SUMMARY: MAXIMUM LIKELIHOOD ESTIMATION

GIVEN: PDF OF BETT. MEAS. y
GIVEN θ : $f_m(y|\theta)$

$$\hat{\theta}_{ML} = \arg\max_{\theta} f_m(y|\theta)$$

$$= \arg\min_{\theta} \underbrace{-\log f_m(y|\theta)}_{\text{NEG. LOG-LIKELIHOOD } = f(\theta)}$$

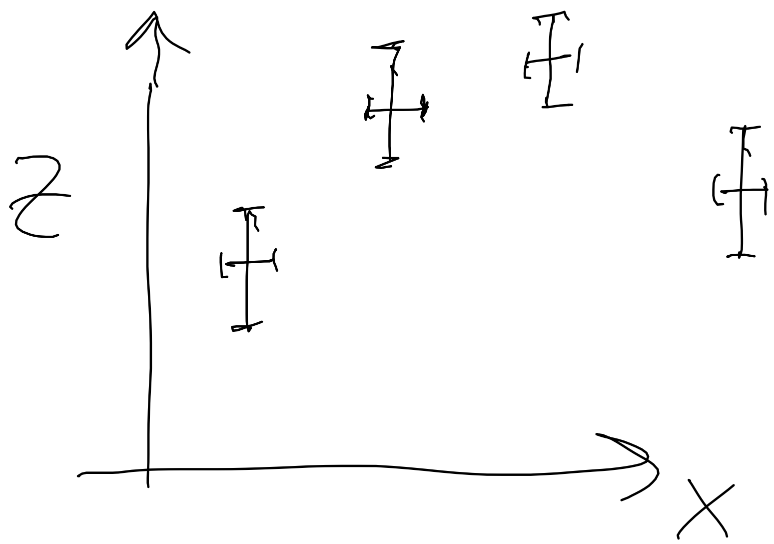
FOR GAUSSIAN MEASUREMENT ERRORS:

$$y(k) = \underbrace{M_k(\theta)}_{\text{MODEL PREDICTION}} + \underbrace{\varepsilon(k)}_{\text{MEASUREMENT NOISE, } \varepsilon(k) \sim \mathcal{N}(0,\sigma^2)}$$

$$\hat{\theta}_{ML} = \arg\min_{\theta} \sum_{k=1}^{N} \frac{(y(k) - M_k(\theta))^2}{2\sigma^2}$$

CH 7: MLE FOR FIRST PRINCIPLE MODELS (WHITE BOX)

