V. Linear & Nonlinear Least-Squares

V.0. Examples, linear/nonlinear least-squares

In practice, one has often to determine unknown parameters of a given function (from natural laws or model assumptions) through a series of measurements.

Usually the number of measurements *m* is much bigger than the number of parameters *n, i.e.*

Measurement

Time

Quantity

i	1	2	3	4	5	6	7	8
t	0.10	0.23	0.36	0.49	0.61	0.74	0.87	1.00
q	0.84	0.30	0.69	0.45	0.31	0.09	-0.17	0.12

$m = 8$

Model:

1 $q(t) = a_1 t + a_2$

2 $q(t) = a_2 e^{a_1 t}$

Goal: Find a_1, a_2

V.0. Examples, linear/nonlinear least-squares

In practice, one has often to determine unknown parameters of a given function (from natural laws or model assumptions) through a series of measurements.

Usually the number of measurements *m* is much bigger than the number of parameters *n, i.e.* $m \gg n$

Measurement

Time

Quantity

i	1	2	3	4	5	6	7	8
t	0.10	0.23	0.36	0.49	0.61	0.74	0.87	1.00
q	0.84	0.30	0.69	0.45	0.31	0.09	-0.17	0.12

$m = 8$

Model:

1 $q(t) = a_1 t + a_2$

2 $q(t) = a_2 e^{a_1 t}$

Goal: Find a_1, a_2 $\quad n = 2$

Problem: more equations than unknowns! … overdetermined

V.0. Examples, linear/nonlinear least-squares

Data:

i	1	2	3	4	5	6	7	8
t	0.10	0.23	0.36	0.49	0.61	0.74	0.87	1.00
q	0.84	0.30	0.69	0.45	0.31	0.09	-0.17	0.12

Model: 1 $q(t) = a_1 t + a_2$ Goal: Find a_1, a_2

$$
\left.\begin{aligned}
q_1 &= a_1 t_1 + a_2 \\
q_2 &= a_1 t_2 + a_2 \\
q_3 &= a_1 t_3 + a_2 \\
q_4 &= a_1 t_4 + a_2 \\
q_5 &= a_1 t_5 + a_2 \\
q_6 &= a_1 t_6 + a_2 \\
q_7 &= a_1 t_7 + a_2 \\
q_8 &= a_1 t_8 + a_2
\end{aligned}\right\} \Rightarrow
$$

Linear in model parameters!

$$
\begin{pmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \\ q_5 \\ q_6 \\ q_7 \\ q_8 \end{pmatrix} = \begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ t_3 & 1 \\ t_4 & 1 \\ t_5 & 1 \\ t_6 & 1 \\ t_7 & 1 \\ t_8 & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}
$$

V.0. Examples, linear/nonlinear least-squares

Data:

i	1	2	3	4	5	6	7	8
t	0.10	0.23	0.36	0.49	0.61	0.74	0.87	1.00
q	0.84	0.30	0.69	0.45	0.31	0.09	-0.17	0.12

Model: 1 $q(t) = a_1 t + a_2$ Goal: Find a_1, a_2

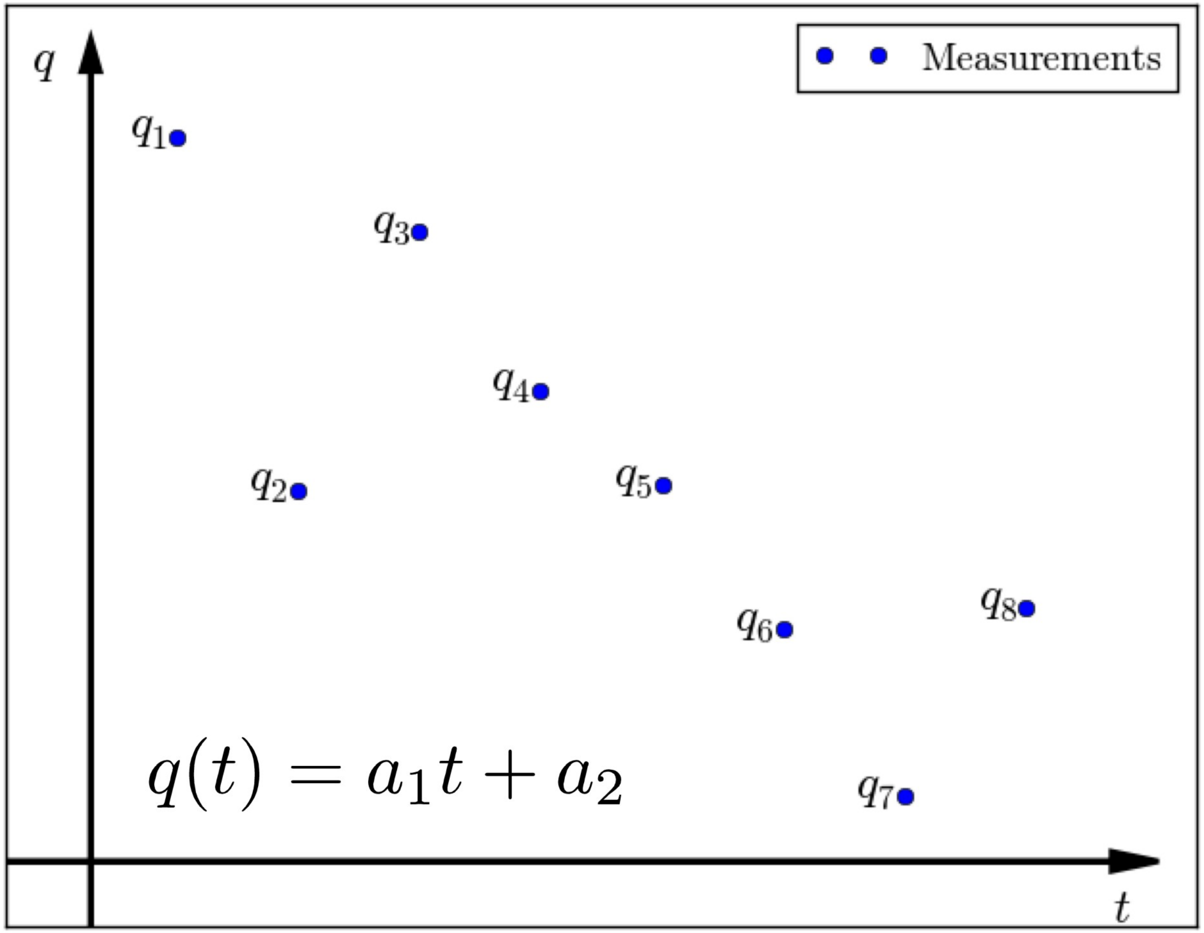
Linear in model parameters!

