<u>Theorem</u>. $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ convex. Then:

$$\boxed{\hat{x} \text{ is a global minimizer}} \iff \boxed{\nabla f(\hat{x}) = 0}$$

<u>Proof</u> $\boxed{\Rightarrow:}$ $\hat{x}$ global minimizer $\Rightarrow$ $\hat{x}$ local minimizer $\Rightarrow$ $\nabla f(\hat{x}) = 0$.

$\boxed{\Leftarrow:}$ Suppose now that $\nabla f(\hat{x}) = 0$.

Take any $x \in C$. Let $d := x - \hat{x}$ and $\varphi(t) = f(\hat{x} + td)$, $t \in \mathbb{R}$.

$$\varphi(1) = \varphi(0) + \varphi'(0) + \frac{1}{2}\varphi''(\theta) \quad \text{for some } \theta \in (0,1)$$

$$\begin{array}{ccccc} \| & & \| & & \\ f(x) & = & f(\hat{x}) & + \underbrace{\nabla f(\hat{x})}_{=0} (x - \hat{x}) + \frac{1}{2}(x-\hat{x})^T \underbrace{F(\hat{x} + \theta(x - \hat{x}))}_{\text{p.s.d}} (x - \hat{x}) \end{array}$$

$$f(x) \geq f(\hat{x})$$

Done! $\square$

Summary.

$$\begin{cases} \text{minimize} & f(x) \\ \text{s.t.} & x \in \mathbb{R}^n \end{cases}$$

Theorem (Necessary condition)

$$x_0 \text{ local minimizer} \Longrightarrow \begin{cases} \nabla f(x_0) = 0 \\ F(x_0) \text{ p.s.d.} \end{cases}$$

Theorem (Sufficient condition)

$$\left.\begin{matrix} \nabla f(x_0) = 0 \\ F(x_0) \text{ p.d.} \end{matrix}\right\} \Longrightarrow x_0 \text{ local minimizer.}$$

Theorem $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ convex.

$$\boxed{\hat{x} \text{ global minimizer}} \Longleftrightarrow \boxed{\nabla f(\hat{x}) = 0}$$

For a convex function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, $\boxed{\hat{x} \text{ is a global minimizer}} \Longleftrightarrow \boxed{\nabla f(\hat{x}) = 0}$

<u>Example</u>

$$f(x) = e^{x} + e^{-x} + \sin x$$

$$f'(x) = e^{x} - e^{-x} + \cos x$$

$$f''(x) = e^{x} + e^{-x} - \sin x$$

$$= e^{x} + e^{-x} - 2\, e^{x/2} \cdot e^{-x/2} + 2 - \sin x$$

$$= \underbrace{\left(e^{x/2} - e^{-x/2}\right)^2}_{\geq 0} + \underbrace{(2 - \sin x)}_{\geq 0} \geq 0 .$$

So f is convex

Know: $\hat{x}$ is a global minimizer $\Longleftrightarrow$ $f'(\hat{x}) = 0$

$\Longleftrightarrow$ $e^{\hat{x}} - e^{-\hat{x}} + \cos \hat{x} = 0$

$\boxed{\hat{x}\ ?}$

$\boxed{\text{Can one find } \hat{x} \text{ numerically?}}$

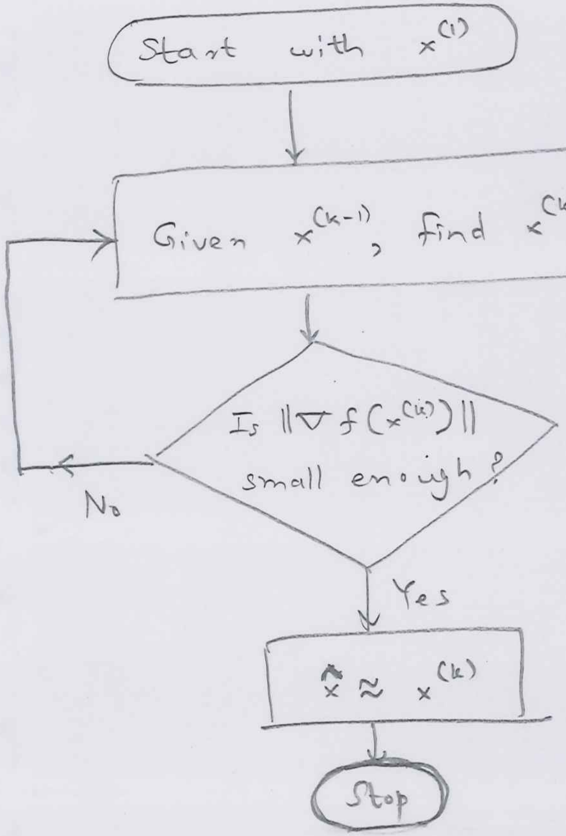
How to find $\hat{x}$ numerically?