EXAMPLE: THROW BALL, MEASURE POSITIONS



$x^m(1), \dots, x^m(N)$ DISTANCE

$z^m(1), \dots, z^m(N)$ HEIGHT

$t^m(1), \dots, t^m(N)$ TIME

AIM: FIND EARTH'S ACCELERATION g $\left(\approx 9.81 \frac{m}{s^2}\right)$

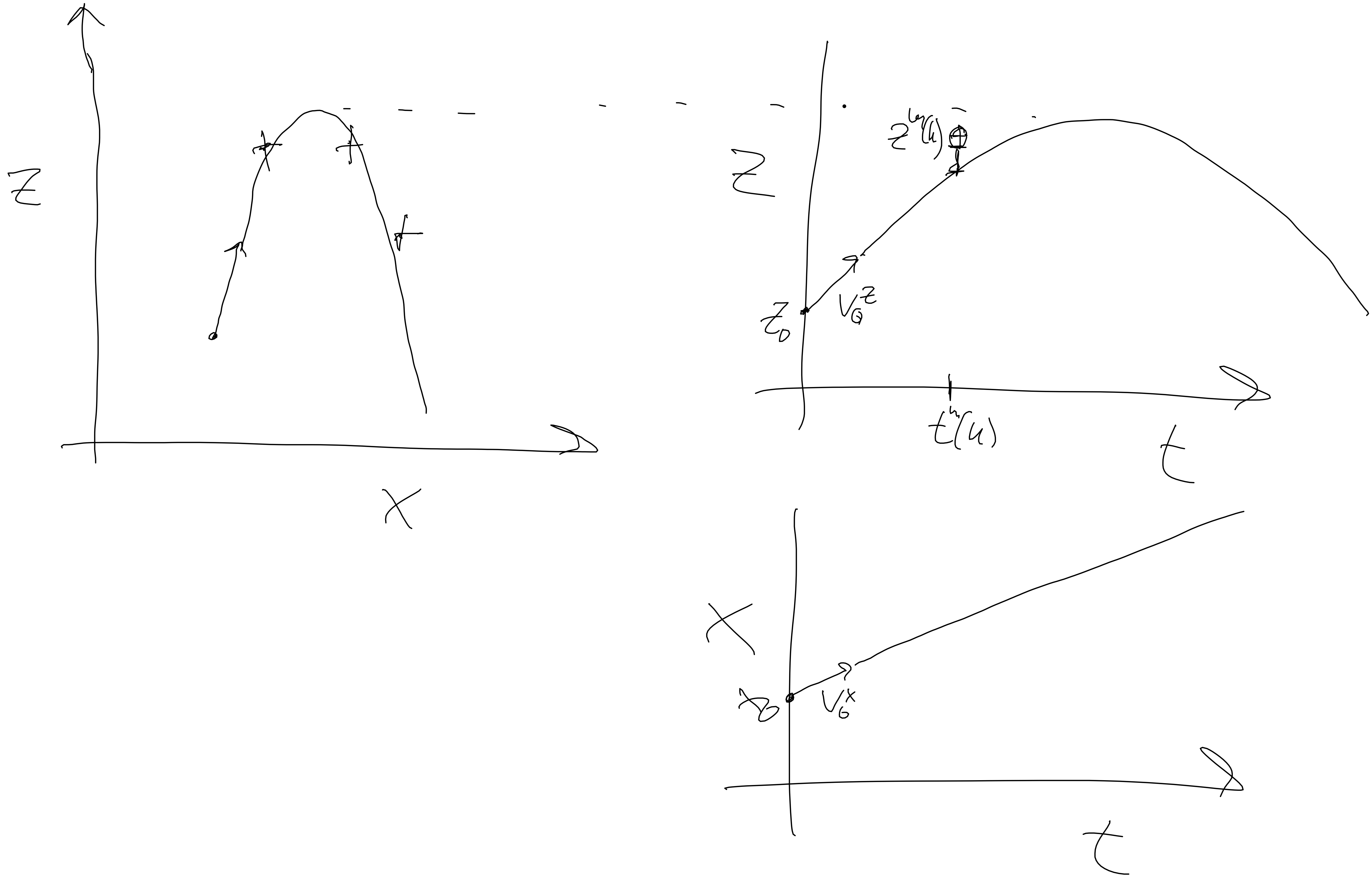
FIRST PRINCIPLE MODEL

$$\boxed{\ddot{x}(t) = 0 \quad , \quad \ddot{z}(t) = -g} \quad \text{(ODE)}$$

SIMPLEST MODEL HAS ANALYTIC SOLUTION

$$\boxed{x(t) = \underline{\underline{x_0}} + \underline{\underline{v_0^x}} \cdot t} \quad , \quad \boxed{z(t) = \underline{\underline{z_0}} + \underline{\underline{v_0^z}} \cdot t - \frac{1}{2} \underline{\underline{g}}\, t^2}$$

$$\theta = [g, x_0, z_0, v_0^x, v_0^z]^T$$


z
x
z
z_0
V_0^z
t
x
x_0
V_0^x
t

7.1 TIMES EXACT

MODEL PREDICTIONS $M_x(k) = x_0 + v_0^x \cdot t^m(k)$

$$M_z(k) = z_0 + v_0^z \cdot t^m(k) - \frac{1}{2} g \cdot \left(t^m(k)\right)^2$$

$$M(\theta) = \begin{bmatrix} M_x(1) \\ \vdots \\ M_x(N) \\ \hline M_z(1) \\ \vdots \\ M_z(N) \end{bmatrix}$$

"MODEL PREDICTIONS"

$$y = \begin{bmatrix} x^m(1) \\ \vdots \\ x^m(N) \\ \hline z^m(1) \\ \vdots \\ z^m(N) \end{bmatrix}$$

"MEASUREMENTS"

ASS: $\boxed{y = M(\theta) + \varepsilon}$

ASS: $\varepsilon \sim \mathcal{N}(0, \Sigma)$

RECALL: $\varepsilon = \begin{bmatrix} \varepsilon^x(1) \\ \vdots \\ \varepsilon^x(N) \\ \hline \varepsilon^z(1) \\ \vdots \\ \varepsilon^z(N) \end{bmatrix}$

independent noise:

$$\Sigma = \begin{bmatrix} \sigma_x(1)^2 & & & & & \\ & \ddots & & & & \\ & & \sigma_x(N)^2 & & & \\ & & & \sigma_z(1)^2 & & \\ & & & & \ddots & \\ & & & & & \sigma_z(N)^2 \end{bmatrix}$$

"independent = diagonal"

PDF $f(y|\theta)$?

$$f(y|\theta) = \text{const} \cdot \exp\left(-\frac{1}{2}(y - M(\theta))^T \Sigma^{-1} (y - M(\theta))\right)$$

$$-\log f(y|\theta) = -\log \text{const} + \underbrace{\frac{1}{2}(y - M(\theta))^T \Sigma^{-1} (y - M(\theta))}_{f(\theta)}$$

Σ pos. def. (ALL EIGS POSITIVE)

$$\Sigma = \Sigma^{\frac{1}{2}} \cdot \Sigma^{\frac{1}{2}}$$ ($\Sigma^{\frac{1}{2}}$ MATRIX SQUARE ROOT)

$\Sigma^{\frac{1}{2}} = \left(\Sigma^{\frac{1}{2}}\right)^T$ SYMMETRIC

EX: $\Sigma = \begin{bmatrix} \sigma^2(1) & & \\ & \ddots & \\ & & \sigma^2(N) \end{bmatrix}$ $\Sigma^{\frac{1}{2}} = \begin{bmatrix} \sigma(1) & & \\ & \ddots & \\ & & \sigma(N) \end{bmatrix}$

$$f(\theta) = \frac{1}{2}\left(M(\theta) - y\right)^T \left[\left(\Sigma^{\frac{1}{2}}\right) \Sigma^{\frac{1}{2}}\right]^{-1} \left(M(\theta) - y\right)$$

$$= \frac{1}{2}\left(M(\theta) - y\right)^T \left(\Sigma^{-\frac{1}{2}}\right) \left(\Sigma^{-\frac{1}{2}}\right) \left(M(\theta) - y\right)$$

$$\Sigma^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{\sigma(1)} & & \\ & \ddots & \\ & & \frac{1}{\sigma(N)} \end{bmatrix}$$

$$= \frac{1}{2} \left\| \Sigma^{-\frac{1}{2}} \left(M(\theta) - y\right) \right\|_2^2$$

"NEG-LOG-LIKELIHOOD"

EX-BALL:

$$\Sigma^{-\frac{1}{2}} = \left[\begin{array}{ccc|ccc} \frac{1}{\sigma^x} & & & & & \\ & \ddots & & & & \\ & & \frac{1}{\sigma^x} & & & \\ \hline & & & \frac{1}{\sigma^z} & & \\ & & & & \ddots & \\ & & & & & \frac{1}{\sigma^z} \end{array}\right]$$

$$f(\theta) = \frac{1}{2} \left\| \Sigma^{-\frac{1}{2}} \cdot \left(M(\theta) - y \right) \right\|_2^2$$

$$= \frac{1}{2} \left[\sum_{k=1}^{N} \left(\frac{1}{\sigma^x} \left(M^x(k;\theta) - x^m(k) \right) \right)^2 + \sum_{k=1}^{N} \left(\frac{1}{\sigma^z} M^z(k;\theta) - z^m(k) \right)^2 \right]$$