$$\left.\begin{aligned} q_1 &= a_1 t_1 + a_2 \\ q_2 &= a_1 t_2 + a_2 \\ q_3 &= a_1 t_3 + a_2 \\ q_4 &= a_1 t_4 + a_2 \\ q_5 &= a_1 t_5 + a_2 \\ q_6 &= a_1 t_6 + a_2 \\ q_7 &= a_1 t_7 + a_2 \\ q_8 &= a_1 t_8 + a_2 \end{aligned}\right\} \Rightarrow \underbrace{\begin{pmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \\ q_5 \\ q_6 \\ q_7 \\ q_8 \end{pmatrix}}_{\mathbf{b}\,,\ \mathbf{y}} = \underbrace{\begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ t_3 & 1 \\ t_4 & 1 \\ t_5 & 1 \\ t_6 & 1 \\ t_7 & 1 \\ t_8 & 1 \end{pmatrix}}_{A\,,\ X} \underbrace{\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}}_{\mathbf{x}\,,\ \boldsymbol{\beta}}$$

JUST NOTATION!

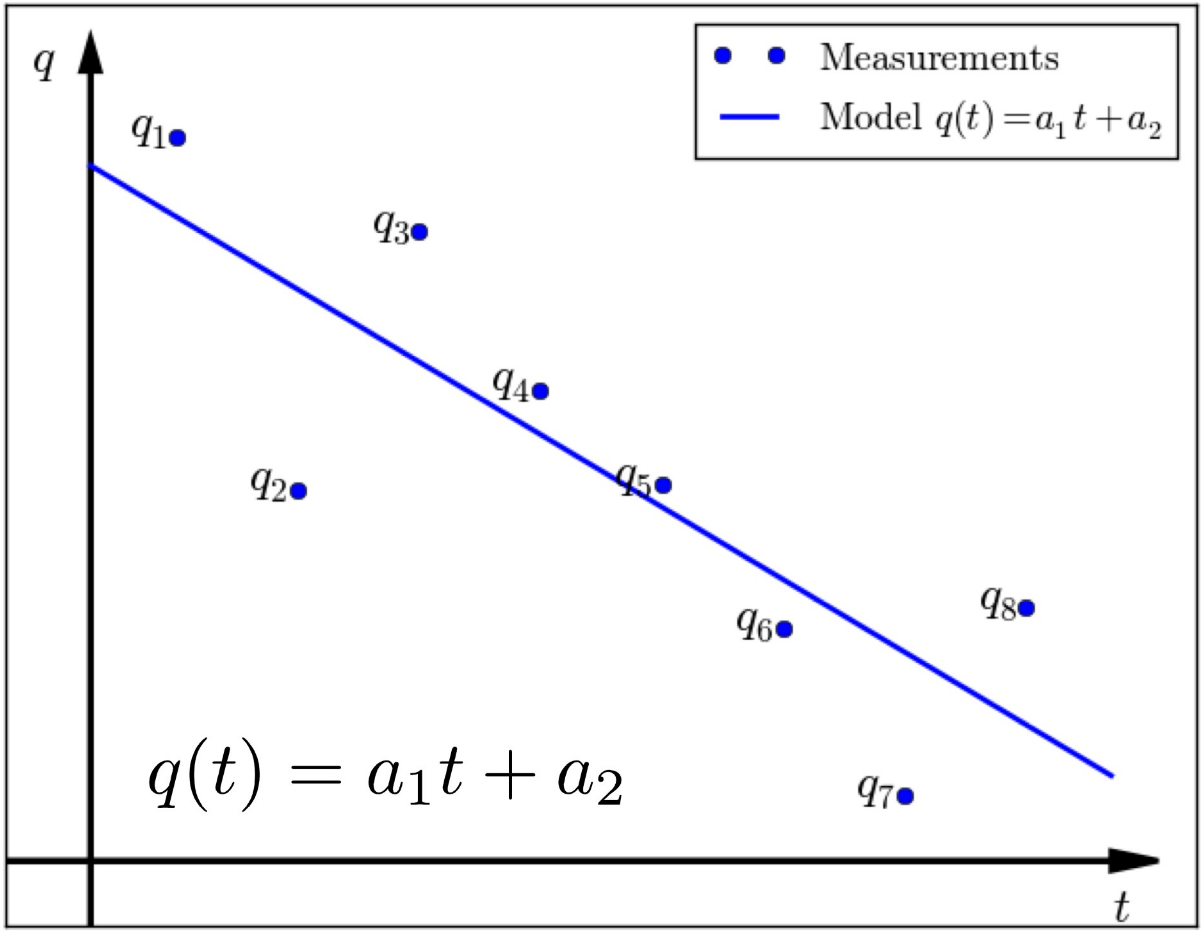
V.0. Examples, linear/nonlinear least-squares

- **Idea:** choose the parameters such that the distance between the data and the curve is minimal, i.e. the curve that fits best.