<u>Newton's method</u>

$$\boxed{\begin{array}{l} f : \mathbb{R}^n \longrightarrow \mathbb{R} \\ F(x) \text{ p.d. } \forall x \in \mathbb{R}^n \end{array}}$$

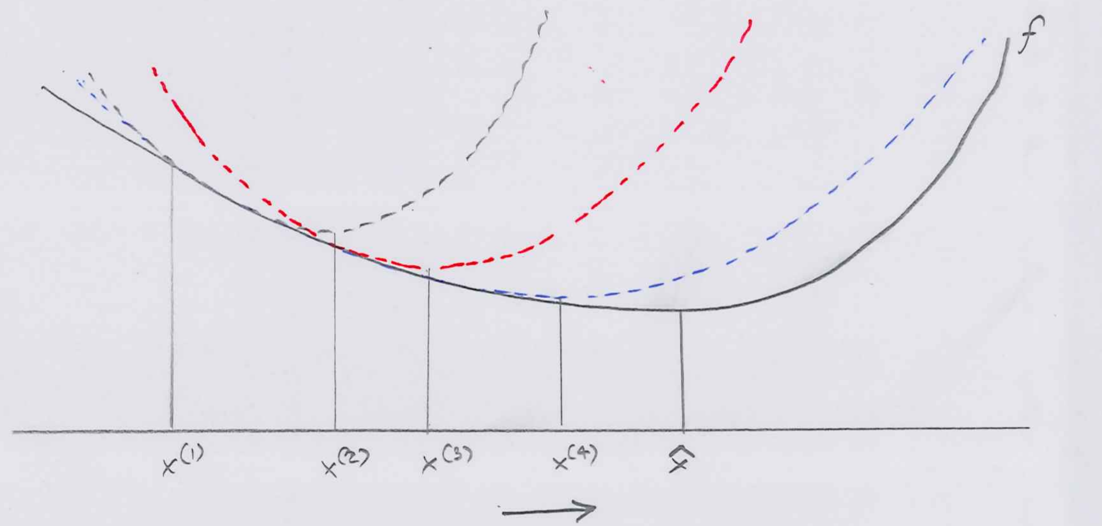
$\Rightarrow f$ is convex.

$\hat{x}$ global minimizer $\Leftrightarrow \nabla f(\hat{x}) = 0$.



?

<u>Idea</u>: Approximate f by a quadratic function q at $x^{(k)}$ and minimize q to find $x^{(k+1)}$



11.4

At a point $x^{(k)}$, look at $f(x^{(k)})$, $\nabla f(x^{(k)})$, $F(x^{(k)})$

Construct a quadratic function $q : \mathbb{R}^n \longrightarrow \mathbb{R}$ s.t.

$$q(x^{(k)}) = f(x^{(k)})$$
$$\nabla q(x^{(k)}) = \nabla f(x^{(k)})$$
$$Q(x^{(k)}) = F(x^{(k)})$$

Take $x^{(k+1)}$ as the minimizer of $q(x)$ $(x \in \mathbb{R}^n)$.

Construction of q:

$$q(x) = \frac{1}{2} x^T H x + c^T x + c_0 .$$
$$\nabla q(x) = (Hx + c)^T$$
$$Q(x) = H .$$

Thus:

$$H = Q(x^{(k)}) = F(x^{(k)})$$
$$Hx^{(k)} + c = (\nabla f(x^{(k)}))^T \Rightarrow c = (\nabla f(x^{(k)}))^T - F(x^{(k)})\, x^{(k)} .$$
$$q(x^{(k)}) = f(x^{(k)}) \Rightarrow c_0 = f(x^{(k)}) - c^T x^{(k)} - \frac{1}{2}(x^{(k)})^T F(x^{(k)}) (x^{(k)})$$

Consider $\begin{cases} \text{minimize} & \frac{1}{2} x^T H x + c^T x + c_0 \\ \text{s.t.} & x \in \mathbb{R}^n \end{cases}$

$H = F(x^{(k)})$ is p.d.