WEIGHTED <u>~~LINEAR~~ LEAST SQUARES</u>

FORTUNATELY, $\boxed{M(\theta) = b + A \cdot \theta}$

$$M^x(k;\theta) = x_0 + V_0^x \cdot t^m(k) = \begin{bmatrix} 0 & 1 & 0 & t^m(k) & 0 \end{bmatrix} \begin{bmatrix} g \\ x_0 \\ z_0 \\ V_0^x \\ V_0^z \end{bmatrix}$$

$$M^z(k;\theta) = \begin{bmatrix} -\frac{1}{2} t^m(k)^2 & 0 & 1 & 0 & t^m(k) \end{bmatrix} \begin{bmatrix} g \\ x_0 \\ z_0 \\ V_0^x \\ V_0^z \end{bmatrix}$$

$$A = \begin{bmatrix} 0 & 1 & 0 & t^m(1) & 0 \\ \vdots \\ \hline -\frac{1}{2} t^m(1)^2 & 0 & 1 & 0 & t^m(1) \\ \vdots \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \hline 0 \\ \vdots \\ 0 \end{bmatrix}$$

EXPLICIT SOLUTION FOR MODELS AFFINE IN PARAMETERS:

$$M(\theta) = b + A\theta$$

$$f(\theta) = \frac{1}{2}\left\| \Sigma^{-\frac{1}{2}}(b + A\theta - y) \right\|_2^2$$

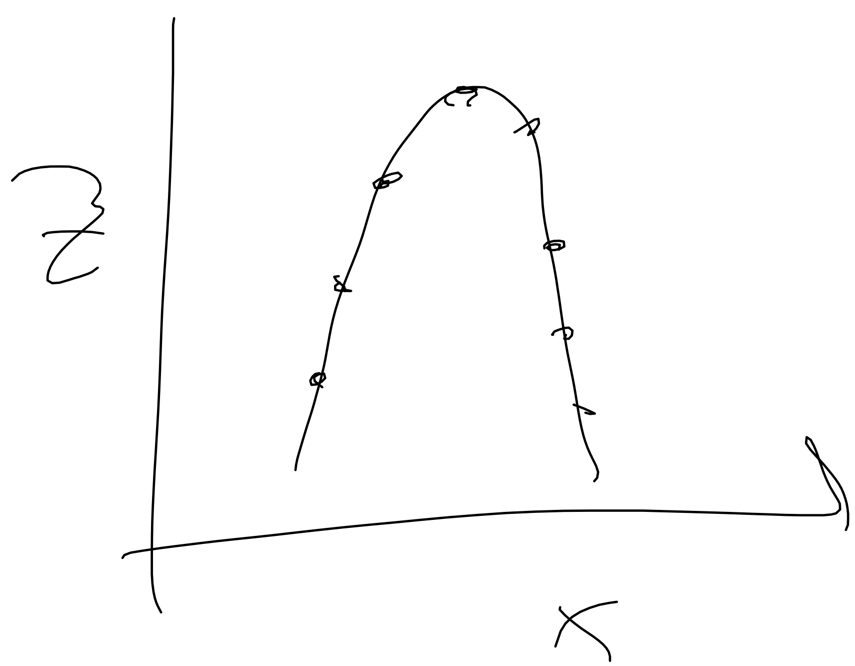
$$= \frac{1}{2}\Big\| \underbrace{\Sigma^{-\frac{1}{2}} A}_{=:J}\theta - \underbrace{\Sigma^{-\frac{1}{2}}(y-b)}_{=:\tilde{y}} \Big\|_2^2$$

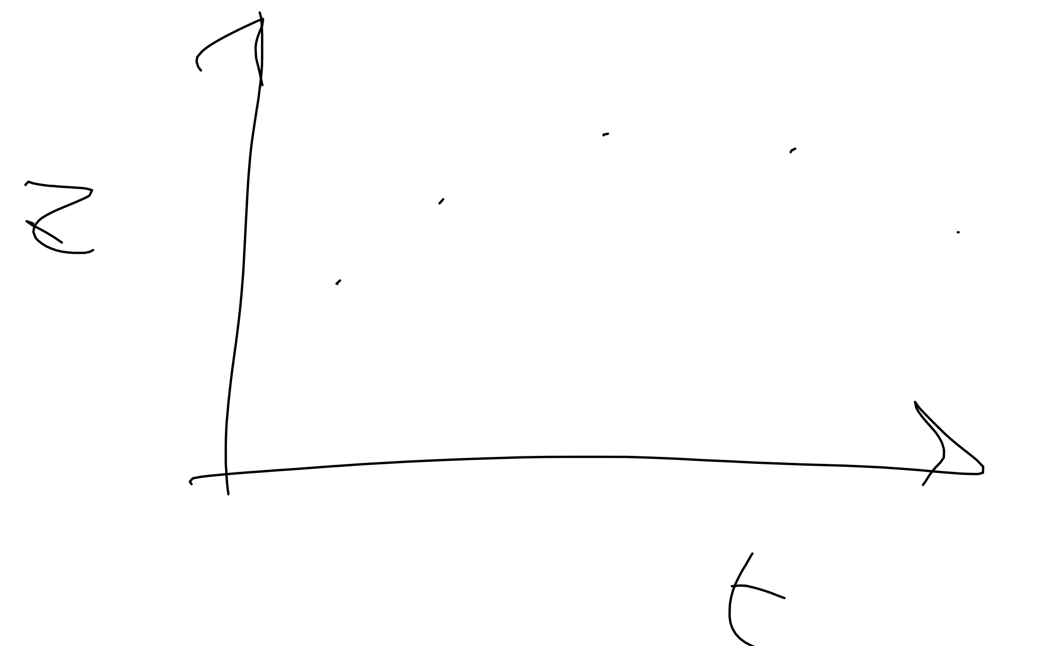
$$\hat{\theta}_{ML} = \arg\min_{\theta} f(\theta) = \arg\min \frac{1}{2}\| J\theta - \tilde{y} \|_2^2$$

