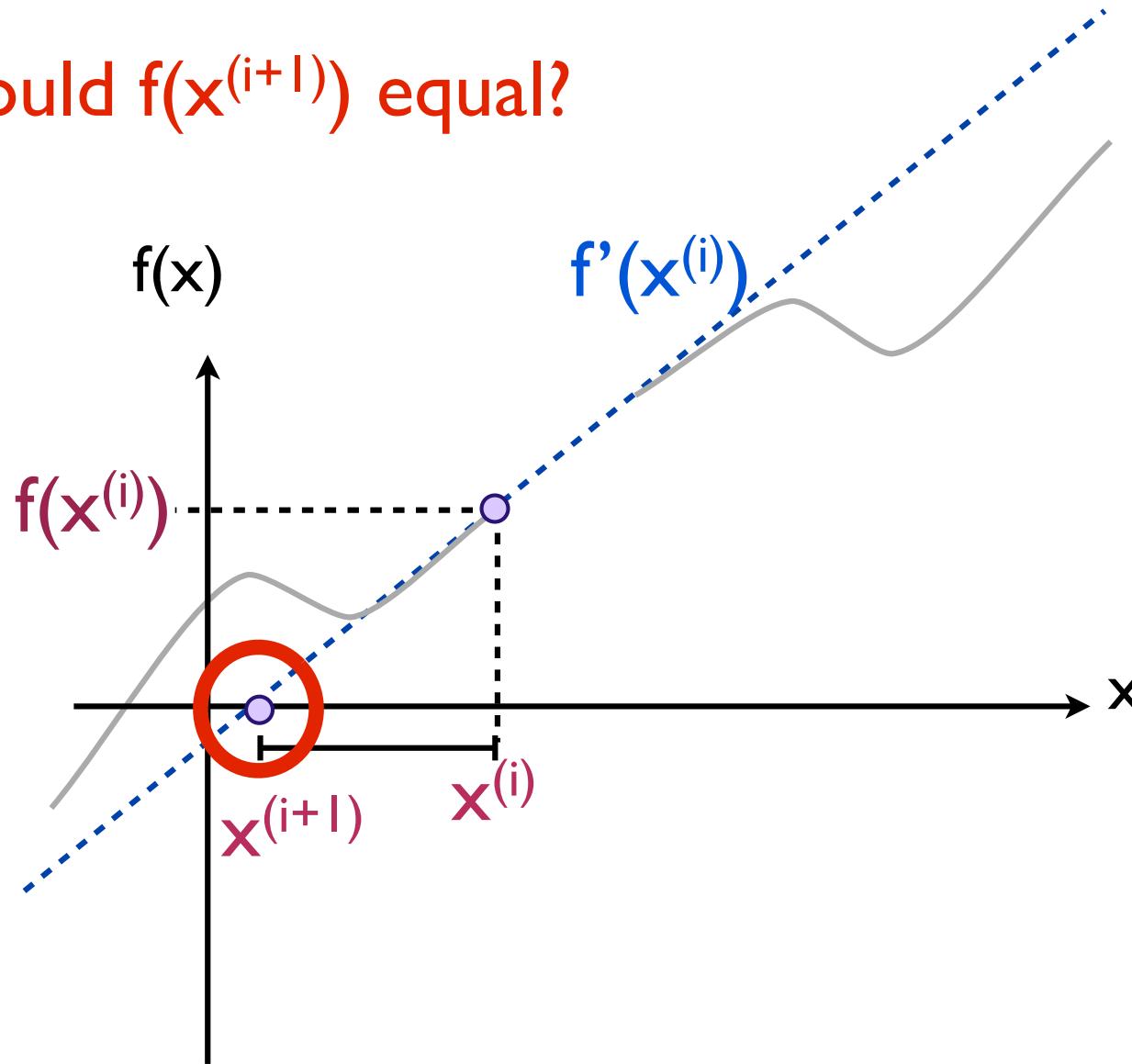
what do we want to find again?



suppose our function was linear: $y = mx + c$

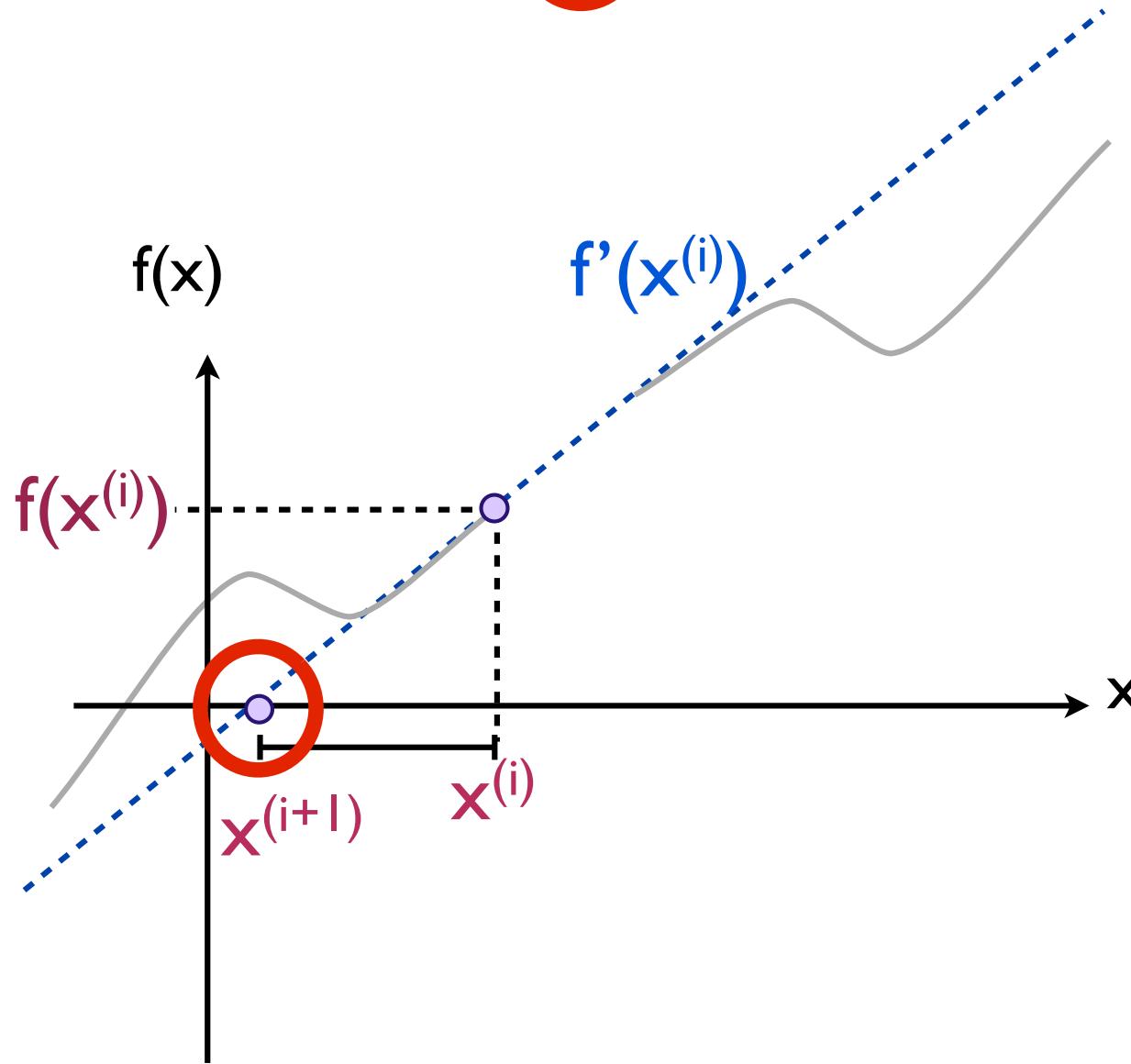
$$f(x^{(i+1)}) = f'(x^{(i)})(x^{(i+1)} - x^{(i)}) + f(x^{(i)})$$

what should $f(x^{(i+1)})$ equal?



suppose our function was linear: $y = mx + c$

$$0 = f'(x^{(i)})(x^{(i+1)} - x^{(i)}) + f(x^{(i)})$$



First-order Taylor series expansion around $x^{(i)}$:

$$f_{\text{linear}}(x) \equiv f'(x^{(i)}) (x - x^{(i)}) + f(x^{(i)})$$

Find root of this linearised system:

$$f_{\text{linear}}(x^{(i+1)}) \equiv f'(x^{(i)}) (x^{(i+1)} - x^{(i)}) + f(x^{(i)}) = 0$$

Rearrange to solve for $x^{(i+1)}$:

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

when do we stop?

in your pseudo-code we iterated for 10 loops

but what other stopping conditions could we check for?

```
double xi = 3
for i = 1 to 10 {
    fx = 0.1*xi^3 + 0.25*xi + 0.5
    f'x = 0.3*xi^2 + 0.25
    xi = xi - (fx / f'x)
}
```

when do we stop?

change your pseudo-code to implement **all three** of these:

- **residual size:**

when function gets “close enough” to root

$$|f(x^{(i)})| < \epsilon$$

- **increment size:**

when step size is “small”

$$|x^{(i+1)} - x^{(i)}| < \epsilon$$

- **number of iterations:**

when done “too many” steps

$$i > i_{\max}$$

Part 3

Newton's method for higher dimensions

we'll use a nice example from:

Math, Numerics, & Programming (for Mechanical Engineers)
by Masayuki Yano, James Douglass Penn, George Konidaris, Anthony T Patera

https://ocw.mit.edu/courses/mechanical-engineering/2-086-numerical-computation-for-mechanical-engineers-fall-2012/readings/MIT2_086F12_notes_unit4.pdf

say we want to solve this system:

$$f_1(z_1, z_2) = z_1^2 + 2z_2^2 - 22 = 0 ,$$

$$f_2(z_1, z_2) = 2z_1^2 + z_2^2 - 17 = 0 .$$

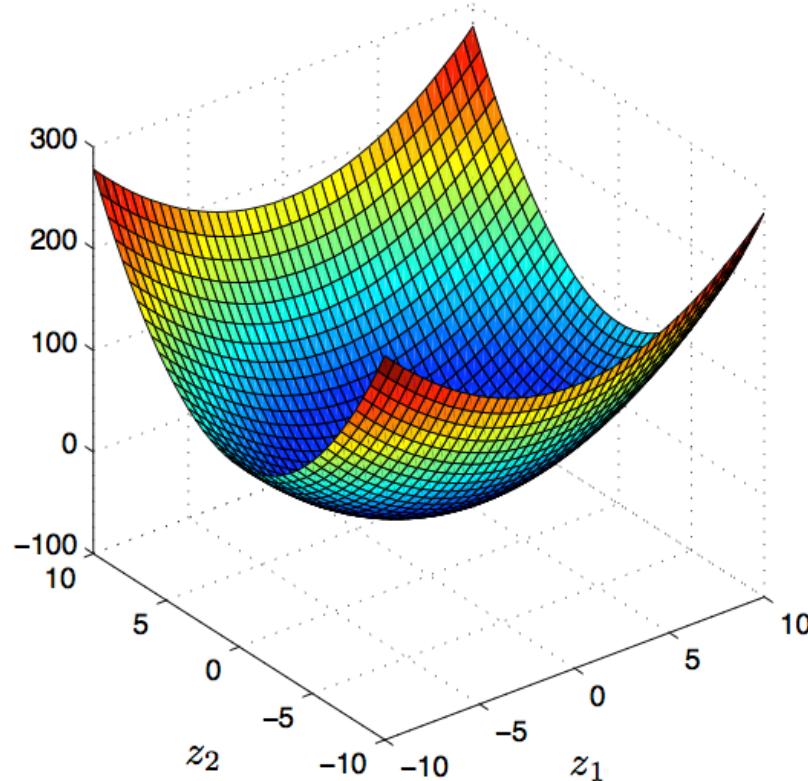
Newton procedure is same:

1. guess initial value $\mathbf{X}^{(0)}$
2. make linear approximation at $\mathbf{X}^{(0)}$
3. solve linearised system to get better guess $\mathbf{X}^{(1)}$
4. repeat from step 2 with $\mathbf{X}^{(1)}$

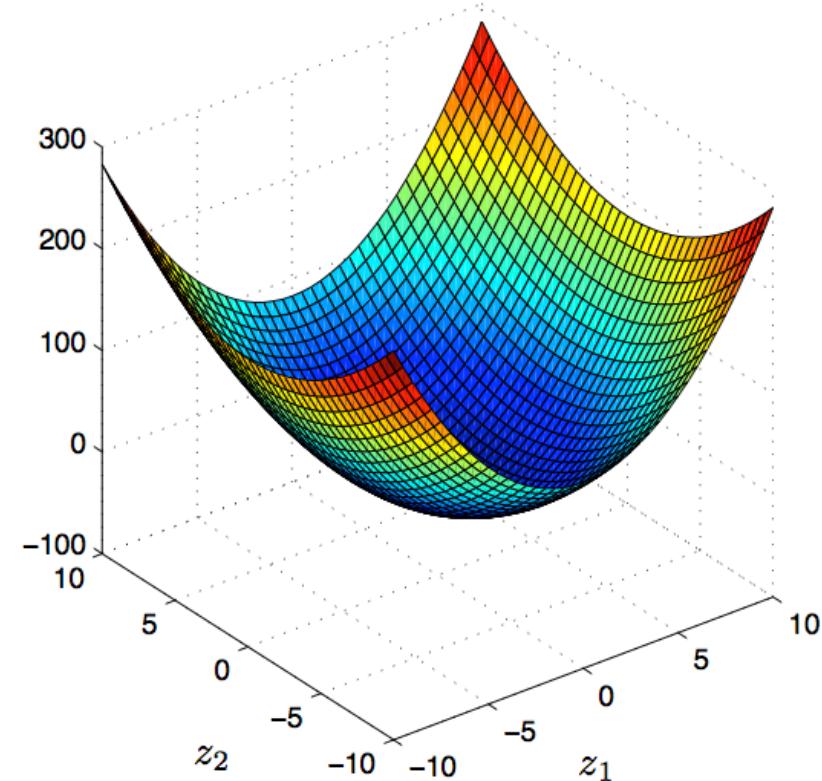
$$f_1(z_1, z_2) = z_1^2 + 2z_2^2 - 22 = 0,$$

$$f_2(z_1, z_2) = 2z_1^2 + z_2^2 - 17 = 0.$$

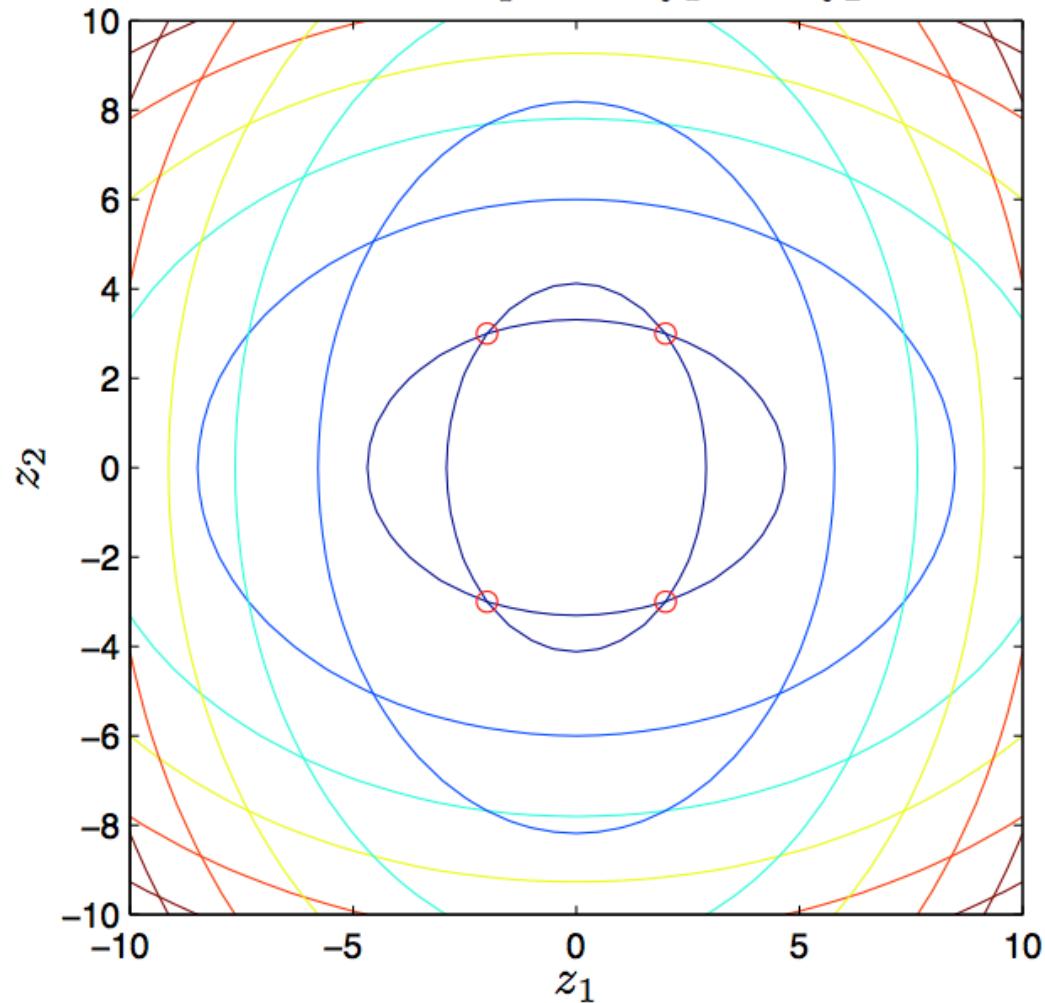
$$f_1(z_1, z_2) = z_1^2 + 2z_2^2 - 22$$



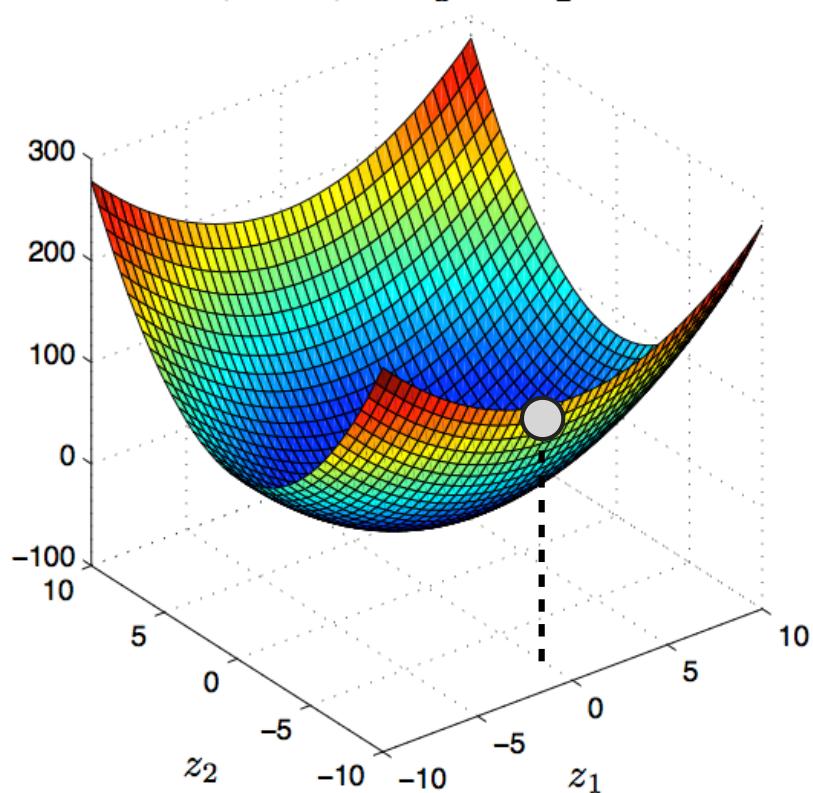
$$f_2(z_1, z_2) = 2z_1^2 + z_2^2 - 17 = 0$$