So $\exists!$ minimizer, given by

$$\begin{aligned} x^{(k+1)} &:= -H^{-1} c \\ &= -\left(F(x^{(k)})\right)^{-1} \left(\left(\nabla f(x^{(k)})\right)^T - F(x^{(k)})\, x^{(k)} \right) \\ &= -\left(F(x^{(k)})\right)^{-1} \left(\nabla f(x^{(k)})\right)^T + x^{(k)} \end{aligned}$$

i.e., $$\boxed{x^{(k+1)} - x^{(k)} = -\left(F(x^{(k)})\right)^{-1} \left(\nabla f(x^{(k)})\right)^T.}$$

Update equation

In the one-variable case:

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

Example (revisited)

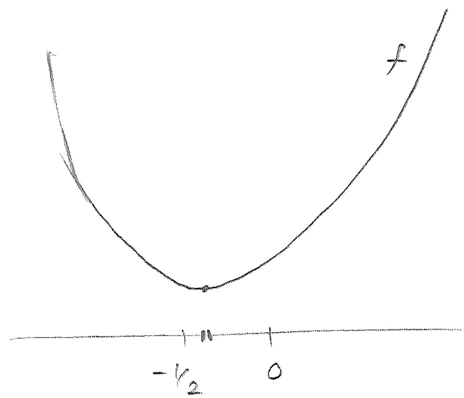
$$f(x) = e^{x} + e^{-x} + \sin x$$

$$f'(x) = e^{x} - e^{-x} + \cos x$$

$$f''(x) = e^{x} + e^{-x} - \sin x$$

$$= \underbrace{\left(e^{x/2} - e^{-x/2}\right)^2}_{\geq 0} + \underbrace{(2 - \sin x)}_{> 0} > 0 .$$

$f''(x)$ is p.d.



$$x^{(k+1)} = x^{(k)} - \frac{e^{x^{(k)}} - e^{-x^{(k)}} + \cos x^{(k)}}{e^{x^{(k)}} + e^{-x^{(k)}} - \sin x^{(k)}} .$$

$$x^{(1)} = 0$$

$$x^{(2)} = 0 - \frac{1 - 1 + 1}{1 + 1 - 0} = -\frac{1}{2} .$$

$$x^{(3)} = -\frac{1}{2} - \frac{e^{-1/2} - e^{1/2} + \cos\left(-\frac{1}{2}\right)}{e^{-1/2} + e^{+1/2} - \sin\left(-\frac{1}{2}\right)} = -0.4398$$

$$x^{(4)} = \ldots \approx -0.4385$$

$$x^{(5)} = \ldots \approx -0.4385$$

$$f'\left(x^{(5)}\right) \approx 1.2851 \times 10^{-5} .$$

<u>Example</u>. (What happens if f is quadratic to begin with?)

$$f(x) = \frac{1}{2} x^T H x + c^T x + c_0 .$$

H p.d.

Take any vector v as $x^{(1)}$.

$$\nabla f(x) = (Hx + c)^T \quad \text{and so} \quad \nabla f(x^{(1)}) = \nabla f(v) = (Hv + c)^T$$

$$F(x) = H \quad \text{and so} \quad F(x^{(1)}) = H.$$

$$x^{(k+1)} = x^{(k)} - \left(F(x^{(k)})\right)^{-1} \left(\nabla f(x^{(k)})\right)^T, \quad \text{and so}$$

$$\begin{aligned} x^{(2)} &= v - H^{-1}(Hv + c) \\ &= v - H^{-1}Hv - H^{-1}c \\ &= \cancel{v} - \cancel{v} - H^{-1}c \\ &= -H^{-1}c . \end{aligned}$$

<u>Recall</u>. Unique optimal solution is $\hat{x} = -H^{-1}c$.

So Newton's algorithm converges in just one step!

But this is expected.

Nonlinear least squares problem.