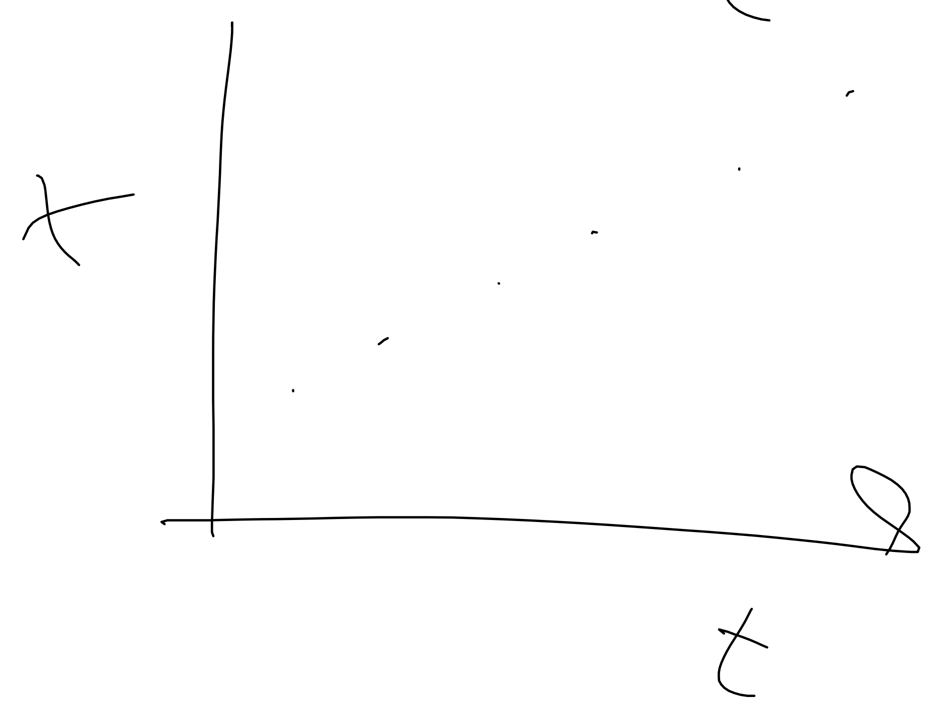
$$= J^{+} \cdot \tilde{y} \qquad J^{+} = (J^T J)^{-1} J^T$$

$$= \arg\max f_M(y|\theta)$$

$$M(\hat{\theta}_{ML}) = \delta + A \cdot \hat{\theta}_{M}$$







7.2 TIMES HAVE i.i.d. ERRORS

(A) $$t^m(k) = t(k) + \varepsilon^t(k)$$

$$\theta = \begin{bmatrix} g \\ x_0 \\ z_0 \\ v_0^x \\ v_0^z \\ \hline t(1) \\ \vdots \\ t(N) \end{bmatrix}$$

$$M^x(k;\theta) = x_0 + v_0^x \cdot t(k)$$

$$M^z(k;\theta) = z_0 + v_0^z\, t(k) - \frac{1}{2} g \left(t(k)\right)^2$$

$$M^t(k;\theta) = t(k)$$

FOR LATER:

(B) $$\left[t^m(k+1) = t^m(k) + \underline{c} + \varepsilon^e(k) \right]$$

$$Y = \begin{bmatrix} x^m(1) \\ \vdots \\ \hline z^m(1) \\ \vdots \\ \hline t^m(1) \\ \vdots \\ t^m(N) \end{bmatrix} \qquad M(\theta) = \begin{bmatrix} M^x(1;\theta) \\ \vdots \\ \hline M^z(1;\theta) \\ \vdots \\ \hline M^t(1;\theta) \\ \vdots \\ M^t(N;\theta) \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \sigma_x^2 & & & & & \\ & \ddots & & & & \\ & & \sigma_z^2 & & & \\ & & & \ddots & & \\ & & & & \sigma_t^2 & \\ & & & & & \ddots \\ & & & & & & \sigma_t^2 \end{bmatrix}$$

$$f(\theta) = \frac{1}{2} \left\| \Sigma^{-\frac{1}{2}} \left(M(\theta) - Y \right) \right\|_2^2$$

$$= \frac{1}{2} \sum_{k=1}^{N} \left[\left(\frac{x^m(k)}{\sigma_x} - \frac{(\underline{\underline{x_0}} + \underline{\underline{v_0^x}}\,\underline{\underline{t}}(k))}{\sigma_x} \right)^2 + \left(\frac{z^m(k) - (\underline{\underline{z_0}} + \underline{\underline{v_0^z}}\,\underline{\underline{t}}(k) - \frac{1}{2}\underline{\underline{g}}\,(\underline{\underline{t}}(k))^2)}{\sigma_z} \right)^2 + \left(\frac{t^m(k) - \underline{\underline{t}}(k)}{\sigma_t} \right)^2 \right]$$

$$\underline{\underline{\theta}}$$

NONLINEAR LEAST SQUARES

CAN DEFINE $\tilde{M}(\theta) = \Sigma^{-\frac{1}{2}} \cdot M(\theta)$

$$\tilde{y} = \Sigma^{-\frac{1}{2}} y$$

$$f(\theta) = \frac{1}{2} \| \tilde{M}(\theta) - \tilde{y} \|_2^2$$

NONLINEAR LEAST-SQUARES (UNWEIGHTED)

(B) : DIFFERENT TIME MEASUREMENT MODEL

$$\boxed{t^m(k+1) = t^m(k) + \underline{\underline{c}} + \underline{\underline{\varepsilon^e(k)}}}$$

$$\varepsilon^e(k) \sim \mathcal{N}(0, \sigma_e^2)$$

$$t(k+1) = t(k) + c \qquad \boxed{t(k) = k \cdot \underline{\underline{c}} + \cancel{\underline{\underline{t_0}}}}$$

$$M^e(k) := \boxed{t^m(k+1) - \left(t^m(k) + c\right) = \underline{\underline{\varepsilon^e(k)}}}$$