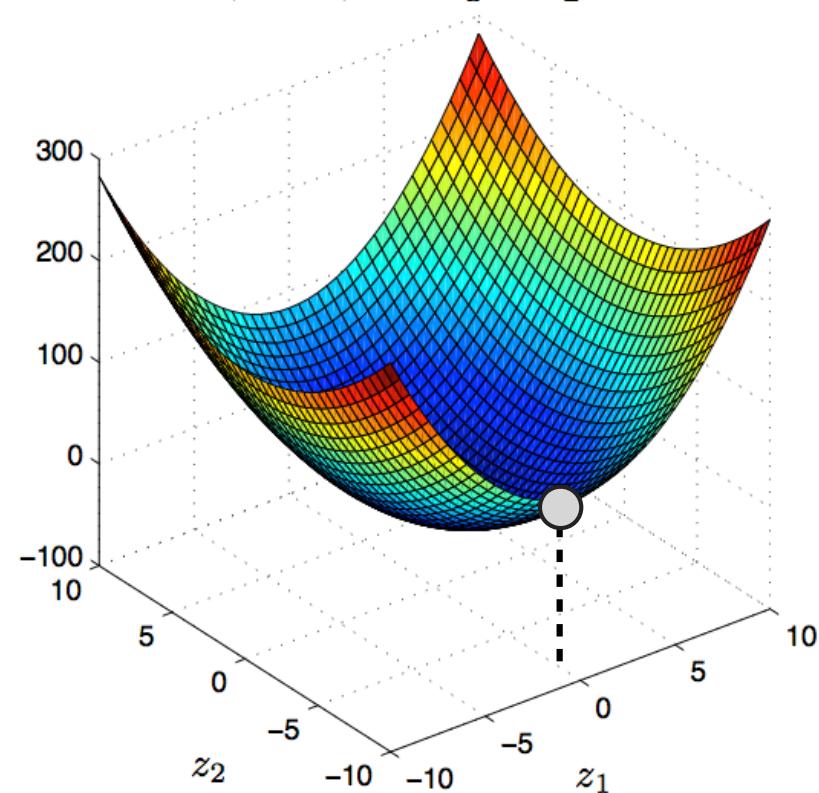
Contour plot of f_1 and f_2



$$f_1(z_1, z_2) = z_1^2 + 2z_2^2 - 22$$

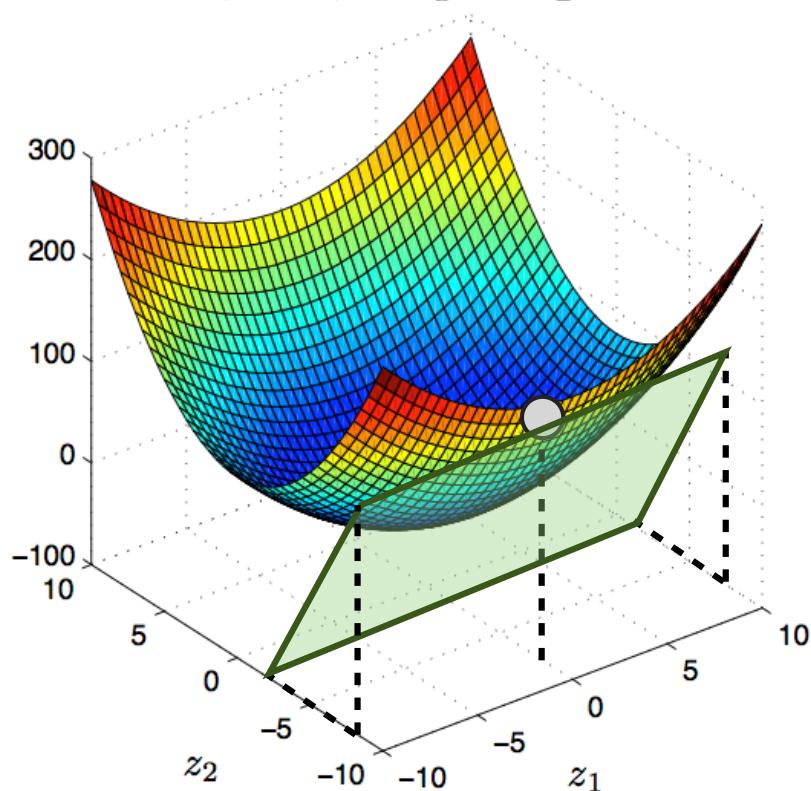


$$f_2(z_1, z_2) = 2z_1^2 + z_2^2 - 17$$

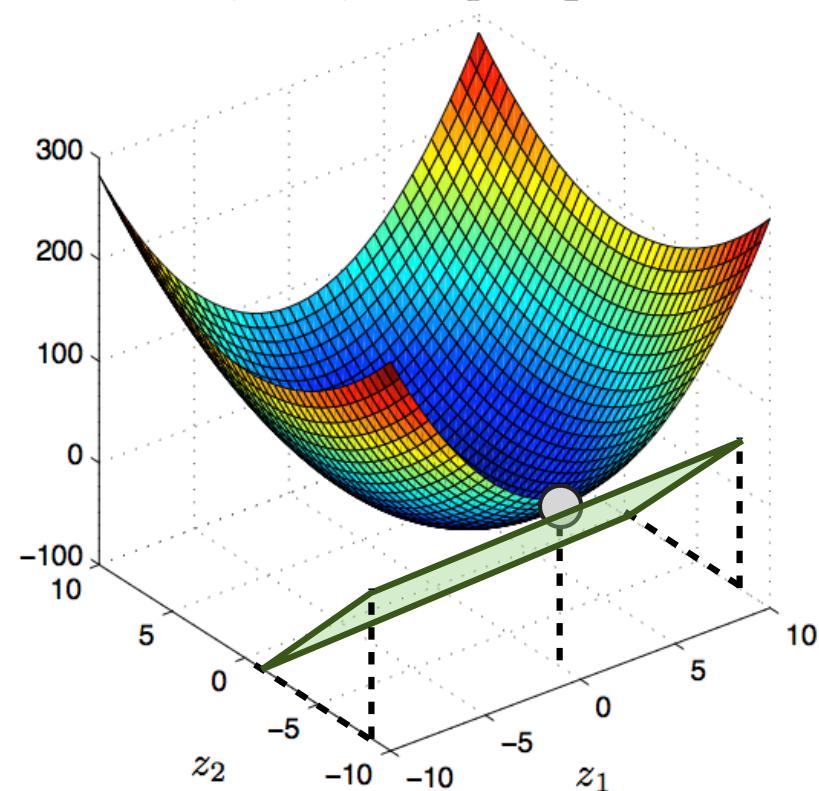


STEP I. guess a point, e.g. $X^{(0)}=(0, -8)$

$$f_1(z_1, z_2) = z_1^2 + 2z_2^2 - 22$$

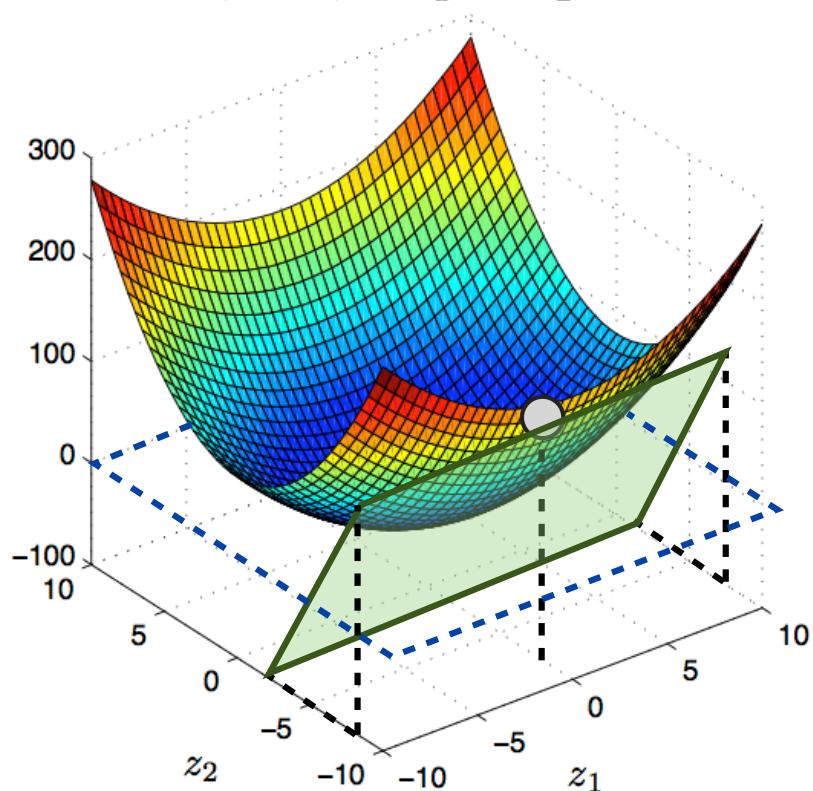


$$f_2(z_1, z_2) = 2z_1^2 + z_2^2 - 17$$

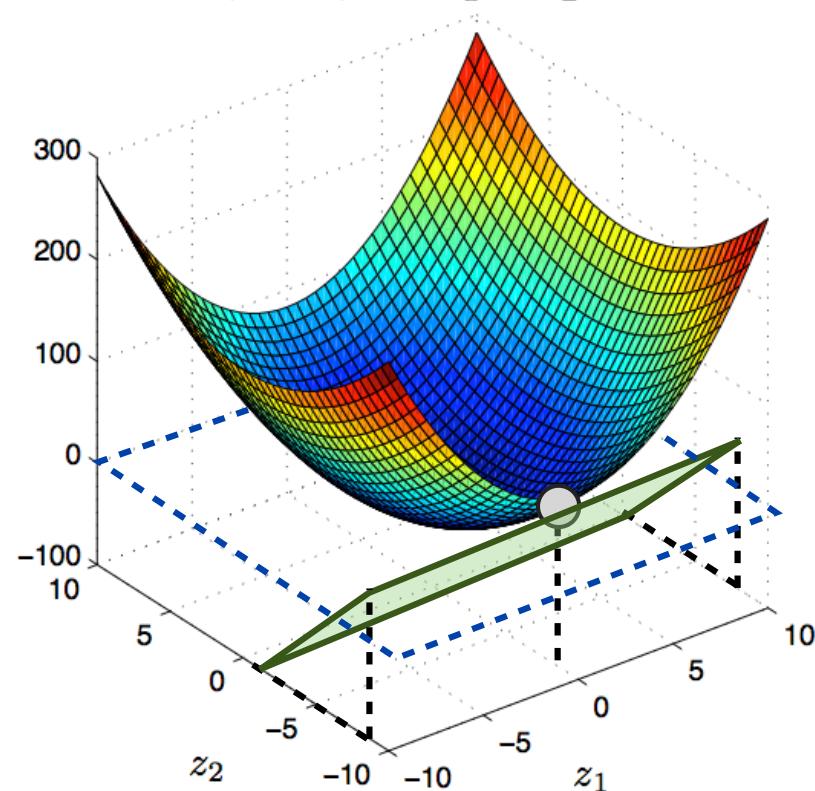


STEP 2. “linearise” function at point $X^{(0)}=(0, -8)$
(linear approximation)

$$f_1(z_1, z_2) = z_1^2 + 2z_2^2 - 22$$

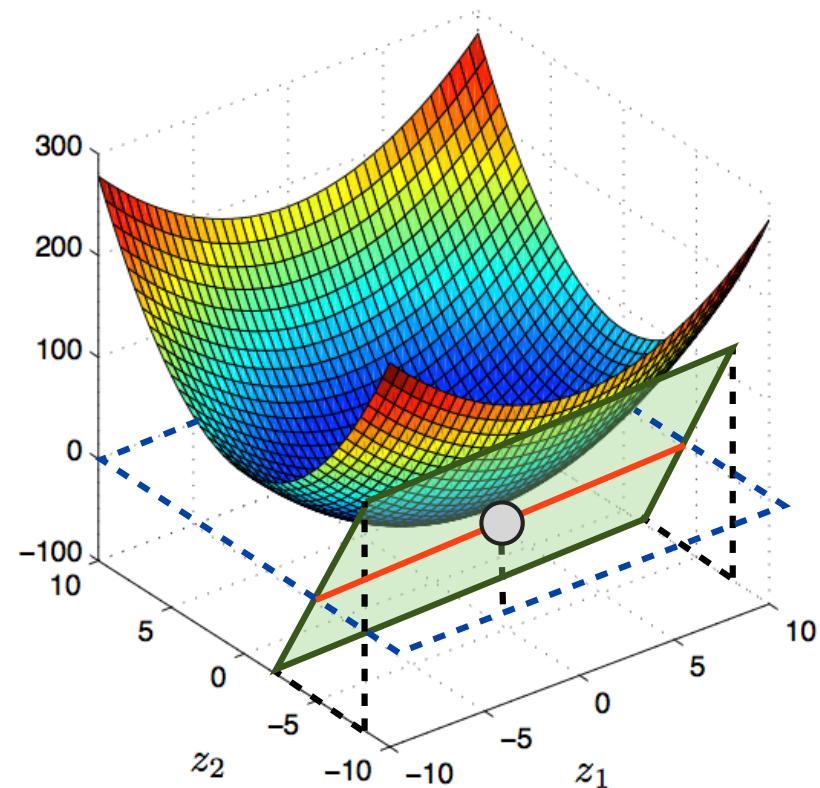


$$f_2(z_1, z_2) = 2z_1^2 + z_2^2 - 17$$

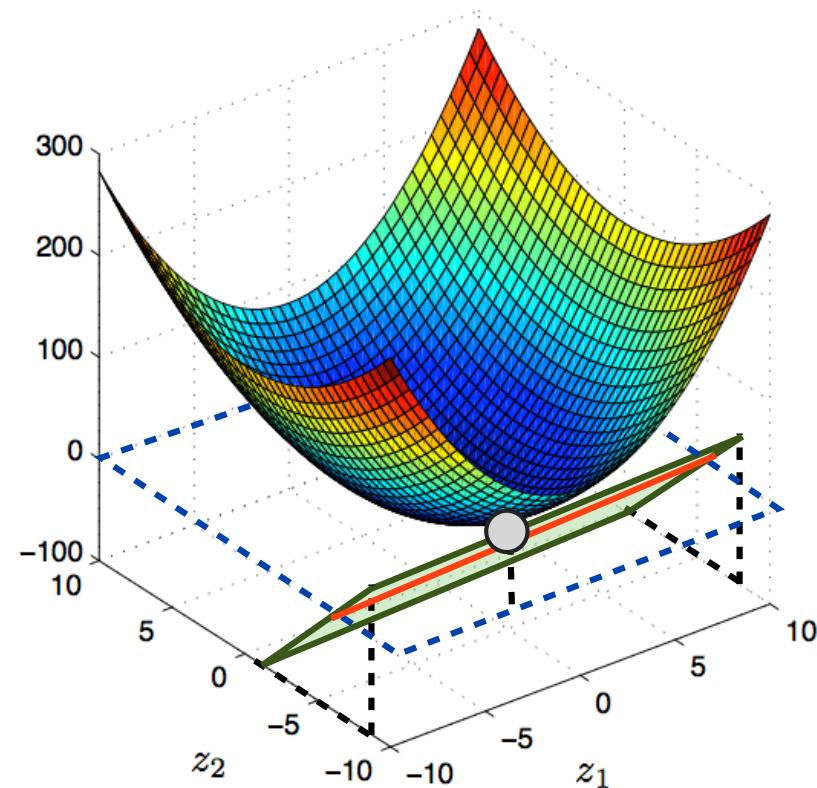


STEP 3. find new $X^{(i+1)}$ where both planes cross 0
(solve linearised system)