Earlier: (LS): $\begin{cases} \text{minimize} & \frac{1}{2}\|Ax-b\|^2 \\ \text{s.t.} & x \in \mathbb{R}^n \end{cases}$ $= \frac{1}{2}\sum_{i=1}^{m}(a_{i1}x_1 + \dots + a_{in}x_n - b_i)^2$

Now: (NLS): $\begin{cases} \text{minimize} & \frac{1}{2}\left((h_1(x))^2 + \dots + (h_m(x))^2\right) = \frac{1}{2}\sum_{i=1}^{m}(h_i(x))^2 . \\ \text{s.t.} & x \in \mathbb{R}^n . \end{cases}$

Why consider this? Suppose we want to solve: $\begin{Bmatrix} h_1(x) = 0 \\ \vdots \\ h_m(x) = 0 \end{Bmatrix}$ Nonlinear system (analogous to $Ax = b$).

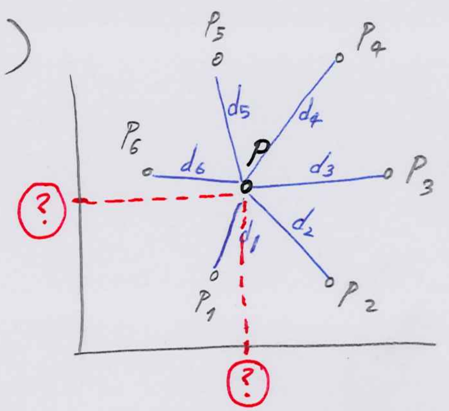
Then try finding an x s.t. the "error"

$$(h_1(x) - 0)^2 + \dots + (h_m(x) - 0)^2$$

is minimized

(Analogous to minimizing $\|Ax-b\|^2$)

Example (Finding coordinates)



m reference points $P_1, \ldots, P_m$ with known coordinates $(x_1, y_1), \ldots, (x_m, y_m)$, respectively.

Problem: Find coordinates of P, given estimates of its distances $d_1, \ldots, d_m$ to the reference points.

Let (x, y) be the coordinates of P, which we seek.

Define
$$h_1(x,y) := \sqrt{(x-x_1)^2 + (y-y_1)^2} - d_1$$
$$\vdots$$
$$h_m(x,y) := \sqrt{(x-x_m)^2 + (y-y_m)^2} - d_m.$$

Ideally we want (x,y) s.t. $\begin{cases} h_1(x,y) = 0 \\ \vdots \\ h_m(x,y) = 0 \end{cases}$.

But we settle for (x,y) which minimizes the error $(h_1(x,y)-0)^2 + \ldots + (h_m(x,y)-0)^2$.

So we consider:
$$\left\{ \begin{array}{ll} \text{minimize} & (h_1(x,y))^2 + \ldots + (h_m(x,y))^2 \\ \text{s.t.} & (x,y) \in \mathbb{R}^2 . \end{array} \right\} \leftarrow \text{Nonlinear least squares problem.}$$

$$\begin{cases} \text{minimize} & \frac{1}{2}\sum_{i=1}^{m} (h_i(x))^2 \\ \text{s.t.} & x \in \mathbb{R}^n . \end{cases}$$

We are interested again in a numerical method for finding an approximate minimizer.

In principle, one could try using Newton's method with $f(x) := \frac{1}{2}\sum_{i=1}^{m} (h_i(x))^2$, but we will learn a simpler method, which uses the special structure of the problem.

This method is called the <u>Gauss-Newton method.</u>

$x^{(k)}$

$\downarrow$

Approximate NLS by a LS problem

$\downarrow$

Take $x^{(k+1)}$ as the minimizer

$\downarrow$

Is $\frac{1}{2}\sum_{i=1}^{m} (h_i(x^{(k+1)}))^2$ small enough $\xrightarrow{\text{Yes}}$ Stop

No

Notation: $h(x) = \begin{bmatrix} h_1(x) \\ \vdots \\ h_m(x) \end{bmatrix}$ $\qquad \nabla h(x) = \begin{bmatrix} \nabla h_1(x) \\ \vdots \\ \nabla h_m(x) \end{bmatrix} \in \mathbb{R}^{m \times n}$.