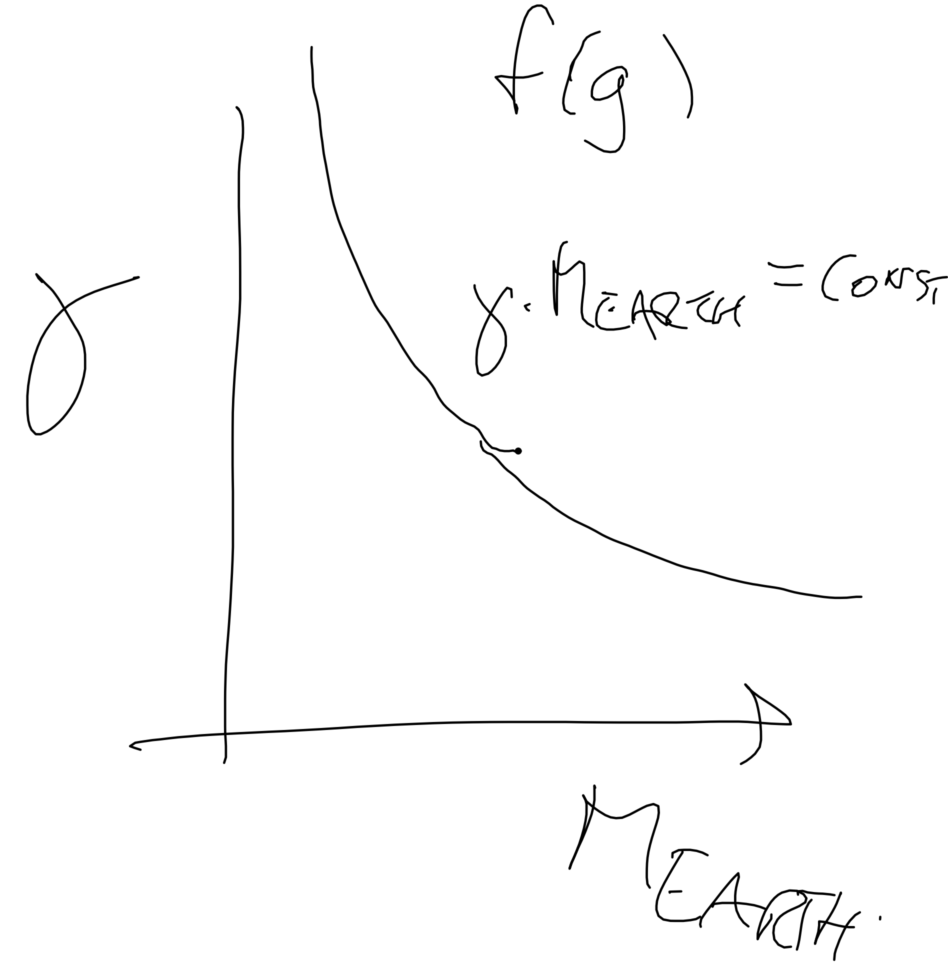
$$\Theta = \begin{bmatrix} g \\ x_0 \\ z_0 \\ v_0^x \\ v_0^z \\ \cancel{t_0} \\ c \end{bmatrix}, \qquad M(\Theta) = \begin{bmatrix} M^x(\Theta) \\ \vdots \\ \hline M^z(\Theta) \\ \vdots \\ \hline M^e(\Theta) \end{bmatrix}$$

BE AWARE OF UNDER-DETERMINED PROBLEMS !

$$g = \gamma \cdot M_{EARTH}$$

γ: GRAV. CONSTAN

M_{EARTH}: MASS EARTH



$$f(g) = f(\gamma \cdot M_{EARTH})$$

$$= f\left(\underbrace{\frac{\gamma}{\alpha}} \cdot \underbrace{\alpha \cdot M_{EARTH}}\right)$$

HOW TO DETECT?

FOR LLS:

IF $(J^T J)^{-1}$ EXISTS, PROBLEM IS OK (NOT UNDERDETERMINED)

OTHERWISE ($J^T J$ HAS ~~ZERO~~ ZERO EIGENVALUE) NOT

EXAMPLE FOR NON-UNIQUE SOLUTION (UNDERDETERMINED)

$$y^m(k) = \frac{L}{C} + \varepsilon(k)$$

σ_e

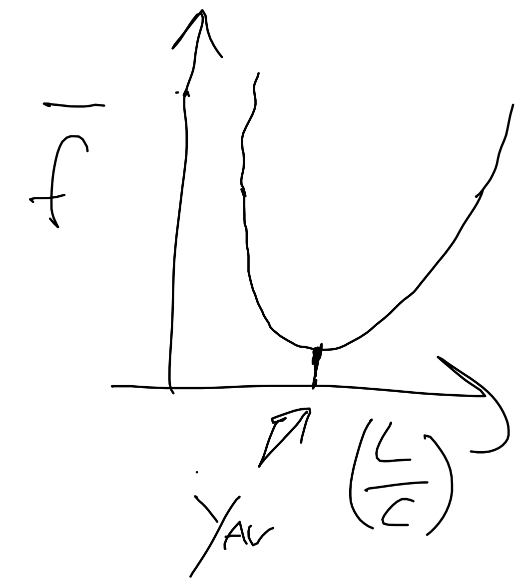
$$\theta = \begin{bmatrix} L \\ c \end{bmatrix}$$

$$f^m(y|\theta) = \prod \exp\left(-\frac{1}{2}\left(\frac{y^m(k) - \frac{L}{c}}{\sigma_e}\right)^2\right)$$