$$f_1(z_1, z_2) = z_1^2 + 2z_2^2 - 22$$

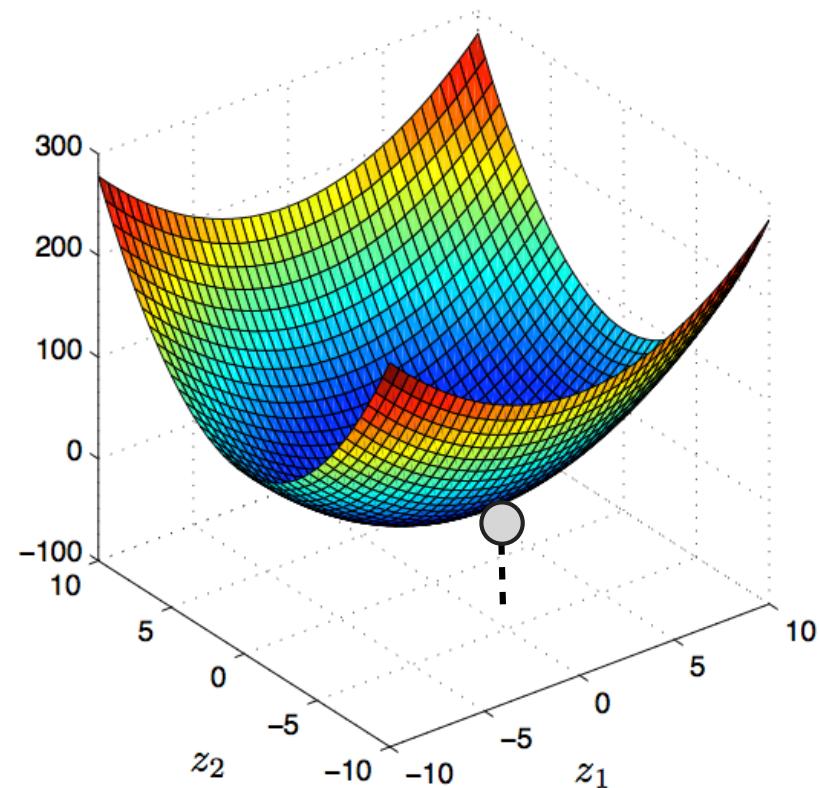


$$f_2(z_1, z_2) = 2z_1^2 + z_2^2 - 17$$

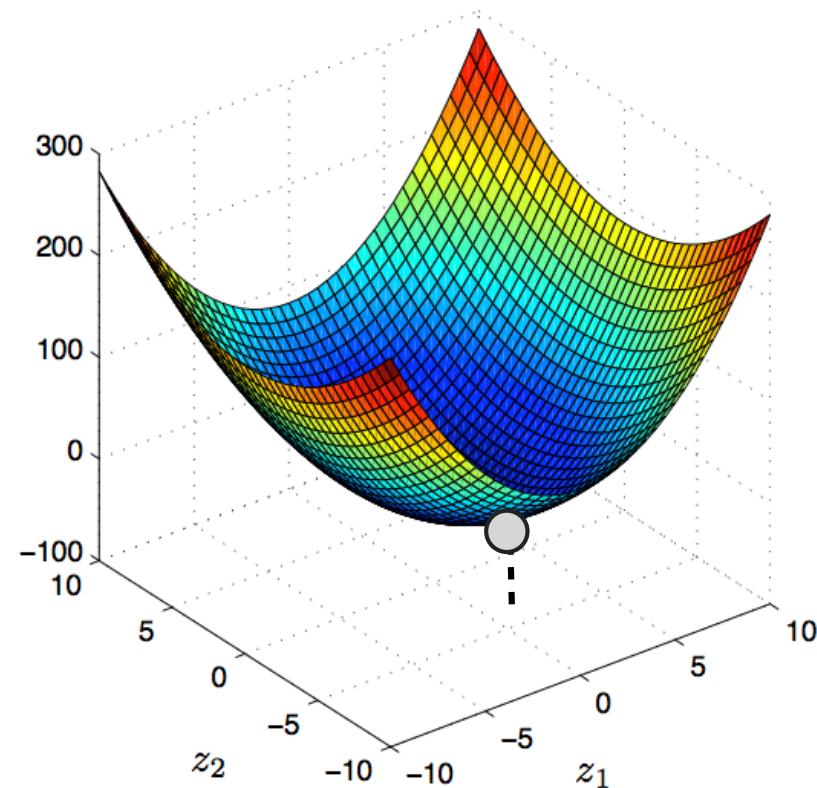


STEP 3. find new $X^{(i+1)}$ where both planes cross 0
(solve linearised system)

$$f_1(z_1, z_2) = z_1^2 + 2z_2^2 - 22$$

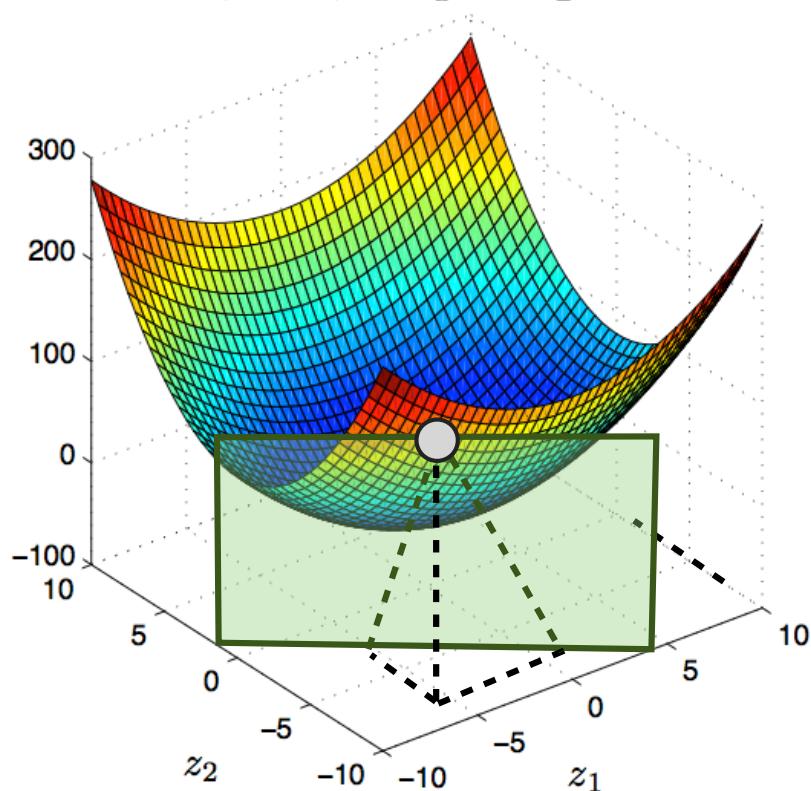


$$f_2(z_1, z_2) = 2z_1^2 + z_2^2 - 17$$

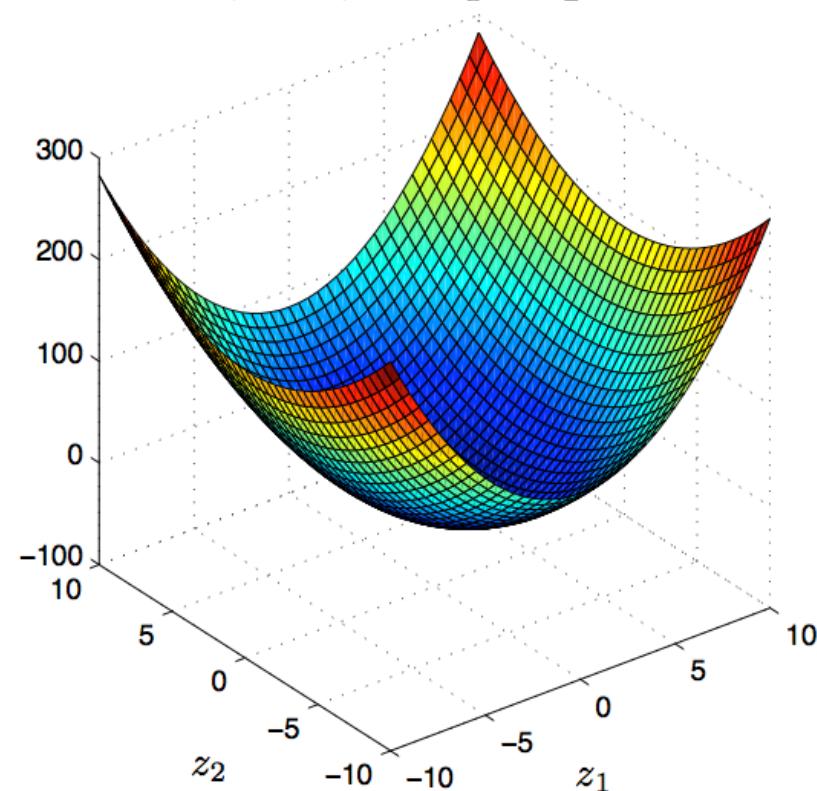


we've found our next guess $X^{(i+1)}$
(repeat from step 2...)

$$f_1(z_1, z_2) = z_1^2 + 2z_2^2 - 22$$

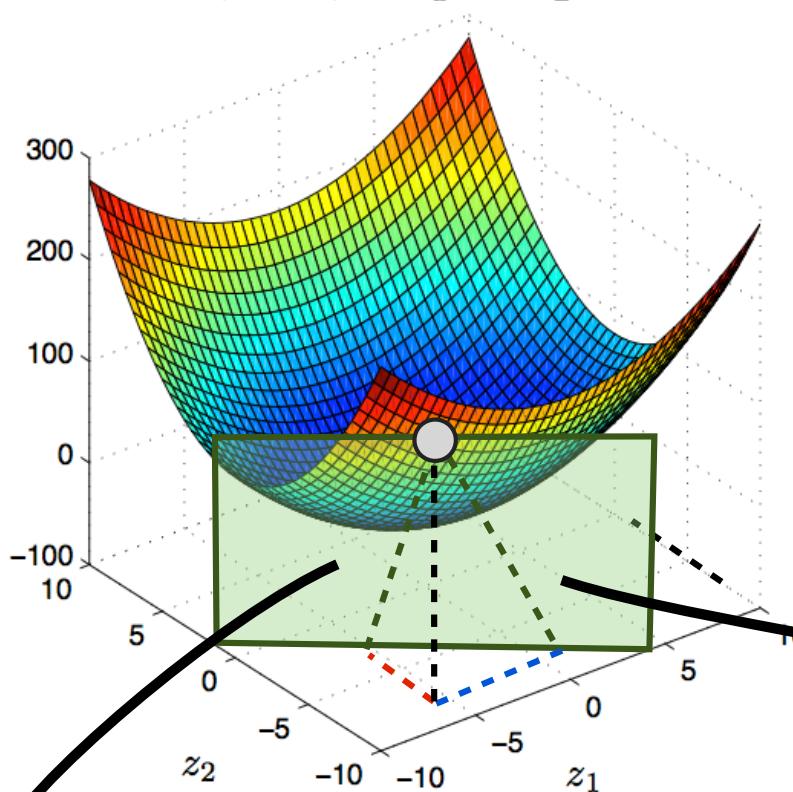


$$f_2(z_1, z_2) = 2z_1^2 + z_2^2 - 17$$



in step 2, instead of derivative (1D case) we need
gradient of f_1, f_2 to linearise

$$f_1(z_1, z_2) = z_1^2 + 2z_2^2 - 22$$



plane defined by:
(partial) derivatives

change
in f_1

change
in f_1

change in z_2

change in z_1

$$f_1(z_1,z_2) \;\; = \;\; z_1^2 + 2z_2^2 - 22 \;\; = \;\; 0 \;,$$

$$f_2(z_1,z_2) \;\; = \;\; 2z_1^2 + z_2^2 - 17 \;\; = \;\; 0 \;.$$

$$f_1(\underline{z_1}, z_2) = z_1^2 + 2z_2^2 - 22 = 0 ,$$

$$f_2(z_1, z_2) = 2z_1^2 + z_2^2 - 17 = 0 .$$

first let's just consider f_1

...and first let's just consider z_1

$$f_1(\underline{z_1}, z_2) = z_1^2 + 2z_2^2 - 22 = 0 ,$$

$$f_2(z_1, z_2) = 2z_1^2 + z_2^2 - 17 = 0 .$$

...treat z_2 like a constant

...then how does f_1 change as z_1 changes?

$$\frac{\partial f_1}{\partial z_1} = ?$$

$$f_1(\underline{z_1}, z_2) = z_1^2 + 2z_2^2 - 22 = 0 ,$$

$$f_2(z_1, z_2) = 2z_1^2 + z_2^2 - 17 = 0 .$$

...treat z_2 like a constant

...then how does f_1 change as z_1 changes?

$$\frac{\partial f_1}{\partial z_1} = 2z_1$$

$$f_1(z_1, \underline{z_2}) = z_1^2 + 2z_2^2 - 22 = 0 ,$$

$$f_2(z_1, z_2) = 2z_1^2 + z_2^2 - 17 = 0 .$$

and now only consider z_2

...treat z_1 like a constant

...then how does f_1 change as z_2 changes?

$$\frac{\partial f_1}{\partial z_2} = ?$$

$$f_1(z_1, z_2) = z_1^2 + 2z_2^2 - 22 = 0 ,$$

$$\underline{f_2(z_1, z_2) = 2z_1^2 + z_2^2 - 17 = 0} .$$

then f_2 and z_1 ...

$$\frac{\partial f_2}{\partial z_1} = ?$$

$$f_1(z_1, z_2) = z_1^2 + 2z_2^2 - 22 = 0 ,$$

$$f_2(z_1, z_2) = 2z_1^2 + z_2^2 - 17 = 0 .$$

..and finally f_2 and z_2

$$\frac{\partial f_2}{\partial z_2} = ?$$

$$f_1(z_1, z_2) = z_1^2 + 2z_2^2 - 22 = 0 ,$$

$$f_2(z_1, z_2) = 2z_1^2 + z_2^2 - 17 = 0 .$$

	∂z_1	∂z_2
∂f_1	$\frac{\partial f_1}{\partial z_1}$	$\frac{\partial f_1}{\partial z_2}$
∂f_2	$\frac{\partial f_2}{\partial z_1}$	$\frac{\partial f_2}{\partial z_2}$

$$f_1(z_1, z_2) = z_1^2 + 2z_2^2 - 22 = 0 ,$$

$$f_2(z_1, z_2) = 2z_1^2 + z_2^2 - 17 = 0 .$$

	∂z_1	∂z_2
∂f_1	$2z_1$	$4z_2$
∂f_2	$\frac{\partial f_2}{\partial z_1}$	$\frac{\partial f_2}{\partial z_2}$

$$f_1(z_1, z_2) = z_1^2 + 2z_2^2 - 22 = 0 ,$$

$$f_2(z_1, z_2) = 2z_1^2 + z_2^2 - 17 = 0 .$$

∂f_1	∂z_1	∂z_2
∂f_2	$2z_1$	$4z_2$
	$4z_1$	$2z_2$

called the “Jacobian” (J)

$$J_{ij} = \frac{\partial f_i}{\partial z_j}$$

∂z_1	∂z_2	
∂f_1	$2z_1$	$4z_2$
∂f_2	$4z_1$	$2z_2$

called the “Jacobian” (J)

$$J_{ij} = \frac{\partial f_i}{\partial z_j}$$

$$J(X^{(i)}) = \begin{array}{|c|c|} \hline J_{11} & J_{12} \\ \hline \hline J_{21} & J_{22} \\ \hline \end{array}$$

...where $X^{(i)} = (z_1, z_2)$ at step i

called the “Jacobian” (J)

$$J_{ij} = \frac{\partial f_i}{\partial z_j}$$

$$J(X^{(i)}) =$$

sometimes might
just write:

$J^{(i)}$

J_{11}	J_{12}
J_{21}	J_{22}

...where $X^{(i)} = (z_1, z_2)$ at step i

1D case:

$$f_{\text{linear}}(x^{(i+1)}) \equiv f'(x^{(i)}) (x^{(i+1)} - x^{(i)}) + f(x^{(i)})$$

nD case:

$$f_{\text{linear}}(X^{(i)}) \equiv J(X^{(i)}) \begin{bmatrix} x_1^{(i+1)} - x_1^{(i)} \\ \vdots \\ x_n^{(i+1)} - x_n^{(i)} \end{bmatrix} + \begin{bmatrix} f_1(X^{(i)}) \\ \vdots \\ f_n(X^{(i)}) \end{bmatrix}$$

$$f_{\text{linear}}(x^{(i+1)}) \equiv f'(x^{(i)}) (x^{(i+1)} - x^{(i)}) + f(x^{(i)})$$

1 variable replaced by
n variables, $X = (x_1, \dots, x_n)$

$$f_{\text{linear}}(X^{(i)}) \equiv J(X^{(i)}) \begin{bmatrix} x_1^{(i+1)} - x_1^{(i)} \\ \vdots \\ x_n^{(i+1)} - x_n^{(i)} \end{bmatrix} + \begin{bmatrix} f_1(X^{(i)}) \\ \vdots \\ f_n(X^{(i)}) \end{bmatrix}$$

$$f_{\text{linear}}(x^{(i+1)}) \equiv f'(x^{(i)}) (x^{(i+1)} - x^{(i)}) + f(x^{(i)})$$

derivative
replaced by
Jakobian

$$f_{\text{linear}}(X^{(i)}) \equiv J(X^{(i)}) \begin{bmatrix} x_1^{(i+1)} - x_1^{(i)} \\ \vdots \\ x_n^{(i+1)} - x_n^{(i)} \end{bmatrix} + \begin{bmatrix} f_1(X^{(i)}) \\ \vdots \\ f_n(X^{(i)}) \end{bmatrix}$$

$$f_{\text{linear}}(x^{(i+1)}) \equiv f'(x^{(i)}) (x^{(i+1)} - x^{(i)}) + f(x^{(i)})$$

now comparing change in
 n variables

$$f_{\text{linear}}(X^{(i)}) \equiv J(X^{(i)}) \begin{bmatrix} x_1^{(i+1)} - x_1^{(i)} \\ \vdots \\ x_n^{(i+1)} - x_n^{(i)} \end{bmatrix} + \begin{bmatrix} f_1(X^{(i)}) \\ \vdots \\ f_n(X^{(i)}) \end{bmatrix}$$

$$f_{\text{linear}}(x^{(i+1)}) \equiv f'(x^{(i)}) (x^{(i+1)} - x^{(i)}) + f(x^{(i)})$$

now evaluating
n functions

$$f_{\text{linear}}(X^{(i)}) \equiv J(X^{(i)}) \begin{bmatrix} x_1^{(i+1)} - x_1^{(i)} \\ \vdots \\ x_n^{(i+1)} - x_n^{(i)} \end{bmatrix} + \begin{bmatrix} f_1(X^{(i)}) \\ \vdots \\ f_n(X^{(i)}) \end{bmatrix}$$