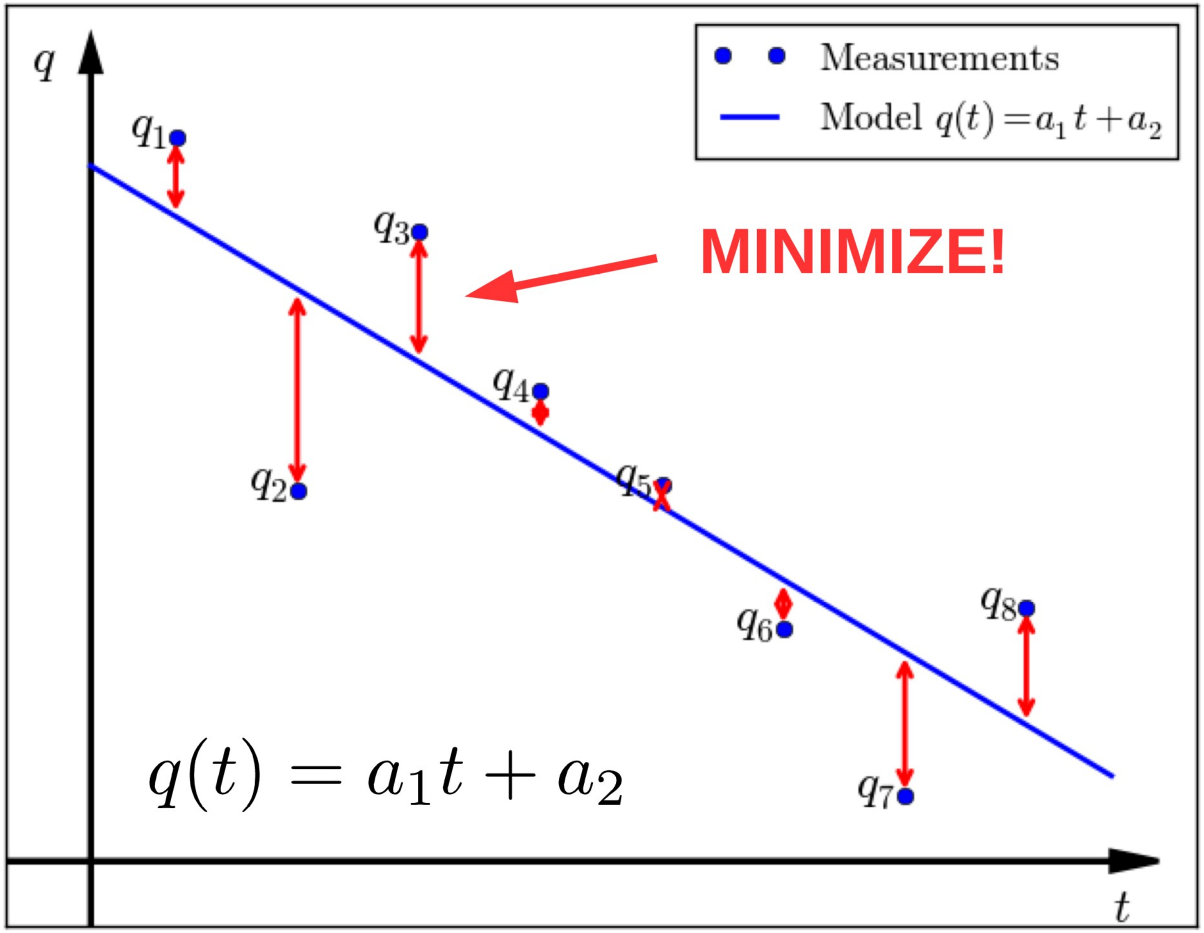
V.0. Examples, linear/nonlinear least-squares

- **Idea:** choose the parameters such that the distance between the data and the curve is minimal, i.e. the curve that fits best.



V.0. Examples, linear/nonlinear least-squares

- **Idea:** choose the parameters such that the distance between the data and the curve is minimal, i.e. the curve that fits best.



V.0. Examples, linear/nonlinear least-squares

- **Idea:** choose the parameters such that the distance between the data and the curve is minimal, i.e. the curve that fits best.
- Least squares solution:

$$\min_{\mathbf{x} \in D} \|A\mathbf{x} - \mathbf{b}\|_2$$

Domain of valid parameters (from application!)

Euclidean distance, a.k.a. 2-norm

$$\|\mathbf{c}\|_2 = \sqrt{\mathbf{c}^T \mathbf{c}} = \sqrt{\sum_{i=1}^{n} c_i^2}$$

V.0. Examples, linear/nonlinear least-squares

Data:

i	1	2	3	4	5	6	7	8
t	0.10	0.23	0.36	0.49	0.61	0.74	0.87	1.00
q	0.84	0.30	0.69	0.45	0.31	0.09	-0.17	0.12

Model: 2 $q(t) = a_2 e^{a_1 t}$

Goal: Find a_1, a_2

$$\left.\begin{aligned}
q_1 &= a_1 e^{a_2 t_1} \\
q_2 &= a_1 e^{a_2 t_2} \\
q_3 &= a_1 e^{a_2 t_3} \\
q_4 &= a_1 e^{a_2 t_4} \\
q_5 &= a_1 e^{a_2 t_5} \\
q_6 &= a_1 e^{a_2 t_6} \\
q_7 &= a_1 e^{a_2 t_7} \\
q_8 &= a_1 e^{a_2 t_8}
\end{aligned}\right\}$$

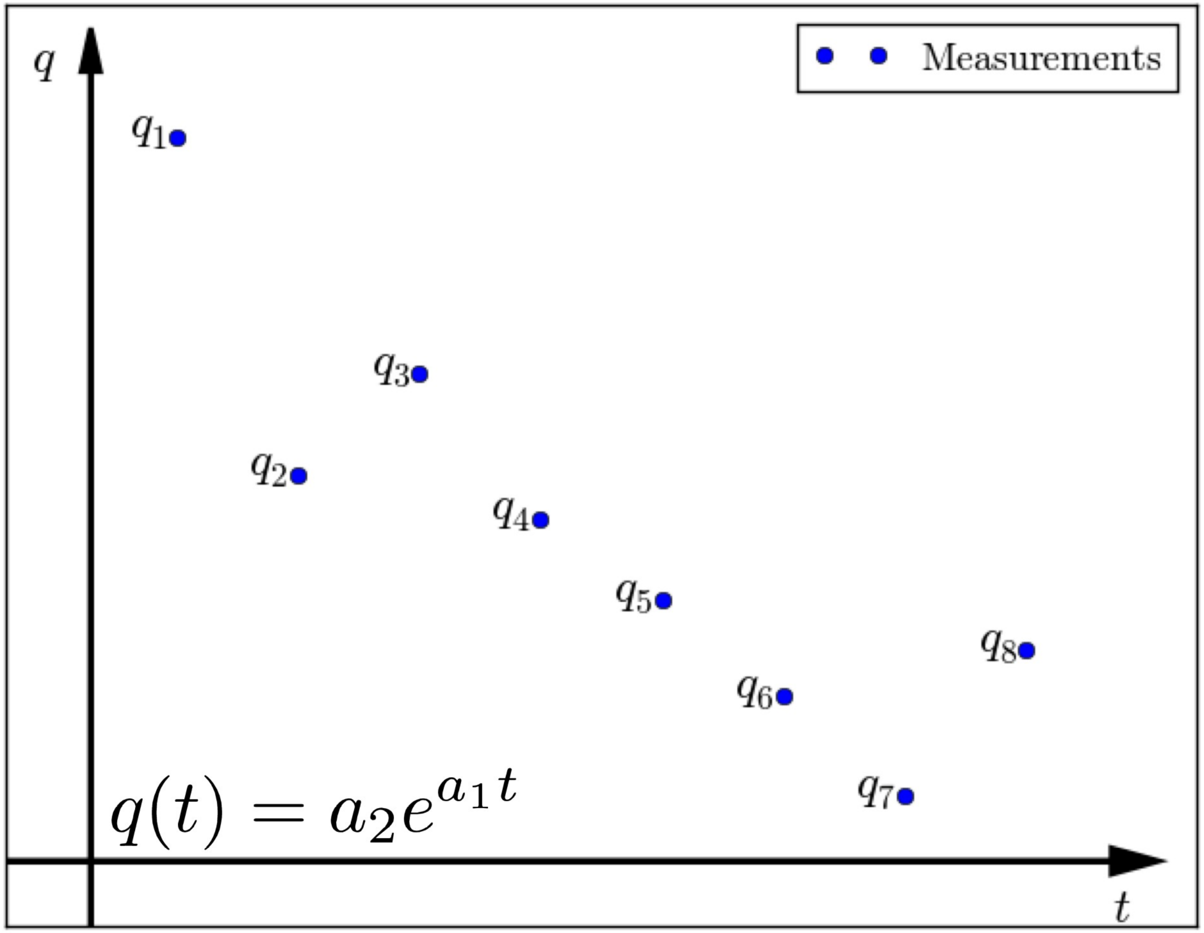
Nonlinear in model parameters!