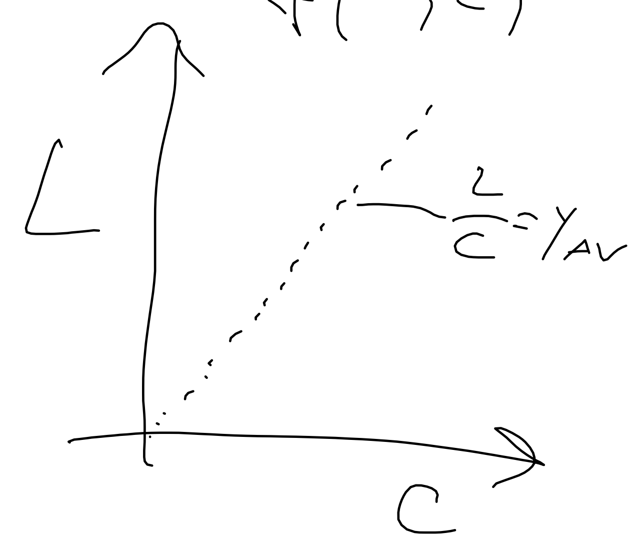
$$-\log f^m(y|\theta) = \text{const} + \underbrace{\frac{1}{2}\sum_{k=1}^{N}\left(\frac{y^m(k) - \left(\frac{L}{c}\right)}{\sigma_e}\right)^2}_{=: f(\theta)}$$

$$= \bar{f}\left(\frac{L}{c}\right)$$

$$f(\theta) = \bar{f}\left(\frac{L}{c}\right)$$



$$f(\theta) = f(L, c)$$



MOTIVATION FOR NEXT CHAPTER 8.

1) HOW TO FIND $\boxed{\arg\min \frac{1}{2}\|\tilde{M}(\theta)-\tilde{y}\|_2^2}$

$\phi^+ = (\phi^T\phi)^{-1}\phi^T$ $\boxed{\theta = \phi^+ y}$

2) HOW TO CHECK THAT SOLUTION IS (LOCALLY) UNIQUE?

$\phi^T\phi$ INVERTIBLE

3) WHAT COVARIANCE DOES MY ESTIMATOR HAVE?

$\boxed{\Sigma_\theta = \phi^+ \Sigma_y (\phi^+)^T}$

FOR LLS

WHAT ABOUT NLS ?

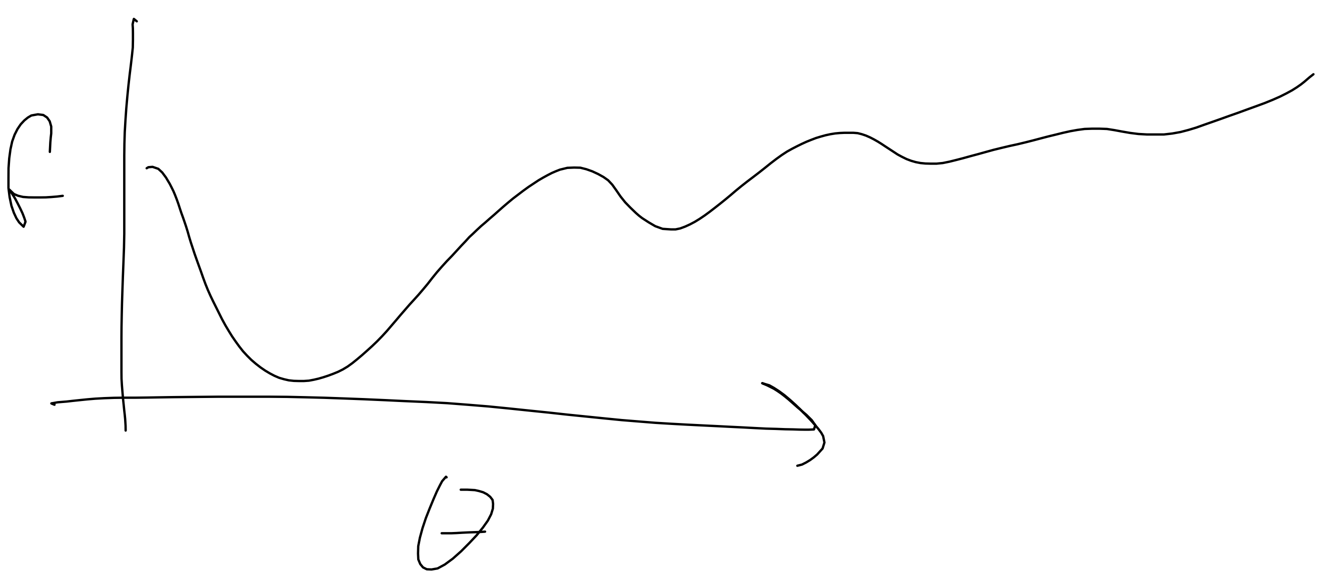
CH 8 SOLVING NONLINEAR OPTIM. PROBLEMS

TASK:

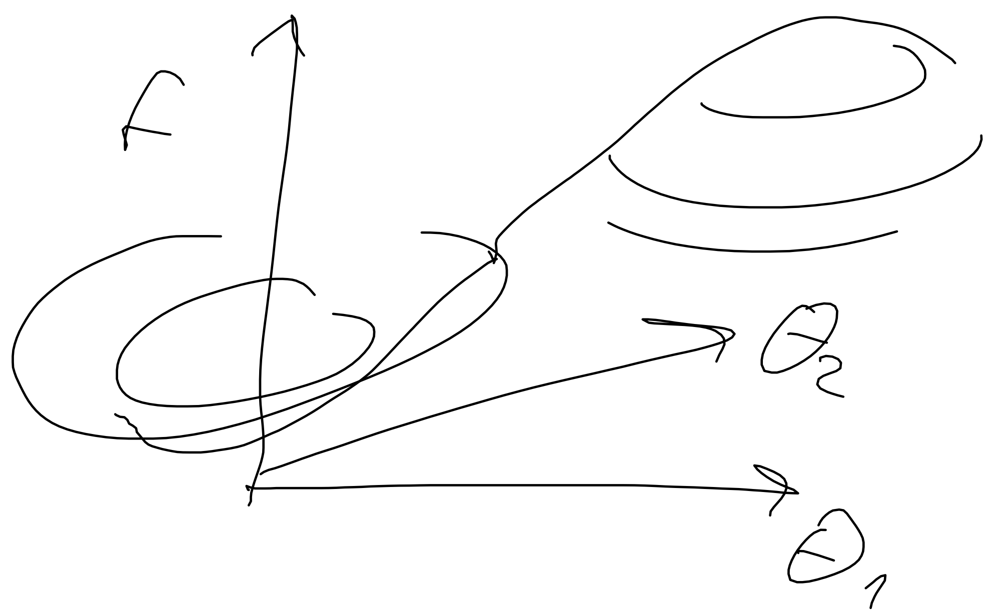
SOLVE $\boxed{\min_{\theta} \underbrace{\frac{1}{2} \| M(\theta) - y \|_2^2}_{f(\theta)}}$