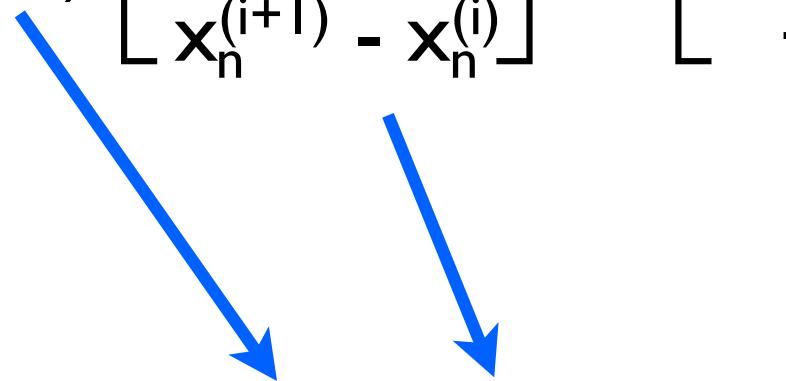
$$f_{\text{linear}}(x^{(i+1)}) \equiv f'(x^{(i)}) (x^{(i+1)} - x^{(i)}) + f(x^{(i)})$$

$$f_{\text{linear}}(X^{(i)}) \equiv J(X^{(i)}) \begin{bmatrix} x_1^{(i+1)} - x_1^{(i)} \\ \vdots \\ x_n^{(i+1)} - x_n^{(i)} \end{bmatrix} + \begin{bmatrix} f_1(X^{(i)}) \\ \vdots \\ f_n(X^{(i)}) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

set to zero to solve linearised system, i.e.
want to find nD point $(x_1^{(i+1)}, \dots, x_n^{(i+1)})$
where all our “hyperplanes” cross 0

$$f_{\text{linear}}(x^{(i+1)}) \equiv f'(x^{(i)}) (x^{(i+1)} - x^{(i)}) + f(x^{(i)})$$

$$f_{\text{linear}}(X^{(i)}) \equiv J(X^{(i)}) \begin{bmatrix} x_1^{(i+1)} - x_1^{(i)} \\ \vdots \\ x_n^{(i+1)} - x_n^{(i)} \end{bmatrix} + \begin{bmatrix} f_1(X^{(i)}) \\ \vdots \\ f_n(X^{(i)}) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$


 $f_{\text{linear}}(X^{(i+1)}) \equiv J(X^{(i)}) \Delta X^{(i)} + f(X^{(i)}) = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$

$$f_{\text{linear}}(x^{(i+1)}) \equiv f'(x^{(i)}) (x^{(i+1)} - x^{(i)}) + f(x^{(i)})$$

$$f_{\text{linear}}(X^{(i)}) \equiv J(X^{(i)}) \begin{bmatrix} x_1^{(i+1)} - x_1^{(i)} \\ \vdots \\ x_n^{(i+1)} - x_n^{(i)} \end{bmatrix} + \begin{bmatrix} f_1(X^{(i)}) \\ \vdots \\ f_n(X^{(i)}) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$J(X^{(i)}) \Delta X^{(i)} = -f(X^{(i)})$$

rearrange to solve for $\Delta X^{(i)}$

initial guess: $z_1=10, z_2=10$

$$J(X) = \begin{array}{|c|c|} \hline 2z_1 & 4z_2 \\ \hline 4z_1 & 2z_2 \\ \hline \end{array}$$

$$f_1(z_1, z_2) = z_1^2 + 2z_2^2 - 22$$

$$f_2(z_1, z_2) = 2z_1^2 + z_2^2 - 17$$

Q 1. what is $X^{(i)}$?

Q 2. what is $J(X^{(i)})$ and $-\mathbf{f}(X^{(i)})$?

$$\frac{J(X^{(i)})}{?} \quad \frac{\Delta X^{(i)}}{?} = \frac{-\mathbf{f}(X^{(i)})}{?}$$

Q 3. suppose: $\Delta X^{(i)} = \begin{bmatrix} -4.8 \\ -4.55 \end{bmatrix}$

then solve: $X^{(i+1)} = X^{(i)} + \Delta X^{(i)}$

initial guess: $z_1=10, z_2=10$

$$J(X) = \begin{array}{|c|c|} \hline 2z_1 & 4z_2 \\ \hline 4z_1 & 2z_2 \\ \hline \end{array} \quad \begin{aligned} f_1(z_1, z_2) &= z_1^2 + 2z_2^2 - 22 \\ f_2(z_1, z_2) &= 2z_1^2 + z_2^2 - 17 \end{aligned}$$

$$\frac{J(X^{(i)})}{?} \Delta X^{(i)} = \frac{-\mathbf{f}(X^{(i)})}{?}$$

initial guess: $z_1=10, z_2=10$

$$J(X) = \begin{array}{|c|c|} \hline 2z_1 & 4z_2 \\ \hline 4z_1 & 2z_2 \\ \hline \end{array}$$
$$\begin{aligned} f_1(z_1, z_2) &= z_1^2 + 2z_2^2 - 22 \\ f_2(z_1, z_2) &= 2z_1^2 + z_2^2 - 17 \end{aligned}$$

$$\begin{bmatrix} 20 & 40 \\ 40 & 20 \end{bmatrix} \Delta X^{(i)} = \begin{bmatrix} -278 \\ -283 \end{bmatrix}$$

initial guess: $z_1=10, z_2=10$

$$J(X) = \begin{array}{|c|c|} \hline 2z_1 & 4z_2 \\ \hline 4z_1 & 2z_2 \\ \hline \end{array}$$
$$\begin{aligned} f_1(z_1, z_2) &= z_1^2 + 2z_2^2 - 22 \\ f_2(z_1, z_2) &= 2z_1^2 + z_2^2 - 17 \end{aligned}$$

$$\begin{bmatrix} 20 & 40 \\ 40 & 20 \end{bmatrix} \Delta X^{(i)} = \begin{bmatrix} -278 \\ -283 \end{bmatrix}$$

$$\Delta X^{(i)} = \begin{bmatrix} -4.8 \\ -4.55 \end{bmatrix}$$

initial guess: $z_1=10, z_2=10$

$$J(X) = \begin{array}{|c|c|} \hline 2z_1 & 4z_2 \\ \hline 4z_1 & 2z_2 \\ \hline \end{array}$$
$$\begin{aligned} f_1(z_1, z_2) &= z_1^2 + 2z_2^2 - 22 \\ f_2(z_1, z_2) &= 2z_1^2 + z_2^2 - 17 \end{aligned}$$

$$\begin{bmatrix} 20 & 40 \\ 40 & 20 \end{bmatrix} \Delta X^{(i)} = \begin{bmatrix} -278 \\ -283 \end{bmatrix}$$

$$\Delta X^{(i)} = \begin{bmatrix} -4.8 \\ -4.55 \end{bmatrix}$$

$$X^{(i+1)} = X^{(i)} + \Delta X^{(i)}$$

initial guess: $z_1=10, z_2=10$

$$J(X) = \begin{array}{|c|c|} \hline 2z_1 & 4z_2 \\ \hline 4z_1 & 2z_2 \\ \hline \end{array} \quad \begin{aligned} f_1(z_1, z_2) &= z_1^2 + 2z_2^2 - 22 \\ f_2(z_1, z_2) &= 2z_1^2 + z_2^2 - 17 \end{aligned}$$

$$\begin{bmatrix} 20 & 40 \\ 40 & 20 \end{bmatrix} \Delta X^{(i)} = \begin{bmatrix} -278 \\ -283 \end{bmatrix}$$

$$\Delta X^{(i)} = \begin{bmatrix} -4.8 \\ -4.55 \end{bmatrix}$$

$$X^{(i+1)} = \begin{bmatrix} 5.2 \\ 5.45 \end{bmatrix}$$

if you want to solve the following system in MATLAB:

$$J \Delta X = -f$$

...then use backslash operator:

$$\Delta X = -J \backslash f$$

the is equivalent to:

$$\Delta X = -J^{-1} f$$

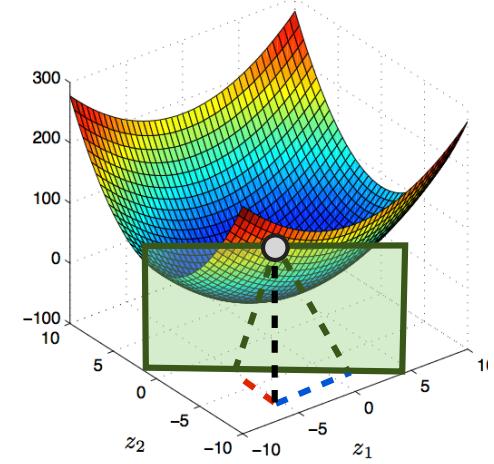
...but faster

notation...

$f'(x)$ 1st-order derivative (1D)

$\nabla f(X)$ gradient (nD)

$g(X)$ gradient (nD)



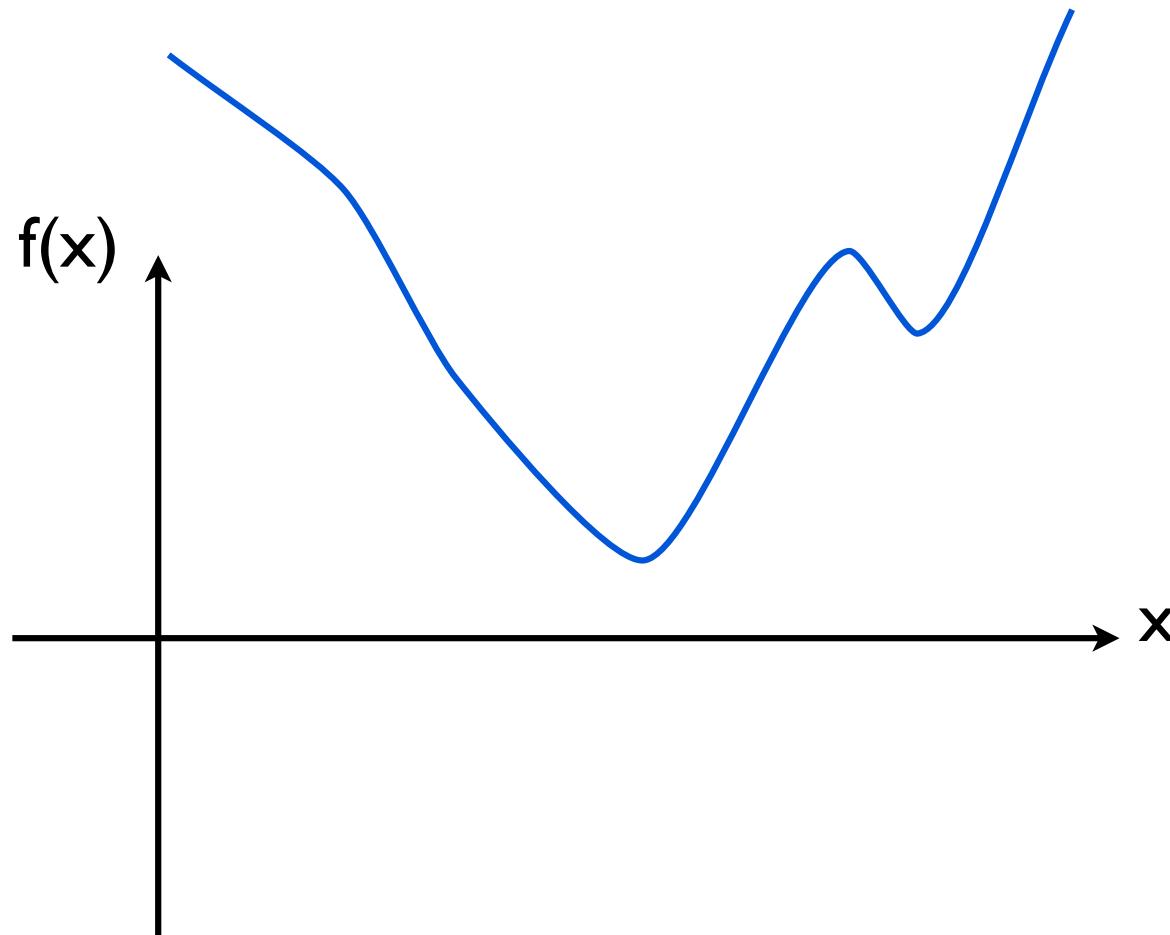
SUMMARY Part 3. nD Newton method

- still linear approximation, but now in n D
- Jacobian

Part 4

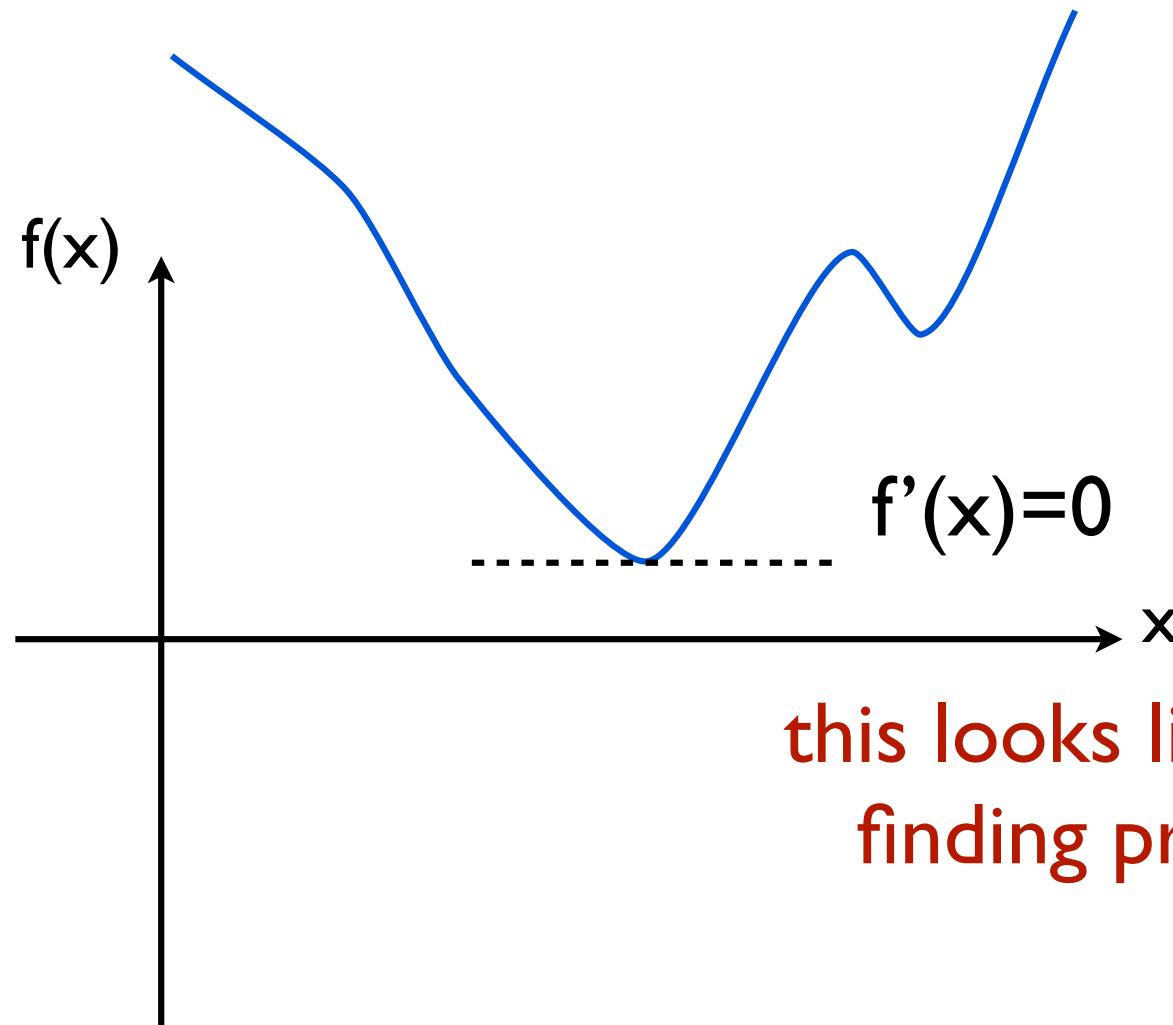
Newton's method for optimisation

can we apply same idea to find minimum?



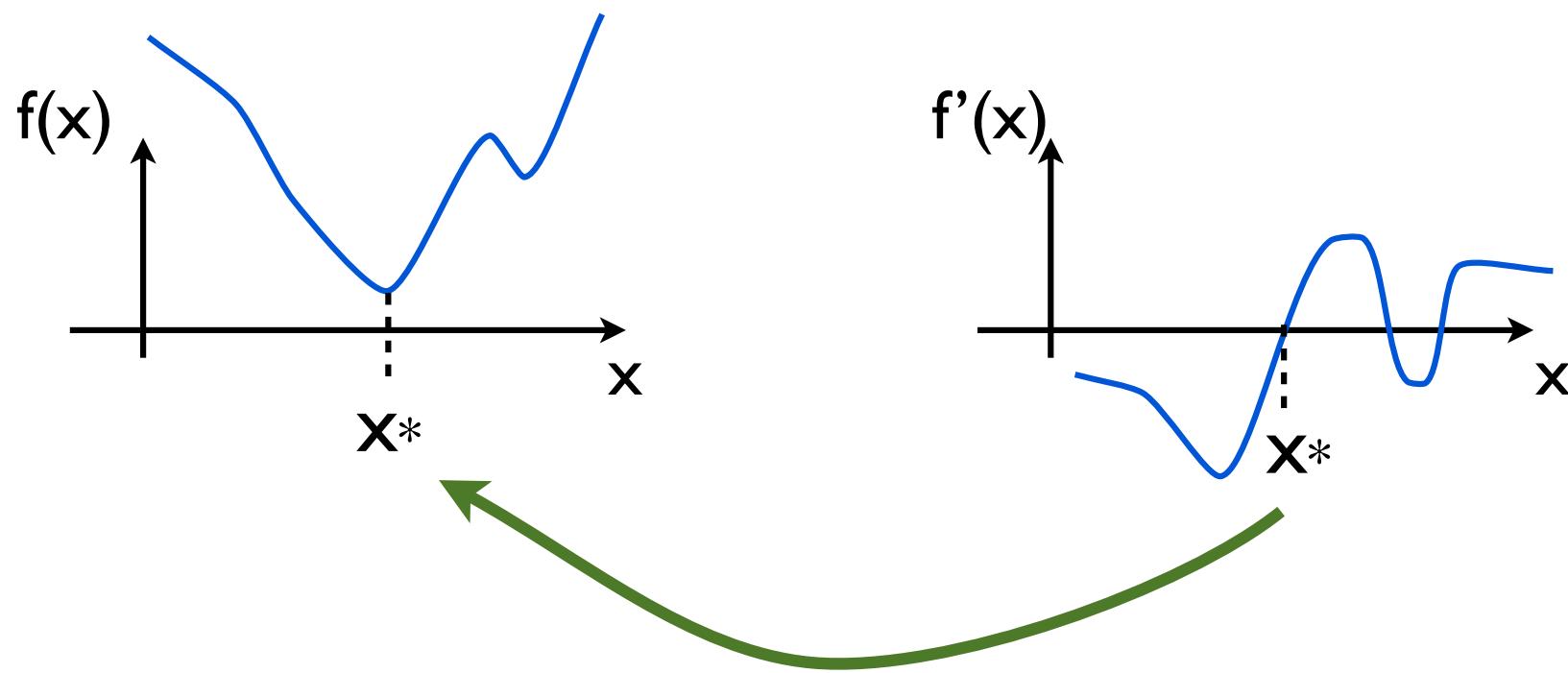
what is true about f at the minimum?

can we apply same idea to find minimum?



this looks like a root-finding problem...