$$\underbrace{\begin{pmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \\ q_5 \\ q_6 \\ q_7 \\ q_8 \end{pmatrix}}_{\mathbf{b}\,,\ \mathbf{y}} = \underbrace{\begin{pmatrix} a_1 e^{a_2 t_1} \\ a_1 e^{a_2 t_2} \\ a_1 e^{a_2 t_3} \\ a_1 e^{a_2 t_4} \\ a_1 e^{a_2 t_5} \\ a_1 e^{a_2 t_6} \\ a_1 e^{a_2 t_7} \\ a_1 e^{a_2 t_8} \end{pmatrix}}_{\mathbf{f}}$$

JUST NOTATION!

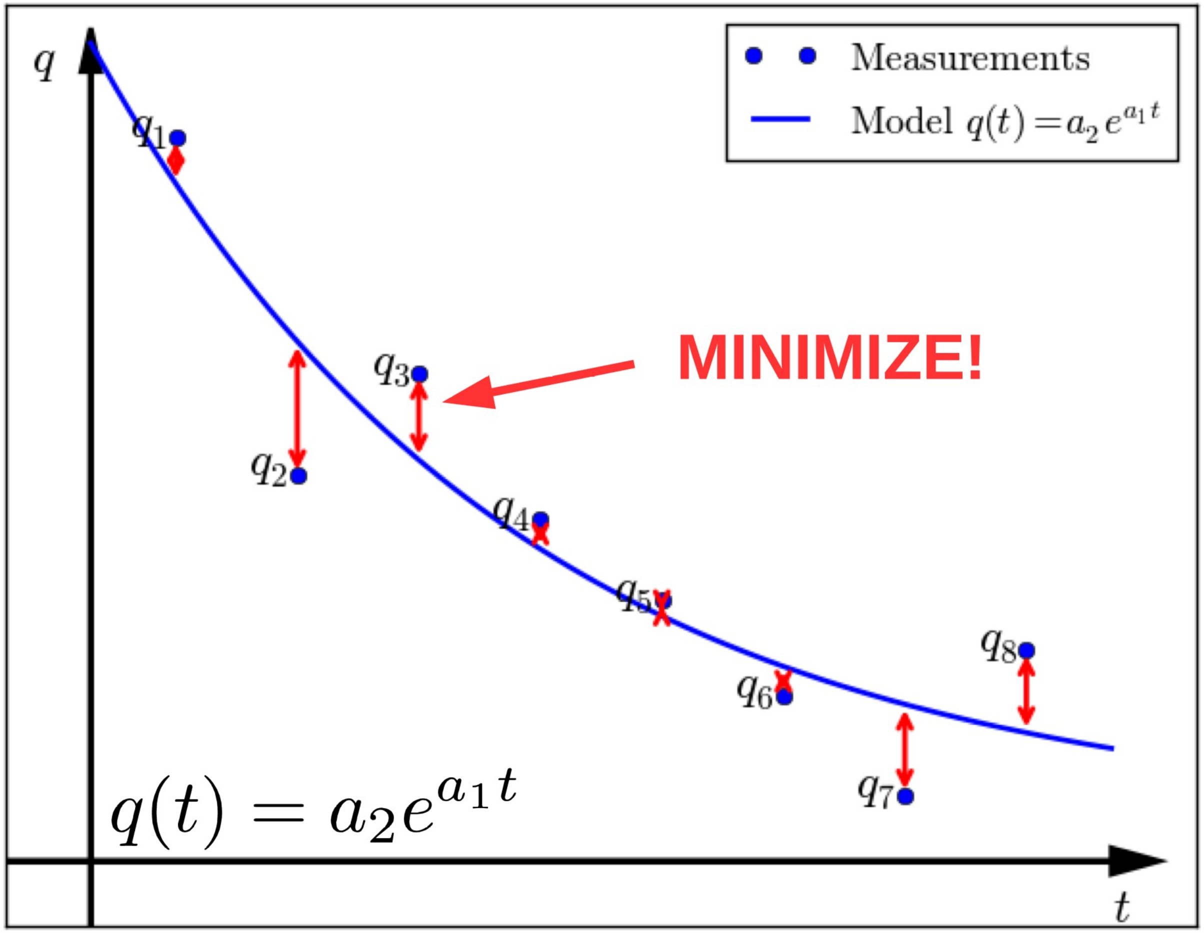
V.0. Examples, linear/nonlinear least-squares

- **Idea:** choose the parameters such that the distance between the data and the curve is minimal, i.e. the curve that fits best.



V.0. Examples, linear/nonlinear least-squares

- **Idea:** choose the parameters such that the distance between the data and the curve is minimal, i.e. the curve that fits best.



V.0. Examples, linear/nonlinear least-squares

- General problem: m measurements n parameters

  $m \gg n$ Overdetermined!

Linear

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}\right\} A\mathbf{x} = \mathbf{b} \qquad A \in \mathbb{R}^{m\times n}$$

$$\mathbf{x} \in \mathbb{R}^n \qquad \mathbf{b} \in \mathbb{R}^m$$

Nonlinear

$$\left.\begin{aligned} f_1(x_1, x_2, ..., x_n) &= b_1 \\ f_2(x_1, x_2, ..., x_n) &= b_2 \\ &\vdots \\ f_m(x_1, x_2, ..., x_n) &= b_m \end{aligned}\right\} \mathbf{f}(\mathbf{x}) = \mathbf{b}$$

$$\mathbf{f} : D \subset \mathbb{R}^n \to \mathbb{R}^m$$

In parameters!