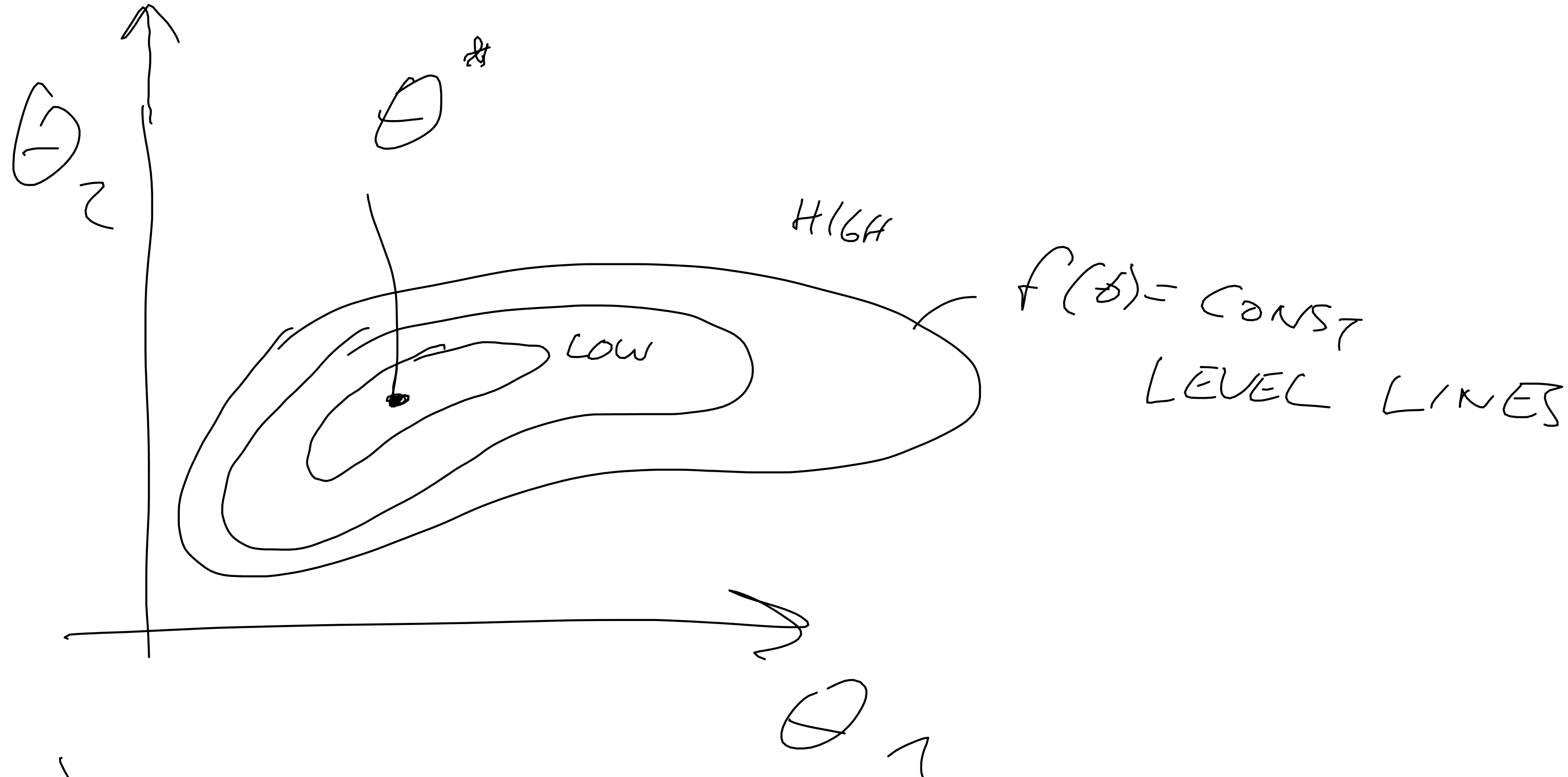
$f(\theta)$: OBJECTIVE $\qquad f: \mathbb{R}^d \to \mathbb{R}$

1-D



2-D





$$\cancel{\theta_{ML}} = \theta^* = \arg\min_{\theta} f(\theta)$$

$$f(\theta^*) \le f(\theta) \qquad \forall \theta \in \mathbb{R}^d$$

ASSUME : f TWICE DIFFERENTIABLE

GRADIENT $\nabla f(\theta)$ EXISTS

<u>HESSIAN $\nabla^2 f(\theta)$</u> EXISTS

EX: $f(\theta) = \theta^3 + \sin\theta$

$\theta \in \mathbb{R}$ $\nabla f(\theta) = 3\cdot\theta^2 + \cos\theta$

$\nabla^2 f(\theta) = 6\cdot\theta - \sin\theta$

EX 2: $f(\theta) = \theta_1 \cdot \sin\theta_2$

$\theta \in \mathbb{R}^2$ $\frac{\partial f}{\partial\theta_1} = \sin\theta_2$ $\frac{\partial f}{\partial\theta_2} = \theta_1 \cdot \cos\theta_2$

$$\nabla f(\theta) = \begin{bmatrix} \frac{\partial f}{\partial\theta_1} \\ \\ \frac{\partial f}{\partial\theta_2} \end{bmatrix} = \begin{bmatrix} \sin\theta_2 \\ \\ \theta_1 \cdot \cos\theta_2 \end{bmatrix}$$