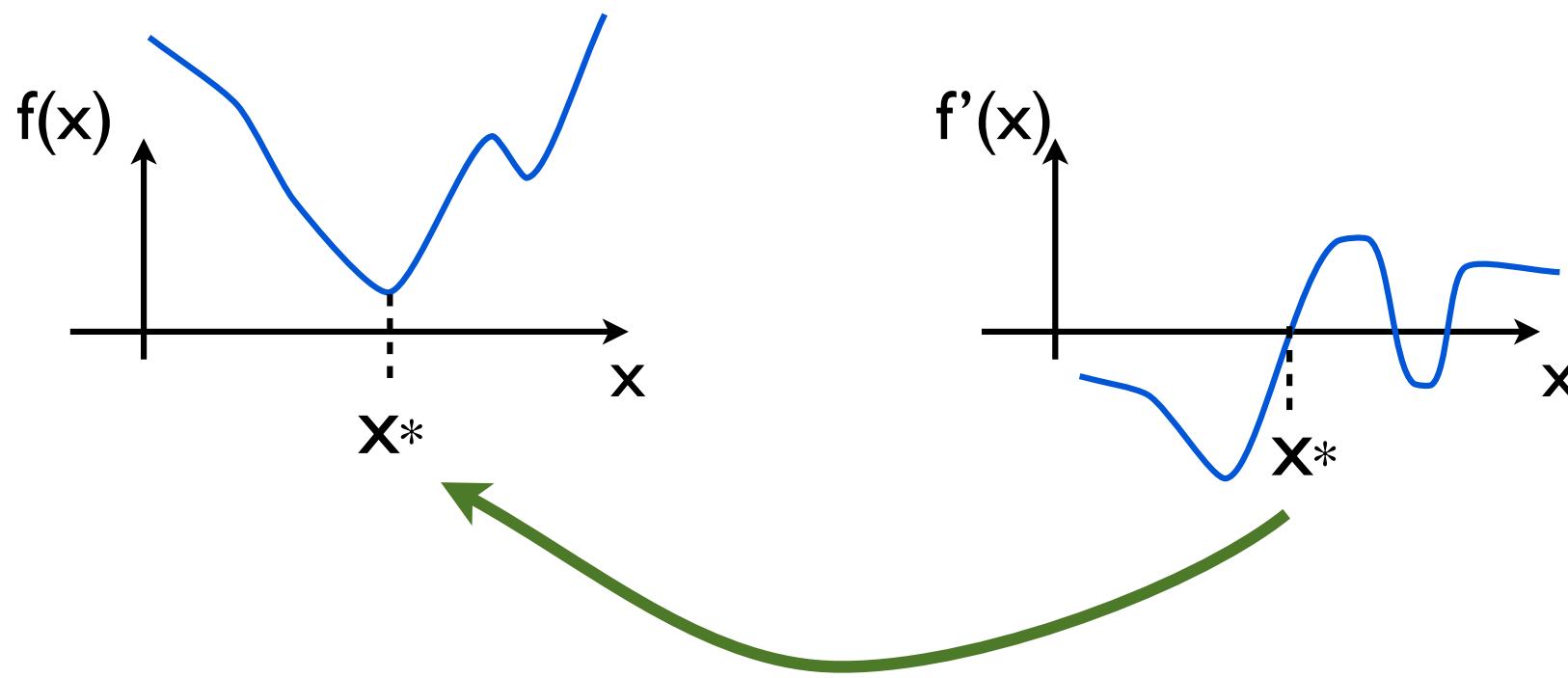
find root of $f'(x)$



that'll give us the minimum of $f(x)$

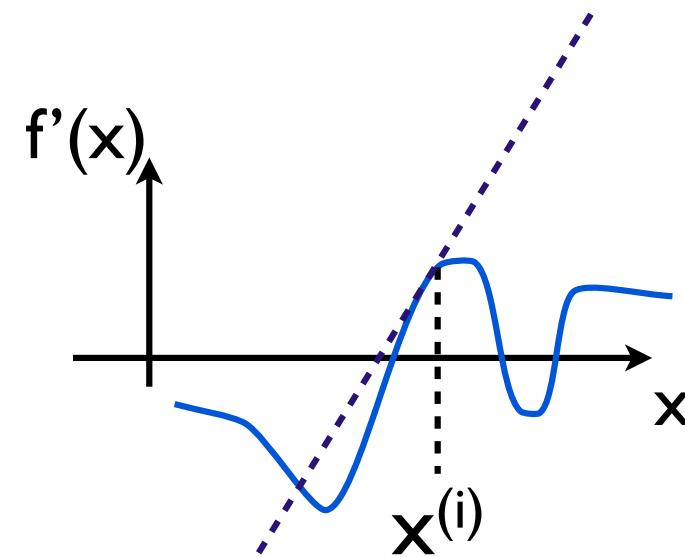
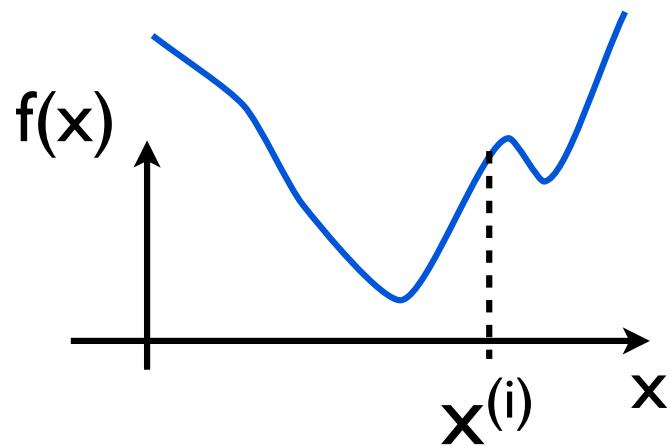
$$x^{(i+1)} = x^{(i)} - \frac{f'(x^{(i)})}{f''(x^{(i)})}$$

“update” formula
to find minimum



that'll give us the minimum of $f(x)$

here Newton's method approximates by **lines**



what is approximation in
original formula?
(see next slide)

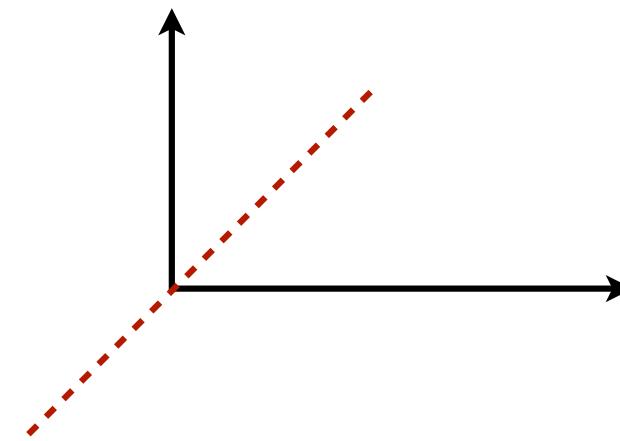
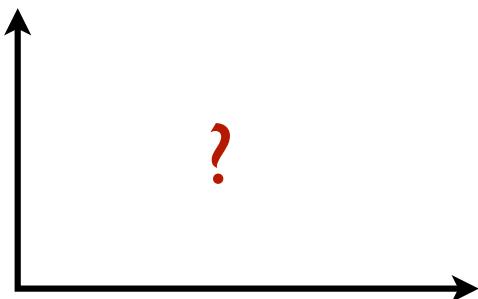
Step 1. discuss with your neighbour (2 mins)

?

$$f(x) = ?$$

line

$$f'(x) = 2x$$



Step 2. ...and in general?

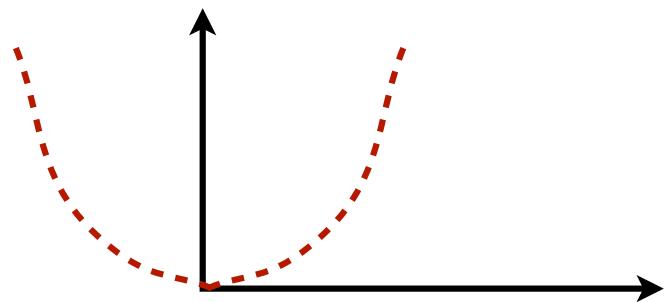
$$f(x) = ?$$

$$f'(x) = mx^1 + c$$



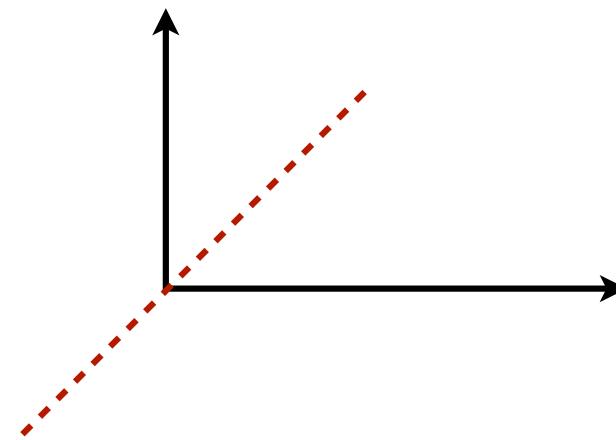
quadratic

$$f(x) = x^2$$



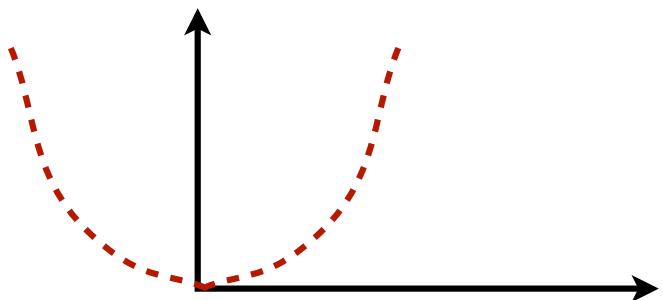
line

$$f'(x) = 2x$$



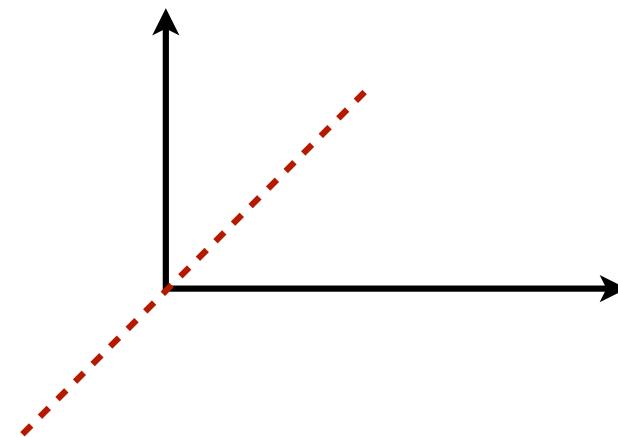
quadratic

$$f(x) = \left(\frac{m}{2}\right)x^2 + cx + k$$

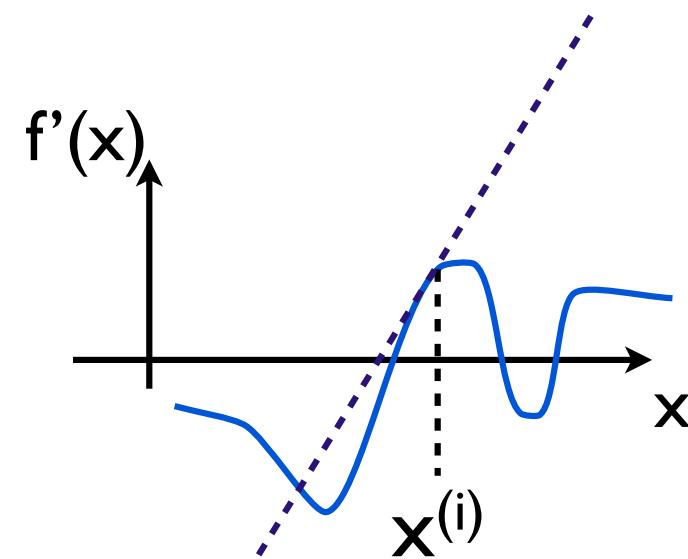
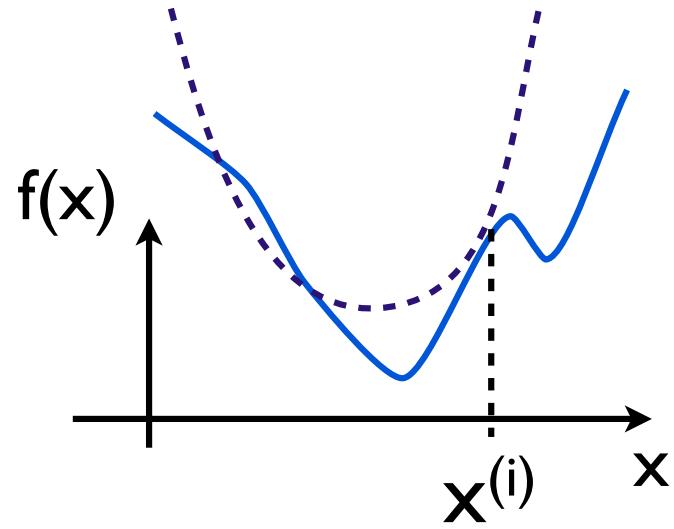


line

$$f'(x) = mx^1 + c$$

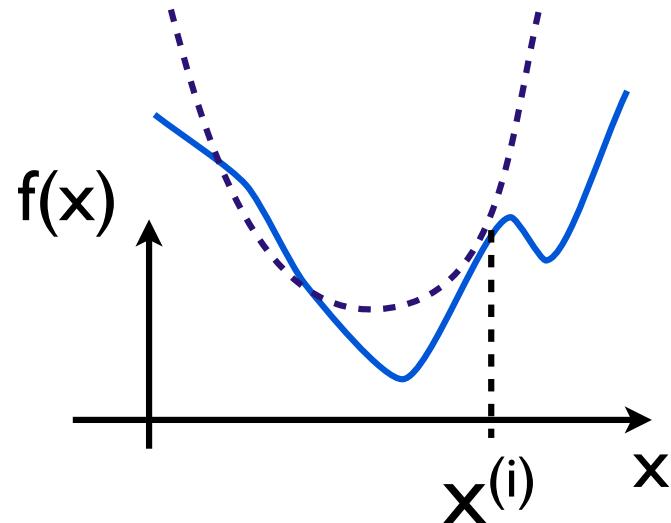


here Newton's method approximates by **lines**

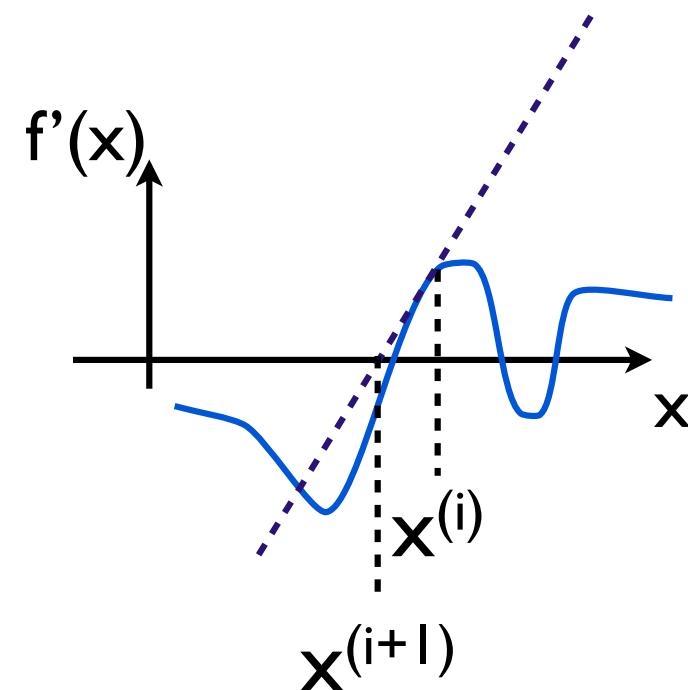


...so here Newton approximates by **quadratics**

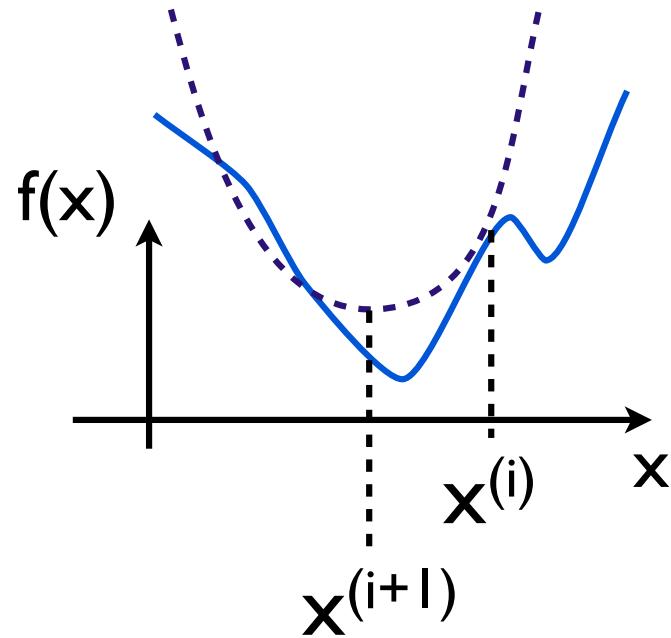
root of line at each step....