$f_1 = f_2 = \cdots = f_m$ usually...

V.0. Examples, linear/nonlinear least-squares

- Least squares solution:

$$\min_{\mathbf{x}\in D} \|A\mathbf{x}-\mathbf{b}\|_2 \qquad \textbf{Linear}$$

$$\min_{\mathbf{x}\in D} \|\mathbf{f}(\mathbf{x})-\mathbf{b}\|_2 \qquad \textbf{Nonlinear}$$

- Define scalar-valued function $\phi(\mathbf{x})$

$$\phi(\mathbf{x}) = \frac{1}{2}\|A\mathbf{x}-\mathbf{b}\|_2^2 \qquad \textbf{Linear}$$

$$\phi(\mathbf{x}) = \frac{1}{2}\|\mathbf{f}(\mathbf{x})-\mathbf{b}\|_2^2 \qquad \textbf{Nonlinear}$$

- Least squares solution: $\min_{\mathbf{x}\in D} \phi(\mathbf{x})$

V.0. Examples, linear/nonlinear least-squares

- Quiz: linear or nonlinear model?

1) $q(t) = a_0 + a_1 t + a_2 t^2$ Model parameters: a_0, a_1, a_2

2) $q(t) = A \sin(\beta t + \varphi)$ Model parameters: A, β, φ

V.0. Examples, linear/nonlinear least-squares

- Quiz: linear or nonlinear model?

1) $q(t) = a_0 + a_1 t + a_2 t^2$ Model parameters: a_0, a_1, a_2

$$A = \begin{pmatrix} \vdots & \vdots & \vdots \\ 1 & t_i & t_i^2 \\ \vdots & \vdots & \vdots \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} \vdots \\ q_i \\ \vdots \end{pmatrix} \quad \mathbf{x} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}$$

2) $q(t) = A\sin(\beta t + \varphi)$ Model parameters: A, β, φ

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} \vdots \\ A\sin(\beta t_i + \varphi) \\ \vdots \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} \vdots \\ q_i \\ \vdots \end{pmatrix} \quad \mathbf{x} = \begin{pmatrix} A \\ \beta \\ \phi \end{pmatrix}$$

V.1. Linear least-squares

- General problem: m measurements n parameters

  $m \gg n$ Overdetermined!

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}\right\} A\mathbf{x} = \mathbf{b} \qquad A \in \mathbb{R}^{m\times n}$$

$$\mathbf{x} \in \mathbb{R}^n \qquad \mathbf{b} \in \mathbb{R}^m$$

Assumption: A has full column rank (linearly indep.)

$$\phi(\mathbf{x}) = \frac{1}{2}\|A\mathbf{x} - \mathbf{b}\|_2^2$$

Least squares solution: $\min\limits_{\mathbf{x}\in D} \phi(\mathbf{x})$

V.1.1. Normal equations

- Least squares solution:

$$\min_{\mathbf{x}\in D} \phi(\mathbf{x}) \qquad\qquad \phi(\mathbf{x}) = \frac{1}{2}\|A\mathbf{x}-\mathbf{b}\|_2^2$$

- Rewrite:

$$\begin{aligned}
\phi(\mathbf{x}) &= \frac{1}{2}\|A\mathbf{x}-\mathbf{b}\|_2^2 \\
&= \frac{1}{2}\left(A\mathbf{x}-\mathbf{b}\right)^T\left(A\mathbf{x}-\mathbf{b}\right) \\
&= \frac{1}{2}\left((A\mathbf{x})^T A\mathbf{x} - (A\mathbf{x})^T\mathbf{b} - \mathbf{b}^T A\mathbf{x} + \mathbf{b}^T\mathbf{b}\right) \\
&= \frac{1}{2}\left(\mathbf{x}^T A^T A\mathbf{x} - \underbrace{\mathbf{x}^T A^T\mathbf{b} - \mathbf{b}^T A\mathbf{x}} + \mathbf{b}^T\mathbf{b}\right) \\
&= \frac{1}{2}\left(\mathbf{x}^T A^T A\mathbf{x} - 2\mathbf{x}^T A^T\mathbf{b} + \mathbf{b}^T\mathbf{b}\right)
\end{aligned}$$

Transpose!

Scalar!

~

$$\frac{1}{2}(ax-b)^2$$

~

V.1.1. Normal equations

- Gradient of $\phi(\mathbf{x}) = \frac{1}{2}\left(\mathbf{x}^T A^T A\mathbf{x} - 2\mathbf{x}^T A^T \mathbf{b} + \mathbf{b}^T \mathbf{b}\right)$

  must vanish at extremum (min/max):

Gradient

$$\nabla\phi(\mathbf{x}) = \begin{pmatrix} \frac{\partial\phi}{\partial x_1} \\ \frac{\partial\phi}{\partial x_2} \\ \vdots \\ \frac{\partial\phi}{\partial x_n} \end{pmatrix} = \cdots = \underbrace{A^T A\mathbf{x} - A^T\mathbf{b} = 0}_{\text{(Gauss') Normal equations}}$$

Nothing difficult…

$$\sim \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{2}(ax-b)^2\right) = aax - ab$$

V.1.1. Normal equations

- Necessary condition: $A^T A\mathbf{x} - A^T\mathbf{b} = 0$

  Not sufficient!

  (Gauss') Normal equations

- Is it a minimum?
  We have to make sure that the matrix $A^T A$ is positive definite

$\tilde{a}a > 0$

1. $\mathbf{x}^T A^T A\mathbf{x} > 0 \quad \text{for all} \quad \mathbf{x} \in \mathbb{R}^n \quad \text{and} \quad \mathbf{x} \neq 0$

   $\|A\mathbf{x}\|_2^2 > 0$ Norm (length!)