Approximation of NLS by LS at $x^{(k)}$:

$$h_1(x) \approx h_1(x^{(k)}) + \nabla h_1(x^{(k)})(x - x^{(k)})$$ ← affine linear function of x

$$\vdots$$

$$h_m(x) \approx h_m(x^{(k)}) + \nabla h_m(x^{(k)})(x - x^{(k)})$$

$$\begin{bmatrix} h_1(x) \\ \vdots \\ h_m(x) \end{bmatrix} \approx \begin{bmatrix} h_1(x^{(k)}) \\ \vdots \\ h_m(x^{(k)}) \end{bmatrix} + \begin{bmatrix} \nabla h_1(x^{(k)}) \\ \vdots \\ \nabla h_m(x^{(k)}) \end{bmatrix} (x - x^{(k)})$$

i.e.,

$$h(x) \approx \underbrace{h(x^{(k)}) + \nabla h(x^{(k)})(x - x^{(k)})}_{Ax - b}\,,$$

$Ax - b$, where $A := \nabla h(x^{(k)})$

$b := \nabla h(x^{(k)})\, x^{(k)} - h(x^{(k)})$.

$$(\text{NLS}): \begin{cases} \text{minimize} & \frac{1}{2}\sum_{i=1}^{m} (h_i(x))^2 = \frac{1}{2}\|h(x)\|^2 \\ \text{s.t.} & x \in \mathbb{R}^n. \end{cases}$$

$$\Downarrow$$

$$(\text{LS}): \begin{cases} \text{minimize} & \frac{1}{2}\|h(x^{(k)}) + \nabla h(x^{(k)})(x - x^{(k)})\|^2 = \frac{1}{2}\|Ax - b\|^2 \\ \text{s.t.} & x \in \mathbb{R}^n \end{cases}$$

where $A := \nabla h(x^{(k)})$

$b := \nabla h(x^{(k)})\, x^{(k)} - h(x^{(k)})$

Take $x^{(k+1)}$ as a minimizer of (LS).

Minimizer $x^{(k+1)}$ is given by a solution to the normal equation associated with (LS):

$$A^T A\, x^{(k+1)} = A^T b,$$

i.e.,

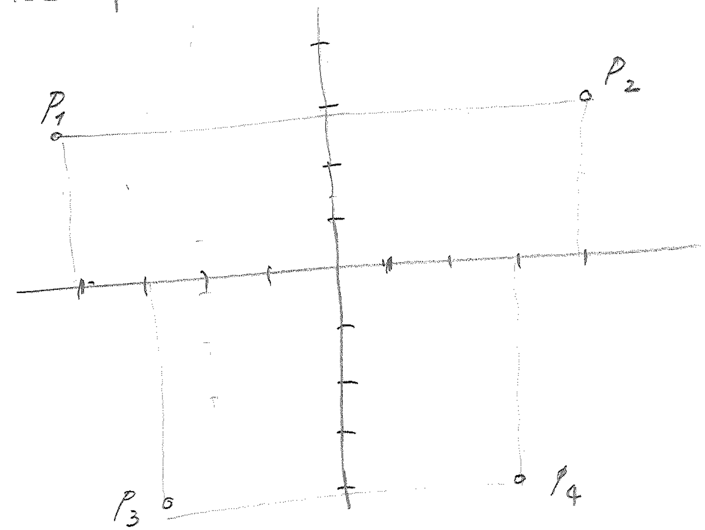
$$(\nabla h(x^{(k)}))^T \nabla h(x^{(k)})\, x^{(k+1)} = (\nabla h(x^{(k)}))^T (\nabla h(x^{(k)})\, x^{(k)} - h(x^{(k)})).$$

<u>Update equation:</u> $(\nabla h(x^{(k)}))^T \nabla h(x^{(k)})\, (x^{(k+1)} - x^{(k)}) = -(\nabla h(x^{(k)}))^T h(x^{(k)})$

Note that this involves only first order derivatives (simpler than Newton's method).

Example (Revisited) Suppose there are $m = 4$ reference points:

$$P_1 \equiv (-40, 30)$$
$$P_2 \equiv (40, 30)$$
$$P_3 \equiv (-30, -40)$$
$$P_4 \equiv (30, -40)$$



We want to find the coordinates of P whose estimated distances to P_1, P_2, P_3, P_4 are $51, 52, 48, 49$, respectively.