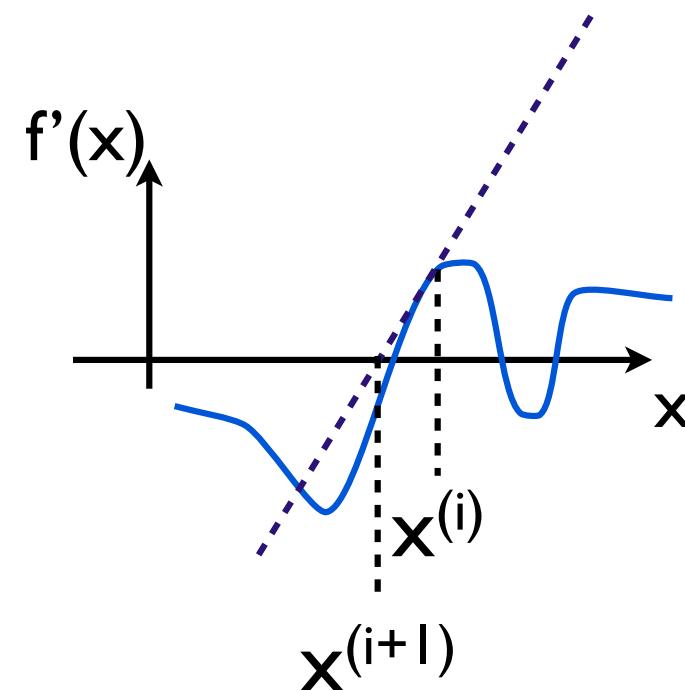
...corresponds to?



root of line at each step....



minimum of quadratic
approximation



Second-order Taylor expansion around x_n

$$f_T(x) = f_T(x_n + \Delta x) \approx f(x_n) + \underline{f'(x_n)\Delta x} + \underline{\frac{1}{2}f''(x_n)\Delta x^2}.$$

want to find Δx so $(x_n + \Delta x)$ is stationary point:

when is $f_T(x_n + \Delta x)$ a stationary point?

Second-order Taylor expansion around x_n

$$f_T(x) = f_T(x_n + \Delta x) \approx f(x_n) + \underline{f'(x_n)\Delta x} + \underline{\frac{1}{2}f''(x_n)\Delta x^2}.$$

want to find Δx so $(x_n + \Delta x)$ is stationary point:

when is $f_T(x_n + \Delta x)$ a stationary point?

...when derivative is **zero**

Second-order Taylor expansion around x_n

$$f_T(x) = f_T(x_n + \Delta x) \approx f(x_n) + \underline{f'(x_n)\Delta x} + \underline{\frac{1}{2}f''(x_n)\Delta x^2}.$$

want to find Δx so $(x_n + \Delta x)$ is stationary point:

$$0 = \frac{d}{d\Delta x} \left(f(x_n) + f'(x_n)\Delta x + \frac{1}{2}f''(x_n)\Delta x^2 \right)$$

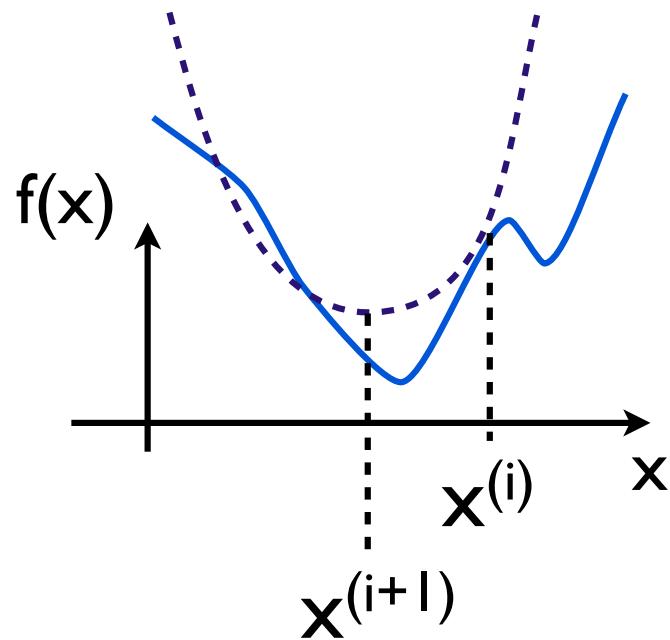
Second-order Taylor expansion around x_n

$$f_T(x) = f_T(x_n + \Delta x) \approx f(x_n) + \underline{f'(x_n)\Delta x} + \underline{\frac{1}{2}f''(x_n)\Delta x^2}.$$

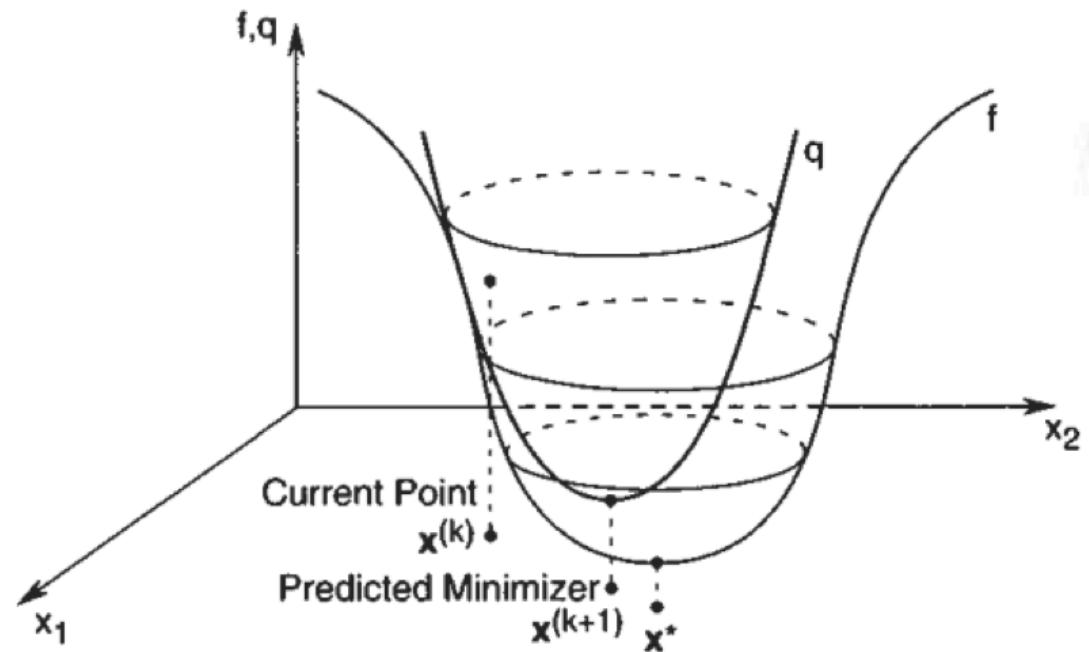
want to find Δx so $(x_n + \Delta x)$ is stationary point:

$$\begin{aligned} 0 &= \frac{d}{d\Delta x} \left(\cancel{f(x_n)} + f'(x_n)\Delta x + \frac{2}{2} \cancel{f''(x_n)\Delta x^2} \right) \\ &= f'(x_n) + f''(x_n)\Delta x. \end{aligned}$$

what about optimisation in higher dimensions?

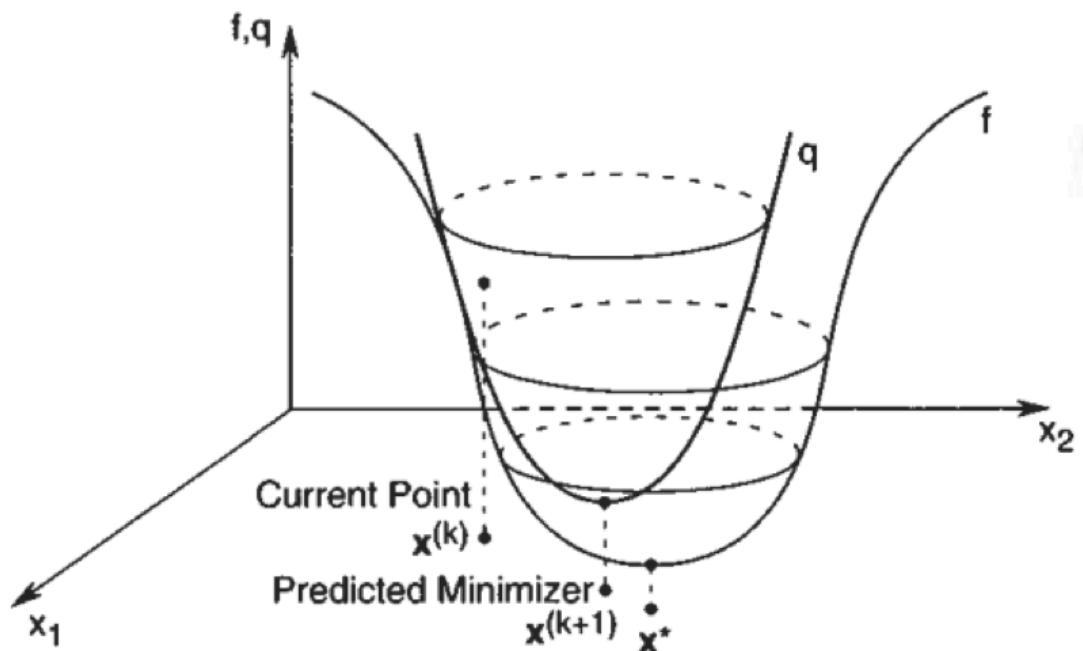
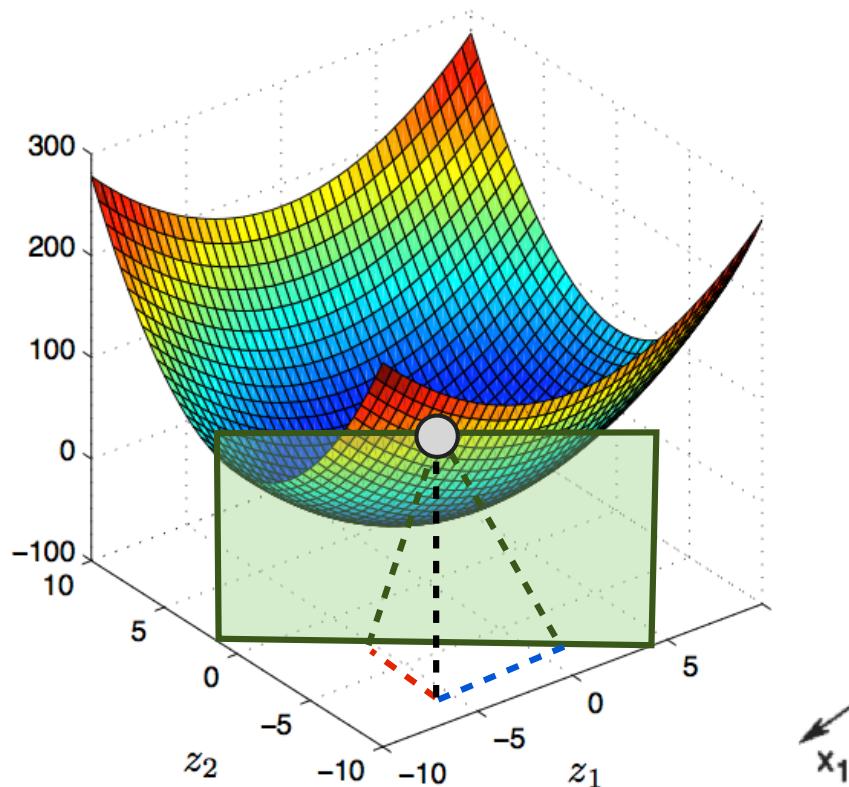


1D case



nD case

what about optimisation in higher dimensions?



nD root-finding:

$$\mathbf{X}^{(i+1)} = \mathbf{X}^{(i)} - \mathbf{J}(\mathbf{X}^{(i)})^{-1} \mathbf{f}(\mathbf{X}^{(i)})$$

“Jakobian”

nD optimisation:

$$\mathbf{X}^{(i+1)} = \mathbf{X}^{(i)} - \mathbf{H}(\mathbf{X}^{(i)})^{-1} \nabla \mathbf{f}(\mathbf{X}^{(i)})$$

“Hessian”

Jakobian

$$J_{ij} = \frac{\partial f_i}{\partial x_j}$$

first-order
partial derivatives

Hessian

$$H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

second-order
partial derivatives

Second-order Taylor expansion 1D case:

$$f_T(x) = f_T(x_n + \Delta x) \approx f(x_n) + f'(x_n)\Delta x + \frac{1}{2}f''(x_n)\Delta x^2.$$

Second-order Taylor expansion nD case:

$$f_T(\mathbf{X}) = f_T(\mathbf{X}^{(i)} + \Delta \mathbf{X}) \approx f(\mathbf{X}^{(i)}) + \nabla f(\mathbf{X}^{(i)}) \Delta \mathbf{X} + \frac{1}{2} \Delta \mathbf{X}^T \mathbf{H}(\mathbf{X}^{(i)}) \Delta \mathbf{X}$$

Second-order Taylor expansion 1D case:

$$f_T(x) = f_T(x_n + \Delta x) \approx f(x_n) + f'(x_n)\Delta x + \frac{1}{2}f''(x_n)\Delta x^2.$$

Second-order Taylor expansion nD case:

$$f_T(\mathbf{X}) = f_T(\mathbf{X}^{(i)} + \Delta \mathbf{X}) \approx f(\mathbf{X}^{(i)}) + \nabla f(\mathbf{X}^{(i)}) \Delta \mathbf{X} + \frac{1}{2} \underline{\Delta \mathbf{X}^T H(\mathbf{X}^{(i)}) \underline{\Delta \mathbf{X}}}$$

notation...

$f''(x)$ 2nd-order derivative (1D)

$H(X)$ Hessian of $f(X)$

$F(X)$ Hessian of $f(X)$

Summary 3.6. Advantages and disadvantages of Newton's method for unconstrained optimization problems

Advantages

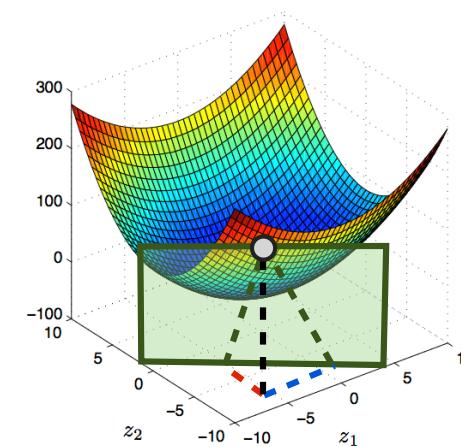
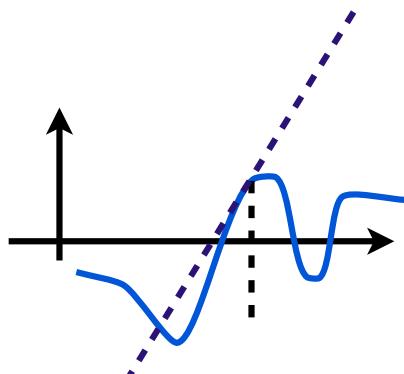
- 1° Quadratically convergent from a good starting point if $\nabla^2 f(\hat{x})$ is positive definite.
- 2° Simple and easy to implement.

Disadvantages

- 1° Not globally convergent for many problems.
- 2° May converge towards a maximum or saddle point of f .
- 3° The system of linear equations to be solved in each iteration may be ill-conditioned or singular.
- 4° Requires second order derivatives of f .