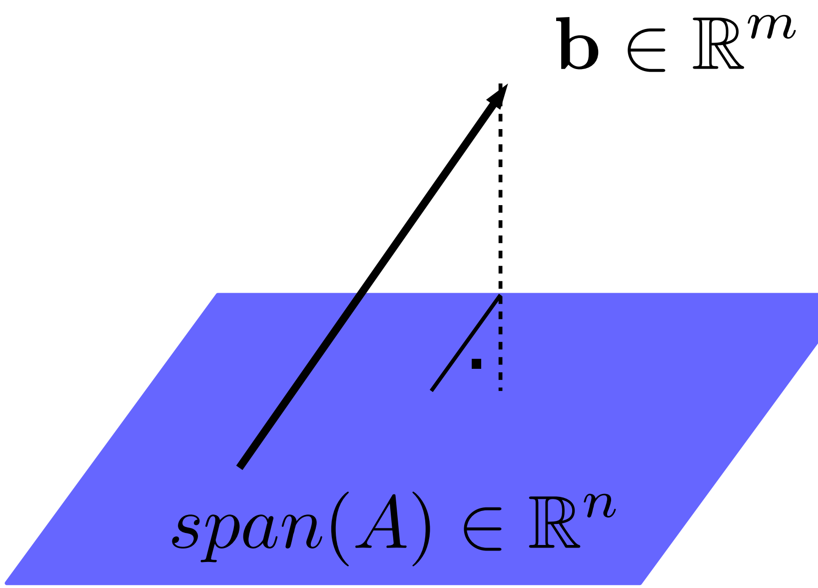
2. $\underbrace{\mathbf{x}^T A^T}_{(A\mathbf{x})^T}\underbrace{A\mathbf{x}}_{\neq 0} = 0 \quad \Leftrightarrow \quad \mathbf{x} = 0$

   $\neq 0$ Unless $\mathbf{x} = 0$

   Because of our assumption of rank n
   Columns linearly independent!

V.1.1. Normal equations

- Geometric interpretation of the normal equations



- It turns out that they lead to a worser conditioned problem, i.e. difficult to solve the normal equations numerically…
Therefore the next method is preferred

V.1.2. The orthogonal decomposition method

- Definition: An **orthogonal matrix** Q is a real square matrix whose columns and rows are orthogonal unit vectors

$$Q^T Q = Q Q^T = I = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix}$$

This means: $Q^{-1} = Q^T$

- Key property: Orthogonal matrices leave the Euclidean length invariant

$$\|Q\mathbf{x}\|_2^2 = (Q\mathbf{x})^T Q\mathbf{x} = \mathbf{x}^T \underbrace{Q^T Q}_{I} \mathbf{x} = \mathbf{x}^T \mathbf{x} = \|\mathbf{x}\|_2^2$$

V.1.2. The orthogonal decomposition method

Fact: Every matrix $A \in \mathbb{R}^{m \times n}$ with full column rank (the columns are linearly independent) has a so-called **QR-decomposition**

$$A = QR$$

where Q is an orthogonal matrix and R an upper triangular matrix

$$\begin{pmatrix} \\ A \\ m \times n \\ \\ \\ \end{pmatrix} = \begin{pmatrix} \\ Q \\ m \times m \\ \\ \\ \end{pmatrix} \underbrace{\begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & & \\ & & \ddots & \\ & & & r_{nn} \\ & & 0 & \end{pmatrix}}_{n} \left.\begin{matrix} \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right\} n \\ \\ \end{matrix}\right\} m$$

Matlab: `[Q,R]=qr(A)`

$r_{ii} \neq 0$ **!**

Non zero diagonal!

V.1.2. The orthogonal decomposition method

Fact: Every matrix $A \in \mathbb{R}^{m \times n}$ with full column rank (the columns are linearly independent) has a so-called **QR-decomposition**

$$A = QR$$

where Q is an orthogonal matrix and R an upper triangular matrix

$$\begin{pmatrix} A \\ m \times n \\ \\ \\ \end{pmatrix} = \begin{pmatrix} Q \\ m \times m \\ \\ \\ \end{pmatrix} \underbrace{\begin{pmatrix} \text{Upper triangular} \\ R_1 \\ n \times n \\ \hdashline 0 \end{pmatrix}}_{n} \left.\begin{matrix} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} n \\ \\ \\ \end{matrix}\right\} m$$

Matlab: `[Q,R]=qr(A)`

V.1.2. The orthogonal decomposition method

Minimize residuum: $$\min_{\mathbf{x}\in D} \|\underbrace{A\mathbf{x}-\mathbf{b}}_{\mathbf{r}}\|_2$$

$$\|A\mathbf{x}-\mathbf{b}\|_2^2 = \|\mathbf{r}\|_2^2$$

$$\|Q^T(A\mathbf{x}-\mathbf{b})\|_2^2 = \|Q^T\mathbf{r}\|_2^2 = \|\mathbf{r}\|_2^2$$

$$\|Q^T(QR\mathbf{x}-\mathbf{b})\|_2^2 = \|\mathbf{r}\|_2^2$$

$$\|R\mathbf{x}-Q^T\mathbf{b}\|_2^2 = \|\mathbf{r}\|_2^2$$

Orthogonal!

We still solve the same minimization problem!!!

$$\left\| \begin{pmatrix} \underset{n\times n}{R_1} \\ \hdashline 0 \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ n\times 1 \end{pmatrix} - \begin{pmatrix} \mathbf{c} \\ \hdashline \mathbf{d} \end{pmatrix} \right\|_2^2 = \overbrace{\|R_1\mathbf{x}-\mathbf{c}\|_2^2}^{0} + \|\mathbf{d}\|_2^2$$

$$= \|\mathbf{d}\|_2^2 = \|\mathbf{r}\|_2^2$$

V d

Mi

$$\begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & & \\ & 0 & \ddots & \\ & & & r_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$$

$$\mathbf{x} = R_1^{-1}\mathbf{c}$$

e the same
problem!!!

$$\left\| \begin{pmatrix} R_1 \\ n \times n \\ \hdashline 0 \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ n \times 1 \end{pmatrix} - \begin{pmatrix} \mathbf{c} \\ \hdashline \mathbf{d} \end{pmatrix} \right\|_2^2 = \overbrace{\|R_1\mathbf{x} - \mathbf{c}\|_2^2}^{0} + \|\mathbf{d}\|_2^2$$

$$= \|\mathbf{d}\|_2^2 = \|\mathbf{r}\|_2^2$$

V.1. Example: linear least-squares

Data:

i	1	2	3	4	5	6	7	8
t	0.10	0.23	0.36	0.49	0.61	0.74	0.87	1.00
q	0.84	0.30	0.69	0.45	0.31	0.09	-0.17	0.12

Model: 1 $q(t) = a_1 t + a_2$ Goal: Find a_1, a_2

Linear in model parameters!