Define

$$h_1(x,y) = \sqrt{(x+40)^2 + (y-30)^2} - 51$$
$$h_2(x,y) = \sqrt{(x-40)^2 + (y-30)^2} - 52$$
$$h_3(x,y) = \sqrt{(x+30)^2 + (y+40)^2} - 48$$
$$h_4(x,y) = \sqrt{(x-30)^2 + (y+40)^2} - 49.$$

Start with $x^{(1)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and find $x^{(2)}, x^{(3)}, \ldots$ using

$$\left(\nabla h(x^{(k)})\right)^T \nabla h(x^{(k)}) \left(x^{(k+1)} - x^{(k)}\right) = -\left(\nabla h(x^{(k)})\right)^T h(x^{(k)}).$$

11.14

	$x^{(k)}$	$h(x^{(k)})$	$\lVert h(x^{(k)}) \rVert$
$k=1$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} -1 \\ -2 \\ 2 \\ 1 \end{bmatrix}$	10
$k=2$	$\begin{bmatrix} -0.7 \\ -2.1 \end{bmatrix}$	$\begin{bmatrix} -0.2565 \\ -0.1647 \\ -0.0949 \\ -0.2260 \end{bmatrix}$	0.1530
$k=3$	$\begin{bmatrix} -0.7075 \\ -2.1068 \end{bmatrix}$	$\begin{bmatrix} -0.2581 \\ -0.1545 \\ -0.1049 \\ -0.2266 \end{bmatrix}$	0.1528

<u>Example</u> (revisited). $f(x) = e^x + e^{-x} + \sin x$.

$$f'(x) = e^x - e^{-x} + \cos x$$

Want an approximate solution to $h(x) = 0$, where

$$h(x) := e^x - e^{-x} + \cos x.$$

So consider $\begin{cases} \text{minimize} & \frac{1}{2}(h(x))^2 \\ \text{s.t.} & x \in \mathbb{R}. \end{cases}$

Update equation:

$$\left(h'(x^{(k)})\right)^2 \left(x^{(k+1)} - x^{(k)}\right) = - h'(x^{(k)})\, h(x^{(k)})$$

i.e., $$x^{(k+1)} = x^{(k)} - \frac{h(x^{(k)})}{h'(x^{(k)})}.$$

$$= x^{(k)} - \frac{\left(e^{x^{(k)}} - e^{-x^{(k)}} + \cos x^{(k)}\right)}{\left(e^{x^{(k)}} + e^{-x^{(k)}} - \sin x^{(k)}\right)},$$

same as in Newton's method before!

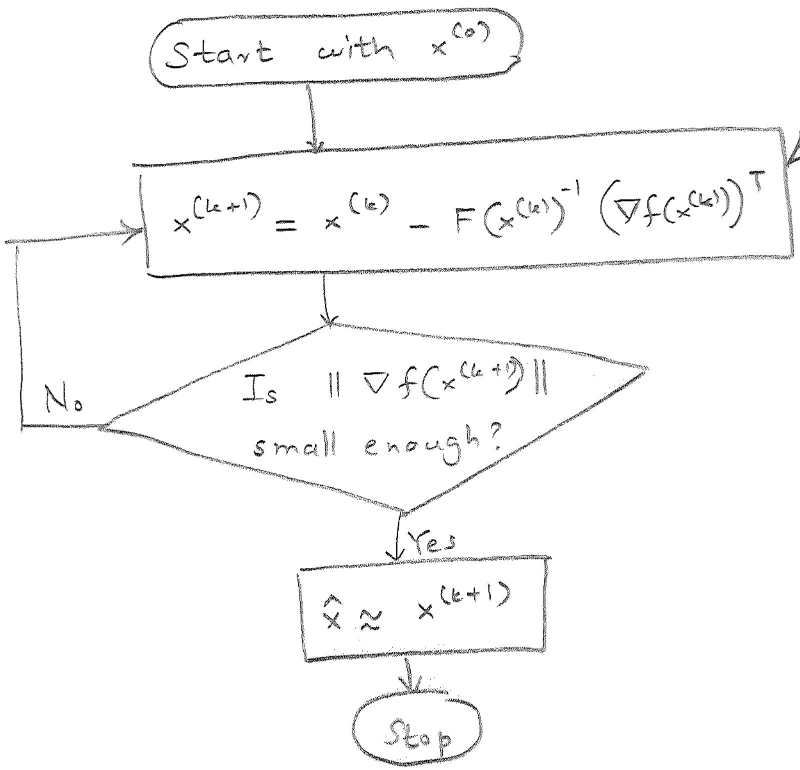
Numerical methods.