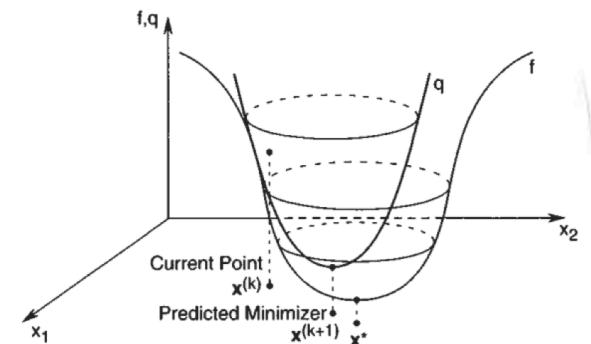
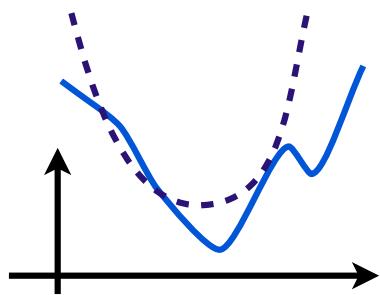
Newton method for root-finding

	ID	nD
at root	$f(x)=0$	$f(X)=0$
Newton's method	$x^{(i+1)} = x^{(i)} - \frac{f(x^{(i)})}{f'(x^{(i)})}$	$X^{(i+1)} = X^{(i)} - J^{(i)-1} f(X^{(i)})$



Newton method for optimisation

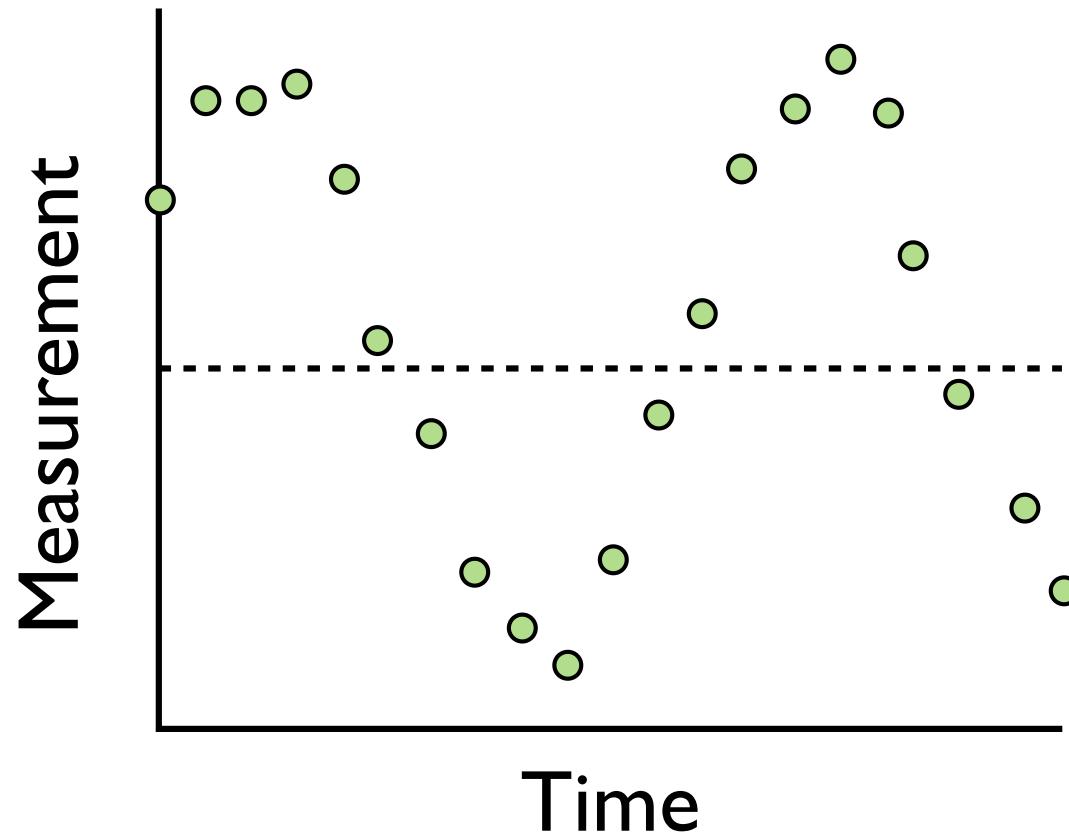
	ID	nD
at optimum	$f'(x)=0$	$\nabla f(X)=0$
Newton's method	$x^{(i+1)} = x^{(i)} - \frac{f'(x^{(i)})}{f''(x^{(i)})}$	$X^{(i+1)} = X^{(i)} - H^{(i)} - I \nabla f(X^{(i)})$



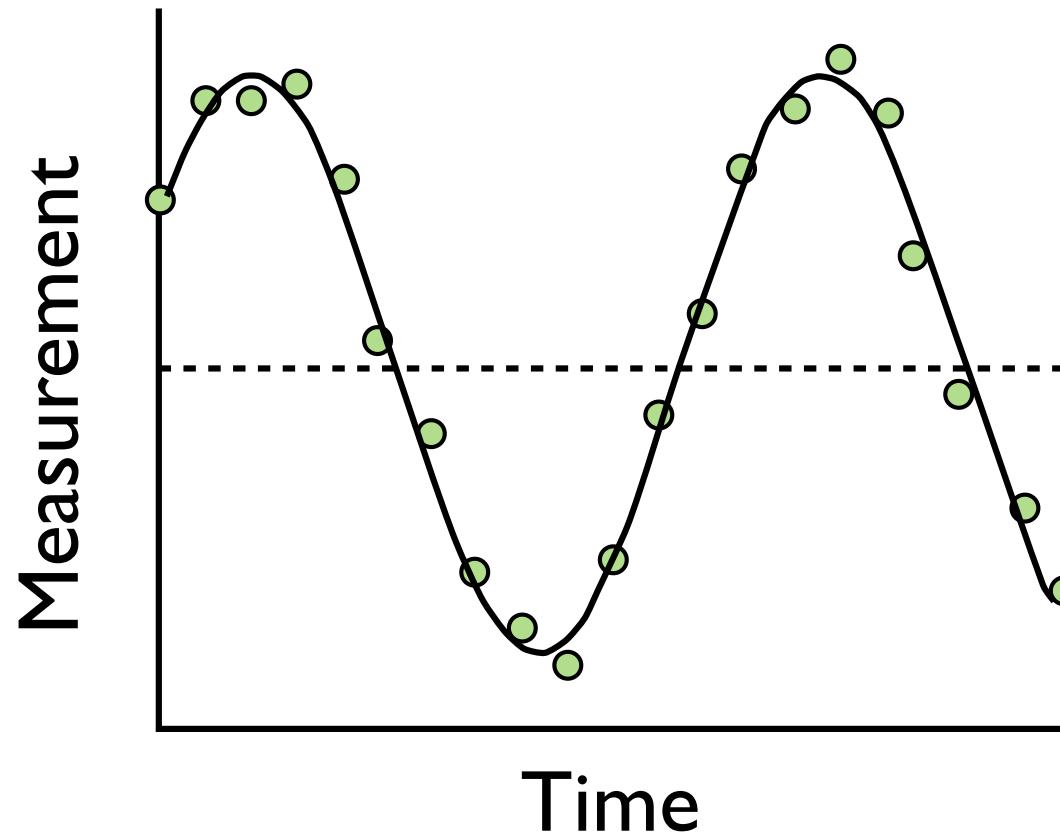
Part 5

curve fitting

curve fitting

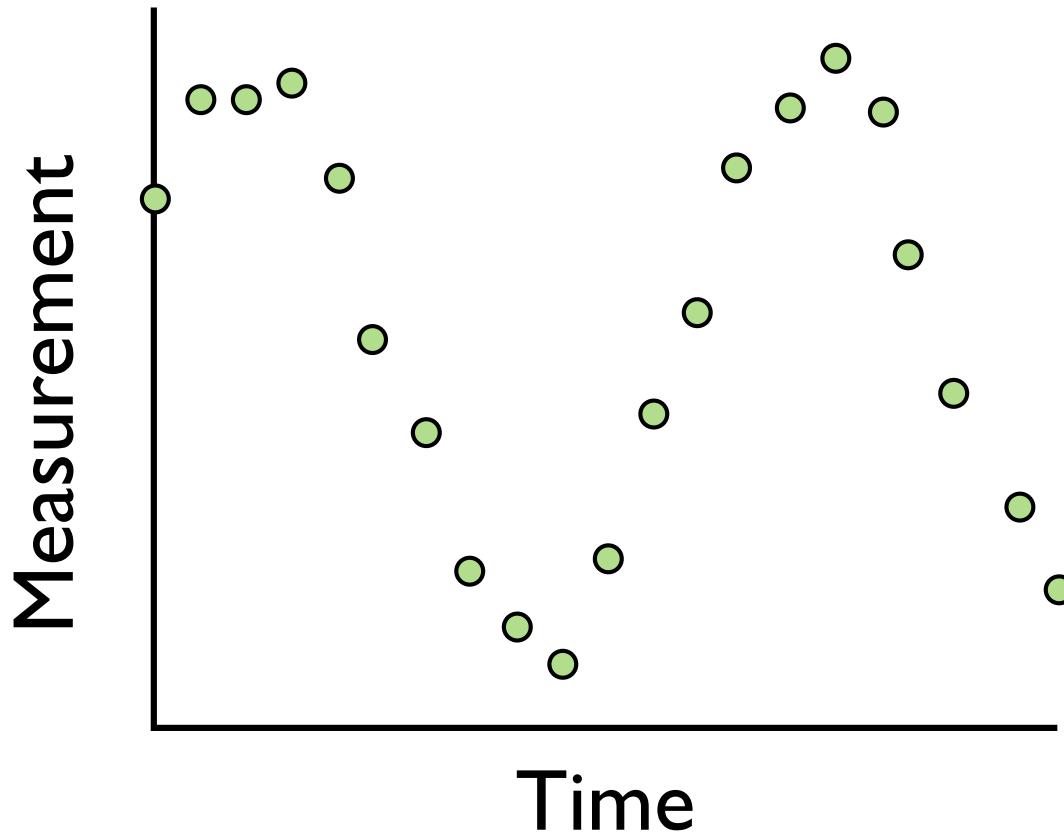


curve fitting



e.g. consider we try fitting the following function with **parameters**:

$$f(t) = A \sin(\omega t + \Phi)$$

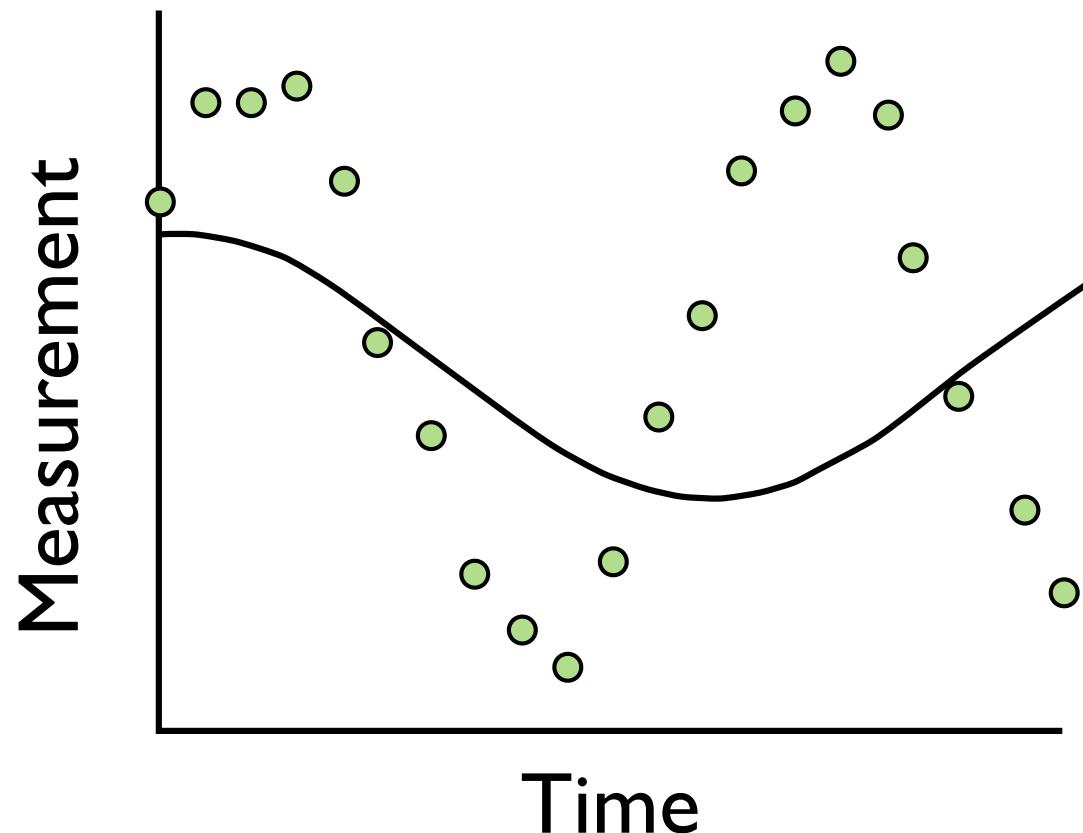


e.g. consider we try fitting the following function with parameters:

$$f(t) = A \sin(\omega t + \Phi)$$

(initial guess)

$$A=1.01 \quad \omega=0.592 \quad \Phi=1.541$$

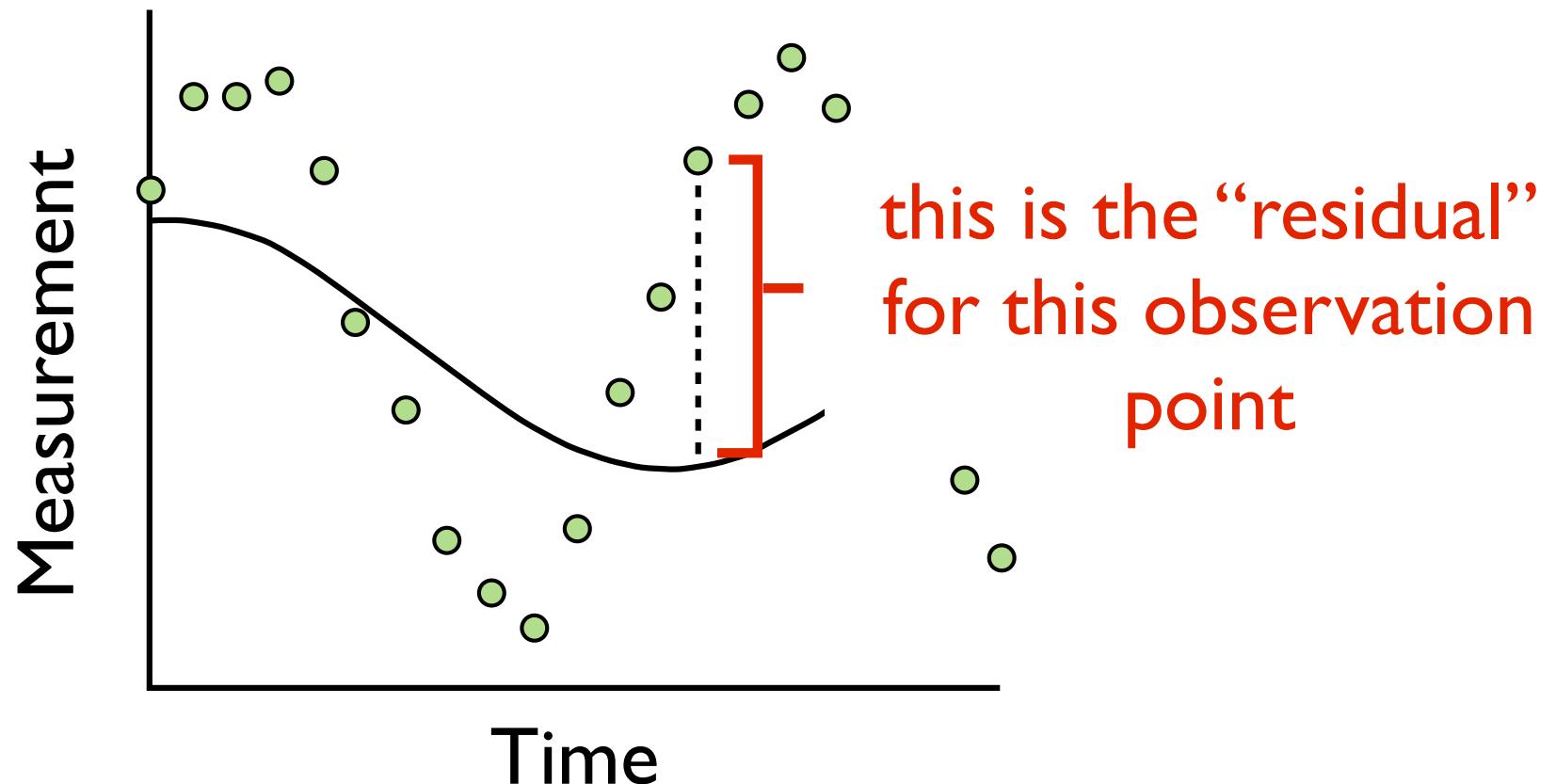


how can we measure how “close” we are to good fit?

e.g. consider we try fitting the following function with **parameters**:

$$f(t) = A \sin(\omega t + \Phi)$$

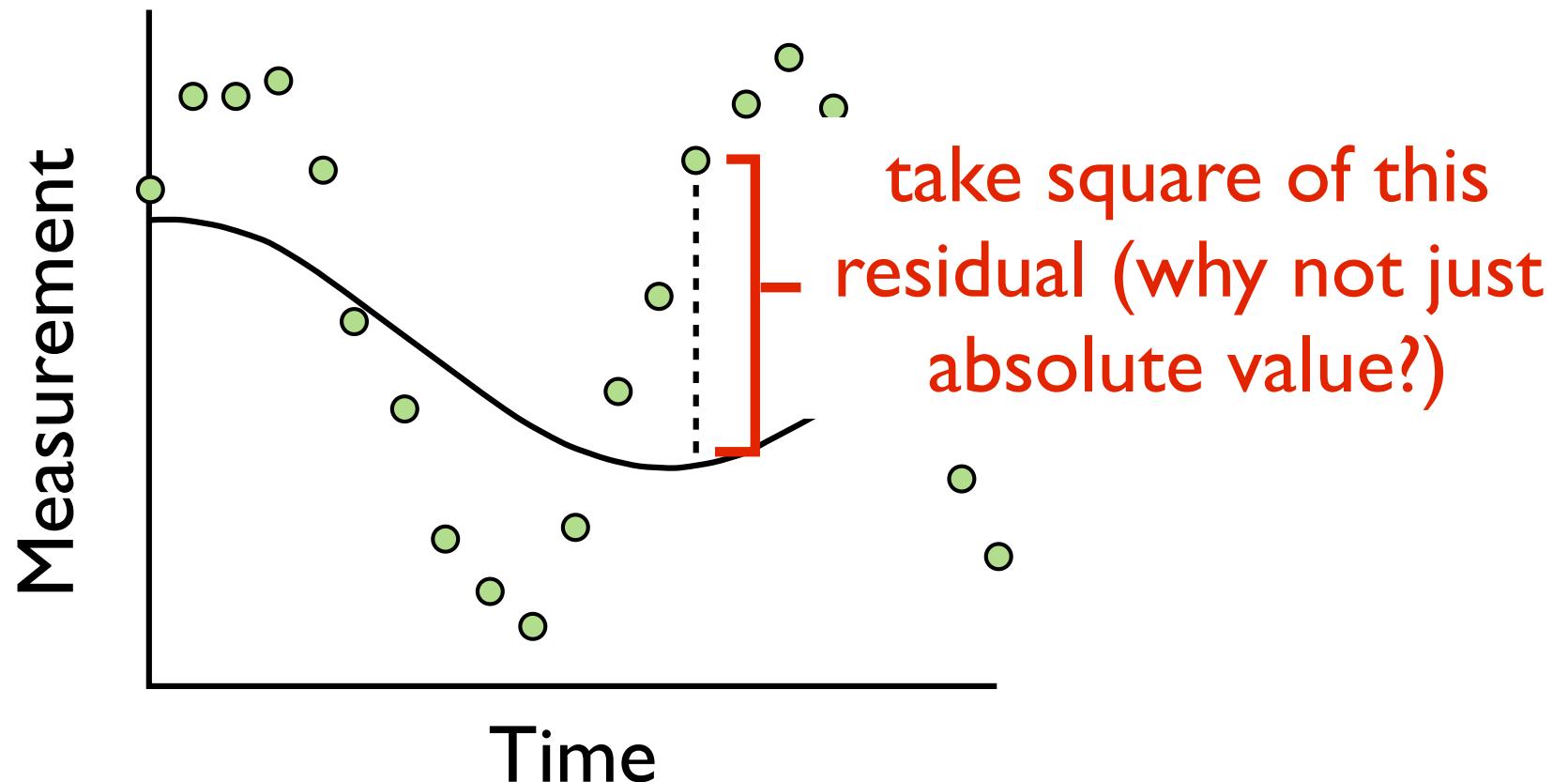
$$A=1.01 \quad \omega=0.592 \quad \Phi=1.541$$