$$\left.\begin{aligned}
0.84 &= 0.10a_1 + a_2\\
0.30 &= 0.23a_1 + a_2\\
0.69 &= 0.36a_1 + a_2\\
0.45 &= 0.49a_1 + a_2\\
0.31 &= 0.61a_1 + a_2\\
0.09 &= 0.74a_1 + a_2\\
-0.17 &= 0.87a_1 + a_2\\
0.12 &= 1.00a_1 + a_2
\end{aligned}\right\} \Rightarrow
\underbrace{\begin{pmatrix} 0.84\\ 0.30\\ 0.69\\ 0.45\\ 0.31\\ 0.09\\ -0.17\\ 0.12 \end{pmatrix}}_{\mathbf{b}} =
\underbrace{\begin{pmatrix} 0.10 & 1\\ 0.23 & 1\\ 0.36 & 1\\ 0.49 & 1\\ 0.61 & 1\\ 0.74 & 1\\ 0.87 & 1\\ 1.00 & 1 \end{pmatrix}}_{A}
\underbrace{\begin{pmatrix} a_1\\ a_2 \end{pmatrix}}_{\mathbf{x}}$$

V.1. Example: linear least-squares

Data:

i	1	2	3	4	5	6	7	8
t	0.10	0.23	0.36	0.49	0.61	0.74	0.87	1.00
q	0.84	0.30	0.69	0.45	0.31	0.09	-0.17	0.12

Model: 1 $q(t) = a_1 t + a_2$ Goal: Find a_1, a_2

Normal equations: $A^T A\mathbf{x} - A^T\mathbf{b} = 0$

$$\begin{pmatrix} 3.1 & 4.4 \\ 4.4 & 8.0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 0.844 \\ 2.62 \end{pmatrix}$$

$$\Rightarrow \quad \mathbf{x} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} -0.86 \\ 0.80 \end{pmatrix}$$

Orthogonal transformation: analogue…

V.2. Nonlinear least-squares

- General problem: m measurements n parameters

  $m \gg n$ Overdetermined!

$$\left.\begin{aligned} f_1(x_1, x_2, ..., x_n) &= b_1 \\ f_2(x_1, x_2, ..., x_n) &= b_2 \\ &\vdots \\ f_m(x_1, x_2, ..., x_n) &= b_m \end{aligned}\right\} \mathbf{f}(\mathbf{x}) = \mathbf{b}$$

$$\mathbf{f} : D \subset \mathbb{R}^n \to \mathbb{R}^m$$

$$\mathbf{x} \in \mathbb{R}^n$$

$$\mathbf{b} \in \mathbb{R}^m$$

$f_1 = f_2 = \cdots = f_m$ usually...

$$\phi(\mathbf{x}) = \frac{1}{2}\|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_2^2$$

Least squares solution: $\min\limits_{\mathbf{x}\in D} \phi(\mathbf{x})$

V.2.1 Newton method

- Gradient of $$\phi(\mathbf{x}) = \frac{1}{2}\|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_2^2$$
$$= \frac{1}{2}\sum_{i=1}^{m}\left(f_i(x_1, ..., x_n) - b_i\right)^2$$

must vanish at extremum (min/max):

Gradient

$$\nabla\phi(\mathbf{x}) = \begin{pmatrix} \frac{\partial\phi}{\partial x_1} \\ \frac{\partial\phi}{\partial x_2} \\ \vdots \\ \frac{\partial\phi}{\partial x_n} \end{pmatrix} = 0$$

n equations

n unknowns

!

V.2.1 Newton method

Apply Newton's method to solve $\nabla\phi(\mathbf{x}) = \mathbf{F}(\mathbf{x}) = 0$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - D\mathbf{F}^{-1}(\mathbf{x}^{(k)})\mathbf{F}(\mathbf{x}^{(k)})$$

Gradient $\mathbf{F}(\mathbf{x}) = \nabla\phi(\mathbf{x}) = \begin{pmatrix} \frac{\partial\phi}{\partial x_1} \\ \frac{\partial\phi}{\partial x_2} \\ \vdots \\ \frac{\partial\phi}{\partial x_n} \end{pmatrix}$

NOT $\mathbf{f}(\mathbf{x})$

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

V.2.2 Gauss-Newton method

Linearize the residuum/error equations

$$\min_{\mathbf{x}\in D}\phi(\mathbf{x}) = \min_{\mathbf{x}\in D}\frac{1}{2}\|\underbrace{\mathbf{f}(\mathbf{x})-\mathbf{b}}\|_2^2$$

Linearize at some $\mathbf{x}^{(k)}$

$$\mathbf{f}(\mathbf{x})-\mathbf{b} \approx \mathbf{f}(\mathbf{x}^{(k)}) + D\mathbf{f}(\mathbf{x}^{(k)})\left(\mathbf{x}-\mathbf{x}^{(k)}\right) - \mathbf{b}$$

$$\approx \underbrace{D\mathbf{f}(\mathbf{x}^{(k)})\mathbf{x}} + \underbrace{\mathbf{f}(\mathbf{x}^{(k)}) - \mathbf{b} - D\mathbf{f}(\mathbf{x}^{(k)})\mathbf{x}^{(k)}}$$

$$\approx A^{(k)}\mathbf{x} - \boldsymbol{\beta}^{(k)}$$

Linear least squares problem!

$$D\mathbf{f} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

V.2.2 Gauss-Newton method

Problem: $$\min_{\mathbf{x}\in D} \phi(\mathbf{x}) = \min_{\mathbf{x}\in D} \frac{1}{2}\|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_2^2$$

Given an initial guess: $\mathbf{x}^{(0)}$

Solve a sequence of linear least squares problems

$$\min_{\mathbf{x}\in D} \frac{1}{2}\|A^{(0)}\mathbf{x} - \boldsymbol{\beta}^{(0)}\|_2^2 \longrightarrow \mathbf{x}^{(1)}$$

$$\min_{\mathbf{x}\in D} \frac{1}{2}\|A^{(1)}\mathbf{x} - \boldsymbol{\beta}^{(1)}\|_2^2 \longrightarrow \mathbf{x}^{(2)}$$

$$\vdots$$

$$\min_{\mathbf{x}\in D} \frac{1}{2}\|A^{(k)}\mathbf{x} - \boldsymbol{\beta}^{(k)}\|_2^2 \longrightarrow \mathbf{x}^{(k+1)}$$ Until convergence

$\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|$ Small enough

V.2. Example: nonlinear least-squares

Data:

i	1	2	3	4	5	6	7	8
t	0.10	0.23	0.36	0.49	0.61	0.74	0.87	1.00
q	0.84	0.30	0.69	0.45	0.31	0.09	-0.17	0.12