(I) Newton's method

$f: \mathbb{R}^n \longrightarrow \mathbb{R}$ s.t. $F(x)$ p.d. $\forall x \in \mathbb{R}^n$

For finding an approximate minimizer of $\begin{cases} \min. f(x) \\ \text{s.t. } x \in \mathbb{R}^n \end{cases}$

i.e, an approximate solution to $\nabla f(\hat{x}) = 0$:

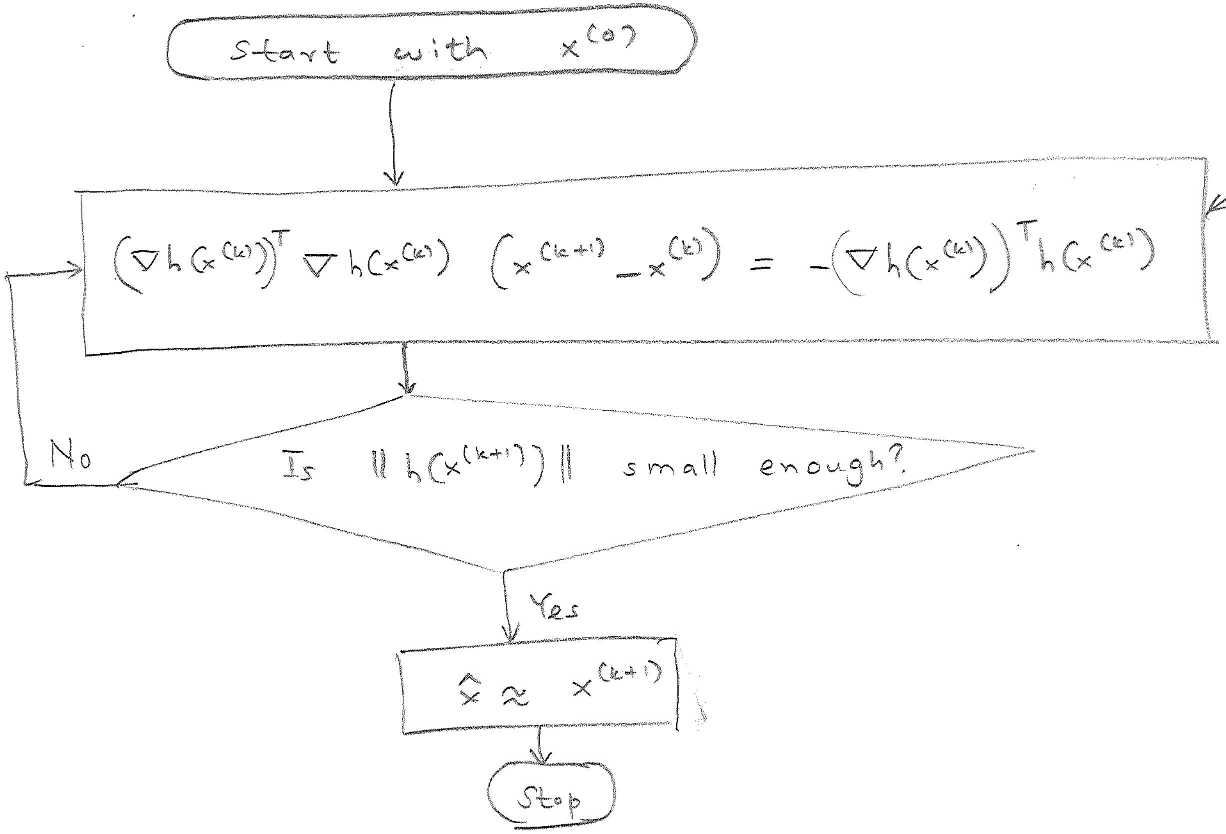


Idea behind this update eqn.:

Approximate f by a quadratic function q at $x^{(k)}$, and minimize q to find $x^{(k+1)}$

II **Gauss-Newton method** for finding an approximate solution to the **nonlinear least squares problem**

$$\begin{cases} \min & \frac{1}{2}\|h(x)\|^2 \\ \text{s.t.} & x \in \mathbb{R}^n \end{cases}, \text{ where } h(x) := \begin{bmatrix} h_1(x) \\ \vdots \\ h_m(x) \end{bmatrix}.$$



Idea behind this update eqn.:

Approximate h by its first order approximation at $x^{(k)}$, and take $x^{(k+1)}$ as a minimizer of the least squares problem hence obtained.