e.g. consider we try fitting the following function with parameters:

$$f(t) = A \sin(\omega t + \Phi)$$

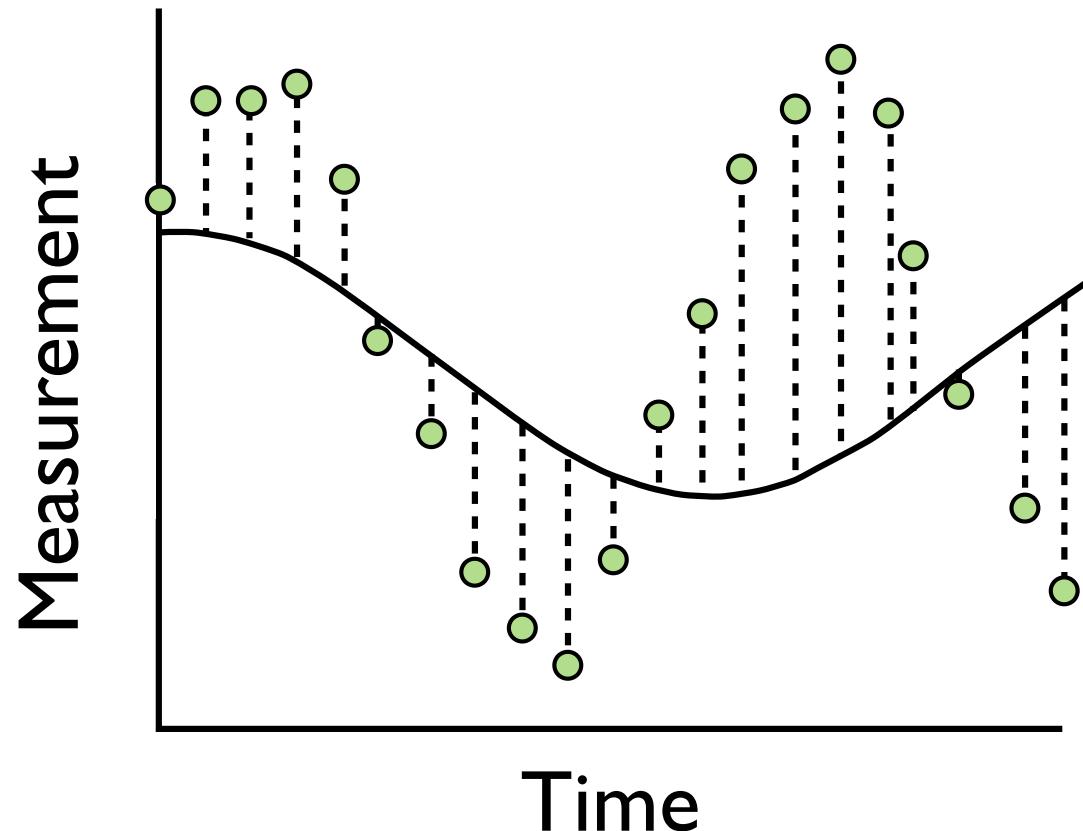
$$A=1.01 \quad \omega=0.592 \quad \Phi=1.541$$



e.g. consider we try fitting the following function with parameters:

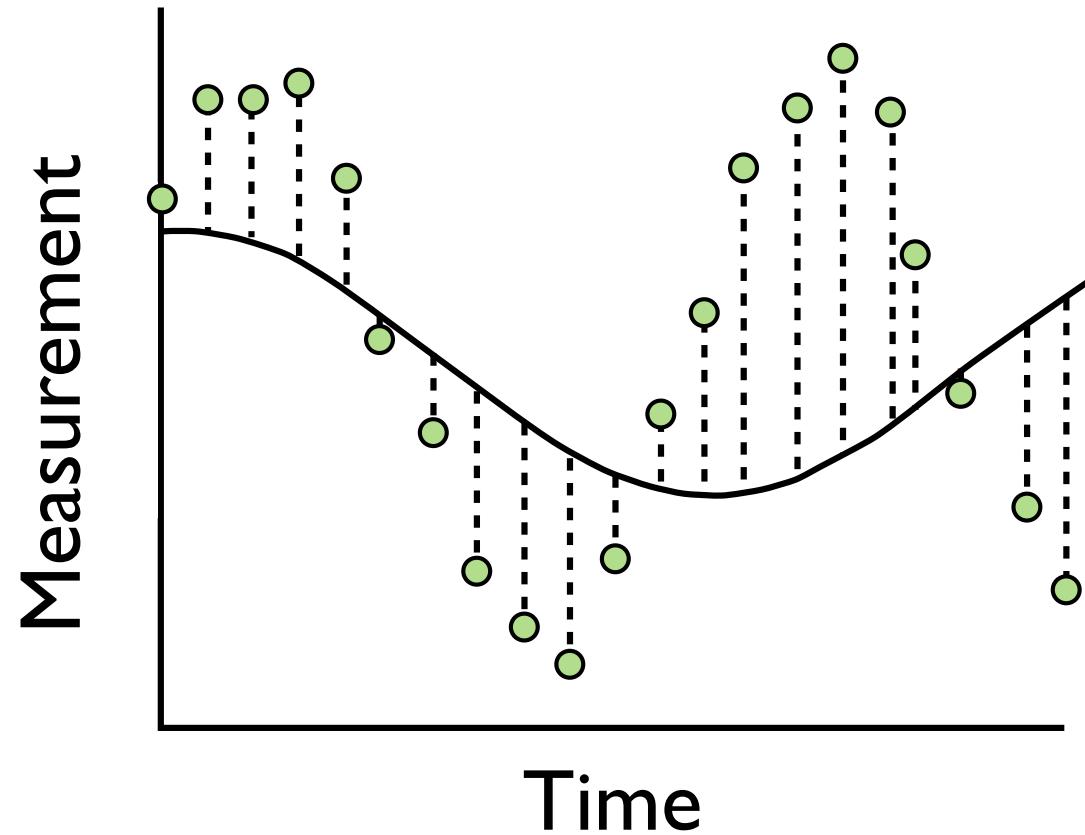
$$f(t) = A \sin(\omega t + \Phi)$$

$$A=1.01 \quad \omega=0.592 \quad \Phi=1.541$$



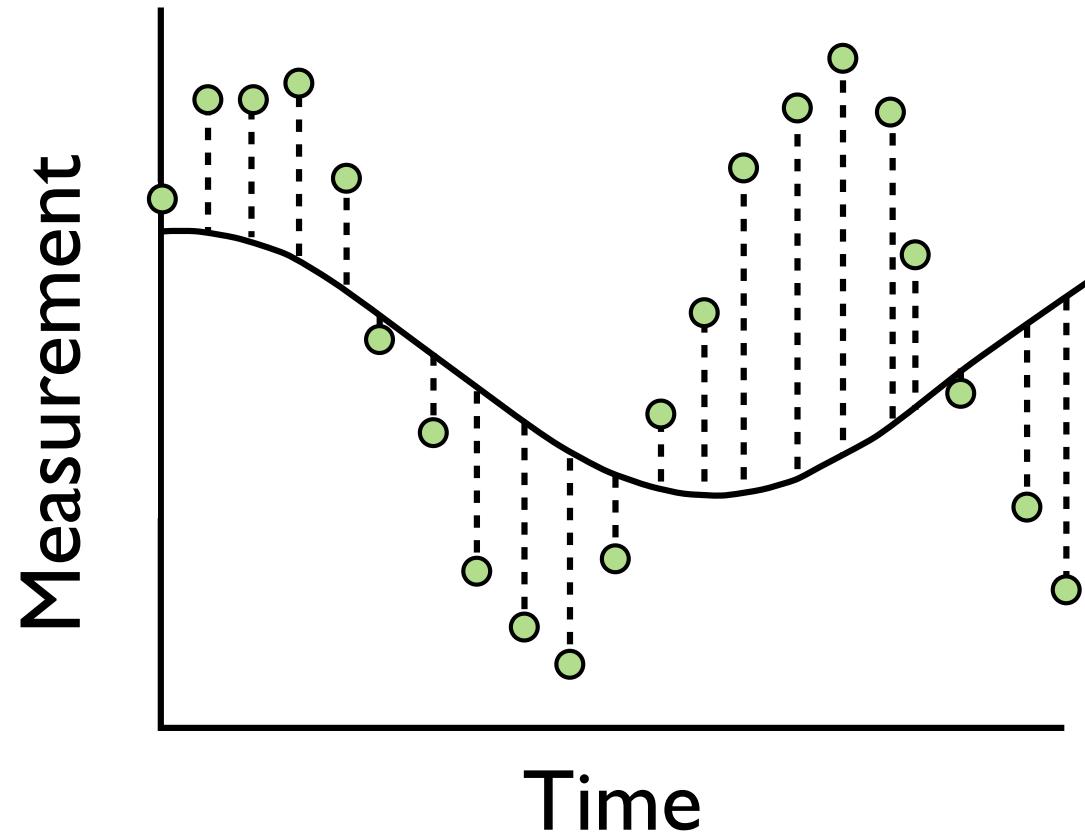
sum all “residual-squared” for all observation points

$$\sum_{j=1}^m (y_j - A \sin(\omega t_j + \Phi))^2$$



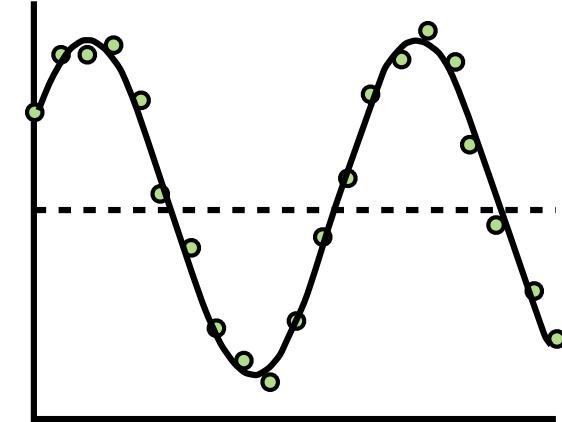
find parameters that **minimise** square residuals

minimise $\sum_{j=1}^m r_j(X)^2$...where $r_j(X) = y_j - A \sin(\omega t_j + \Phi)$
...and $X=(A, \omega, \Phi)$



find parameters that **minimise** square residuals

SUMMARY Part 5. curve fitting



- curve fitting “least squares”
- next time: Gauss-Newton, approximate Hessian

See the last slides for an (optional) homework exercise, get some practice with Hessians.

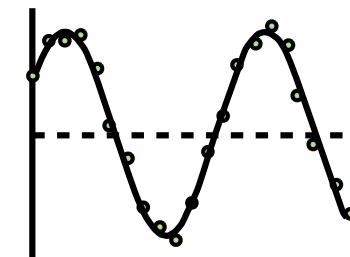
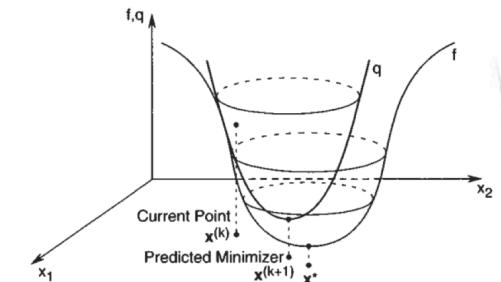
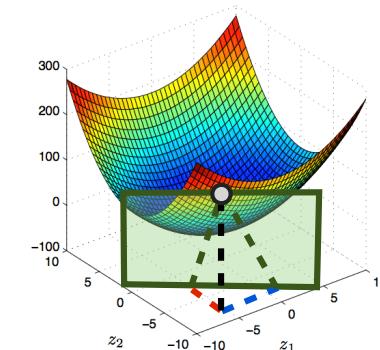
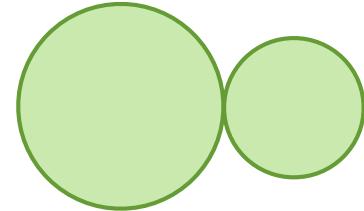
Together we did pseudo-code for 1D Newton root-finding.

You could try implementing in MATLAB/Python:

- 1D Newton root-finding
- nD Newton root-finding (Jakobian)
- 1D Newton optimisation
- nD Newton optimisation (Hessian)

SUMMARY

- preliminaries
- Newton from **1D to nD**
- *from root finding to optimization*
- *from Jakobian to Hessian*



exercise - some Newton optimization steps on this one:

$$\mathbf{X} = (x_1, x_2, x_3, x_4)$$

$$f(\mathbf{X}) = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4$$

exercise - some Newton optimization steps on this one:

$$X = (x_1, x_2, x_3, x_4)$$

$$f(X) = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4$$

STEP I. partial derivatives

$\nabla f(X) =$	partial derivative of f w.r.t x_1 ($\partial f / \partial x_1$)
	partial derivative of f w.r.t x_2 ($\partial f / \partial x_2$)
	partial derivative of f w.r.t x_3 ($\partial f / \partial x_3$)
	partial derivative of f w.r.t x_4 ($\partial f / \partial x_4$)

exercise - some Newton optimization steps on this one:

$$X = (x_1, x_2, x_3, x_4)$$

$$f(X) = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4$$

STEP I. partial derivatives

$2(x_1 + 10x_2) + 40(x_1 - x_4)^3$
partial derivative of f w.r.t x_2 ($\partial f / \partial x_2$)
partial derivative of f w.r.t x_3 ($\partial f / \partial x_3$)
partial derivative of f w.r.t x_4 ($\partial f / \partial x_4$)

exercise - some Newton optimization steps on this one:

$$\mathbf{X} = (x_1, x_2, x_3, x_4)$$

$$f(\mathbf{X}) = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4$$

STEP I. partial derivatives

$$\nabla f(\mathbf{X}) = \begin{matrix} 2(x_1 + 10x_2) + 40(x_1 - x_4)^3 \\ 20(x_1 + 10x_2) + 4(x_2 - 2x_3)^3 \\ 10(x_3 - x_4) - 8(x_2 - 2x_3)^3 \\ -10(x_3 - x_4) - 40(x_1 - x_4)^3 \end{matrix}$$

STEP 2. partial derivatives of each “partial derivative” (Hessian)

$\partial f / \partial x_1 =$	$2(x_1 + 10x_2) + 40(x_1 - x_4)^3$
$\partial f / \partial x_2 =$	$20(x_1 + 10x_2) + 4(x_2 - 2x_3)^3$
$\partial f / \partial x_3 =$	$10(x_3 - x_4) - 8(x_2 - 2x_3)^3$
$\partial f / \partial x_4 =$	$-10(x_3 - x_4) - 40(x_1 - x_4)^3$

	∂x_1	∂x_2	∂x_3	∂x_4
$\partial f / \partial x_1$?	?	?	?
$\partial f / \partial x_2$				
$\partial f / \partial x_3$				
$\partial f / \partial x_4$				

STEP 2. partial derivatives of each “partial derivative” (Hessian)

$\partial f / \partial x_1 =$	$2(x_1 + 10x_2) + 40(x_1 - x_4)^3$
$\partial f / \partial x_2 =$	$20(x_1 + 10x_2) + 4(x_2 - 2x_3)^3$
$\partial f / \partial x_3 =$	$10(x_3 - x_4) - 8(x_2 - 2x_3)^3$
$\partial f / \partial x_4 =$	$-10(x_3 - x_4) - 40(x_1 - x_4)^3$

	∂x_1	∂x_2	∂x_3	∂x_4
$\partial f / \partial x_1$	$2 + 120(x_1 - x_4)^2$	20	0	$-120(x_1 - x_4)^2$
$\partial f / \partial x_2$				
$\partial f / \partial x_3$				
$\partial f / \partial x_4$				

STEP 2. partial derivatives of each “partial derivative” (Hessian)

$\partial f / \partial x_1 =$	$2(x_1 + 10x_2) + 40(x_1 - x_4)^3$
$\partial f / \partial x_2 =$	$20(x_1 + 10x_2) + 4(x_2 - 2x_3)^3$
$\partial f / \partial x_3 =$	$10(x_3 - x_4) - 8(x_2 - 2x_3)^3$
$\partial f / \partial x_4 =$	$-10(x_3 - x_4) - 40(x_1 - x_4)^3$

	∂x_1	∂x_2	∂x_3	∂x_4
$\partial f / \partial x_1$	$2 + 120(x_1 - x_4)^2$	20	0	$-120(x_1 - x_4)^2$
$\partial f / \partial x_2$	20	$200 + 12(x_2 - 2x_3)^2$	$-24(x_2 - 2x_3)^2$	0
$\partial f / \partial x_3$	0	$-24(x_2 - 2x_3)^2$	$10 + 48(x_2 - 2x_3)^2$	-10
$\partial f / \partial x_4$	$-120(x_1 - x_4)^2$	0	-10	$10 + 120(x_1 - x_4)^2$

STEP 3. do three iterations (X_1, X_2, X_3)

$$X_0 = [3, -1, 0, 1]$$

$$X^{(i+1)} = X^{(i)} - H(X^{(i)})^{-1} \nabla f(X^{(i)})$$

$\partial f / \partial x_1 =$	$2(x_1 + 10x_2) + 40(x_1 - x_4)^3$
$\partial f / \partial x_2 =$	$20(x_1 + 10x_2) + 4(x_2 - 2x_3)^3$
$\partial f / \partial x_3 =$	$10(x_3 - x_4) - 8(x_2 - 2x_3)^3$
$\partial f / \partial x_4 =$	$-10(x_3 - x_4) - 40(x_1 - x_4)^3$

	∂x_1	∂x_2	∂x_3	∂x_4
$\partial f / \partial x_1$	$2 + 120(x_1 - x_4)^2$	20	0	$-120(x_1 - x_4)^2$
$\partial f / \partial x_2$	20	$200 + 12(x_2 - 2x_3)^2$	$-24(x_2 - 2x_3)^2$	0
$\partial f / \partial x_3$	0	$-24(x_2 - 2x_3)^2$	$10 + 48(x_2 - 2x_3)^2$	-10
$\partial f / \partial x_4$	$-120(x_1 - x_4)^2$	0	-10	$10 + 120(x_1 - x_4)^2$