Model: 2 $q(t) = a_2 e^{a_1 t}$

Goal: Find a_1, a_2

$$\left.\begin{aligned} q_1 &= a_1 e^{a_2 t_1} \\ q_2 &= a_1 e^{a_2 t_2} \\ q_3 &= a_1 e^{a_2 t_3} \\ q_4 &= a_1 e^{a_2 t_4} \\ q_5 &= a_1 e^{a_2 t_5} \\ q_6 &= a_1 e^{a_2 t_6} \\ q_7 &= a_1 e^{a_2 t_7} \\ q_8 &= a_1 e^{a_2 t_8} \end{aligned}\right\} \Rightarrow$$

Nonlinear in model parameters!

$$\underbrace{\begin{pmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \\ q_5 \\ q_6 \\ q_7 \\ q_8 \end{pmatrix}}_{\mathbf{b}} = \underbrace{\begin{pmatrix} a_1 e^{a_2 t_1} \\ a_1 e^{a_2 t_2} \\ a_1 e^{a_2 t_3} \\ a_1 e^{a_2 t_4} \\ a_1 e^{a_2 t_5} \\ a_1 e^{a_2 t_6} \\ a_1 e^{a_2 t_7} \\ a_1 e^{a_2 t_8} \end{pmatrix}}_{\mathbf{f}}$$

V.2. Example: nonlinear least-squares

Data:

i	1	2	3	4	5	6	7	8
t	0.10	0.23	0.36	0.49	0.61	0.74	0.87	1.00
q	0.84	0.30	0.69	0.45	0.31	0.09	-0.17	0.12

Model: 2 $q(t) = a_2 e^{a_1 t}$

Goal: Find a_1, a_2

Newton method: $\min_{\mathbf{x} \in D} \phi(\mathbf{x})$

$$\mathbf{x} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

$$\begin{aligned}
\phi(\mathbf{x}) &= \frac{1}{2} \|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_2^2 \\
&= \frac{1}{2} \sum_{i=1}^{m} (f_i(x_1, \ldots, x_n) - b_i)^2 \\
&= \frac{1}{2} \sum_{i=1}^{8} \left(a_2 e^{a_1 t_i} - q_i\right)^2
\end{aligned}$$

V.2. Example: nonlinear least-squares

Data:

i	1	2	3	4	5	6	7	8
t	0.10	0.23	0.36	0.49	0.61	0.74	0.87	1.00
q	0.84	0.30	0.69	0.45	0.31	0.09	-0.17	0.12

Model: 2 $q(t) = a_2 e^{a_1 t}$ Goal: Find a_1, a_2

Newton method: $\min_{\mathbf{x} \in D} \phi(\mathbf{x})$

$$\phi(\mathbf{x}) = \frac{1}{2} \sum_{i=1}^{8} \left(a_2 e^{a_1 t_i} - q_i\right)^2$$

Gradient

$$\mathbf{F}(\mathbf{x}) = \nabla \phi(\mathbf{x}) = \begin{pmatrix} \frac{\partial \phi}{\partial x_1} \\ \frac{\partial \phi}{\partial x_2} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{8} \left(a_2 e^{a_1 t_i} - q_i\right)\left(a_2 e^{a_1 t_i} t_i\right) \\ \sum_{i=1}^{8} \left(a_2 e^{a_1 t_i} - q_i\right)\left(e^{a_1 t_i}\right) \end{pmatrix}$$

$\mathbf{F}(\mathbf{x}) = \nabla \phi(\mathbf{x}) = 0$ Two equations in two unknowns!

V.2. Example: nonlinear least-squares

Data:

i	1	2	3	4	5	6	7	8
t	0.10	0.23	0.36	0.49	0.61	0.74	0.87	1.00
q	0.84	0.30	0.69	0.45	0.31	0.09	-0.17	0.12

Model: 2 $q(t) = a_2 e^{a_1 t}$ Goal: Find a_1, a_2

Newton method: $\min_{\mathbf{x} \in D} \phi(\mathbf{x})$

$$\phi(\mathbf{x}) = \frac{1}{2} \sum_{i=1}^{8} \left(a_2 e^{a_1 t_i} - q_i\right)^2$$

Hessian

$$D\mathbf{F}(\mathbf{x}) = H(\phi(\mathbf{x})) = \begin{pmatrix} \frac{\partial^2 \phi}{\partial x_1^2} & \frac{\partial^2 \phi}{\partial x_1 \partial x_2} \\ \frac{\partial^2 \phi}{\partial x_2 \partial x_1} & \frac{\partial^2 \phi}{\partial x_2^2} \end{pmatrix} = (\ldots)$$

V.2. Example: nonlinear least-squares

Data:

i	1	2	3	4	5	6	7	8
t	0.10	0.23	0.36	0.49	0.61	0.74	0.87	1.00
q	0.84	0.30	0.69	0.45	0.31	0.09	-0.17	0.12

Model: 2 $q(t) = a_2 e^{a_1 t}$ Goal: Find a_1, a_2

Newton method: $\min\limits_{\mathbf{x} \in D} \phi(\mathbf{x})$

$$\phi(\mathbf{x}) = \frac{1}{2} \sum_{i=1}^{8} \left(a_2 e^{a_1 t_i} - q_i \right)^2$$

Initial guess: $\mathbf{x}^{(0)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

Iterate! $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - D\mathbf{F}^{-1}(\mathbf{x}^{(k)})\mathbf{F}(\mathbf{x}^{(k)})$

V.2. Example: nonlinear least-squares

Data:

i	1	2	3	4	5	6	7	8
t	0.10	0.23	0.36	0.49	0.61	0.74	0.87	1.00
q	0.84	0.30	0.69	0.45	0.31	0.09	-0.17	0.12

Model: 2 $q(t) = a_2 e^{a_1 t}$

Goal: Find a_1, a_2

Gauss-Newton method: $$\min_{\mathbf{x} \in D} \phi(\mathbf{x})$$

$$\phi(\mathbf{x}) = \frac{1}{2} \sum_{i=1}^{8} \left(a_2 e^{a_1 t_i} - q_i \right)^2$$

Linearize residuum/error equations:

$$f_i(x_1, x_2) - b_i = x_2 e^{x_1 t_i} - b_i$$

$$\approx x_2^{(k)} e^{x_1^{(k)} t_i} - b_i + x_2^{(k)} e^{x_1^{(k)} t_i} t_i (x_1 - x_1^{(k)}) + e^{x_1^{(k)} t_i} (x_2 - x_2^{(k)})$$

Linear in parameters now!

V. Summary

- Linear/Nonlinear in parameters
- Overdetermined system of equations!
  Determine parameters in the least-squares sense...
- Linear:
  - Normal equations
  - Orthogonal transformation method
  - Example
- Nonlinear:
  - Newton method
  - Gauss-Newton method
  - Example