$\frac{\partial^2 f}{\partial\theta_1\partial\theta_1} = 0$, $\frac{\partial^2 f}{\partial\theta_2\partial\theta_1} = \cos\theta_2$, $\frac{\partial^2 f}{\partial\theta_1\partial\theta_2} = \cos\theta_2$, $\frac{\partial^2 f}{\partial\theta_2\partial\theta_2} = -\theta_1\cdot\sin\theta_2$

$=$

$$\nabla^2 f(\theta) = \begin{bmatrix} \frac{\partial^2 f}{\partial\theta_1\partial\theta_1} & \frac{\partial^2 f}{\partial\theta_1\partial\theta_2} \\ \frac{\partial^2 f}{\partial\theta_2\partial\theta_1} & \frac{\partial^2 f}{\partial\theta_2\partial\theta_2} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & \cos\theta_2 \\ \cos\theta_2 & -\theta_1 \sin\theta_2 \end{bmatrix}$$

"HESSIAN MATRIX"

ALWAYS SYMMETRIC $\nabla^2 f(\theta) = \nabla^2 f(\theta)^T$

$\rightarrow$ ALL EIGS ARE REAL!

$$\nabla^2 f(\theta) = Q \cdot D \cdot Q^T$$

Q ORTHONORMAL, (ROTATION!) $\quad Q^{-1} = Q^T \quad Q^T Q = \mathbb{I}$

IDENTITY

D DIAGONAL $\quad D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{bmatrix} \quad \lambda_i \in \mathbb{R}$

8.1 NECESSARY OPTIMALITY CONDITION

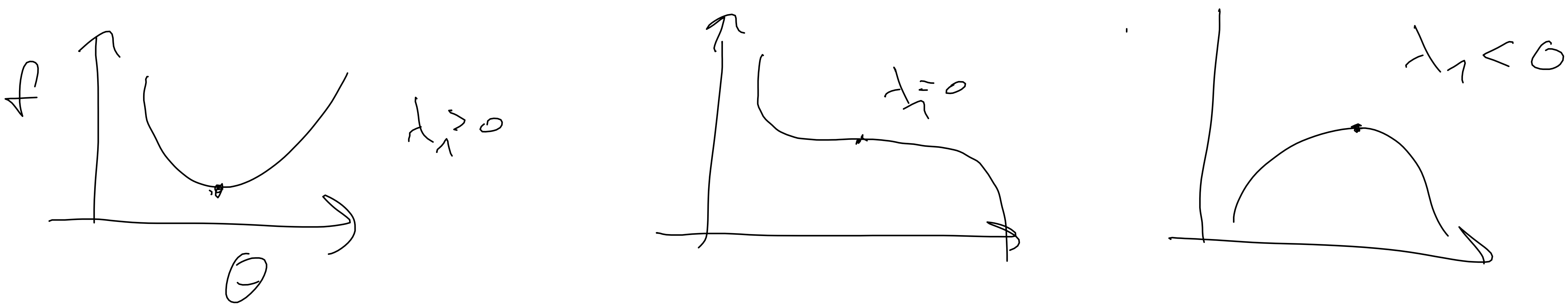
FIRST ORDER $\quad \nabla f(\theta^*) = 0 \quad$ "STATIONARY"

SECOND ORDER $\quad \nabla^2 f(\theta^*) \succeq 0 \iff \lambda_1, \ldots, \lambda_d \geq 0$

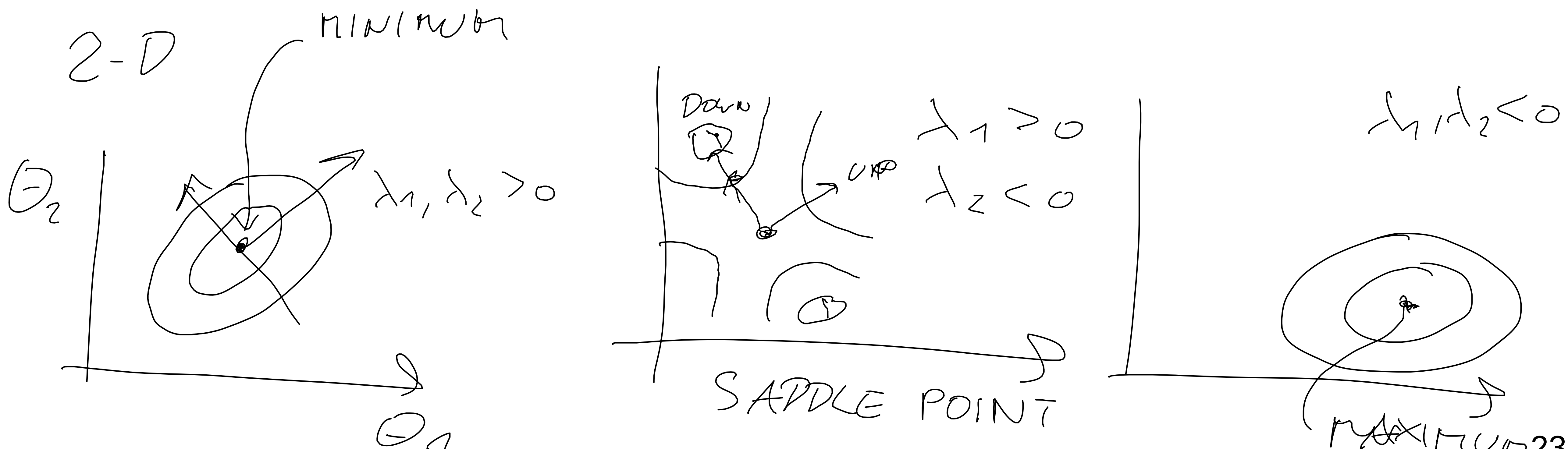
$\iff$ HESSIAN POSITIVE SEMI-DEFINITE

VISUALIZATION OF STATIONARY POINTS

1-D: $\lambda_1 = \frac{\partial^2 f}{\partial \theta^2}(\theta) = \nabla^2 f(\theta)$



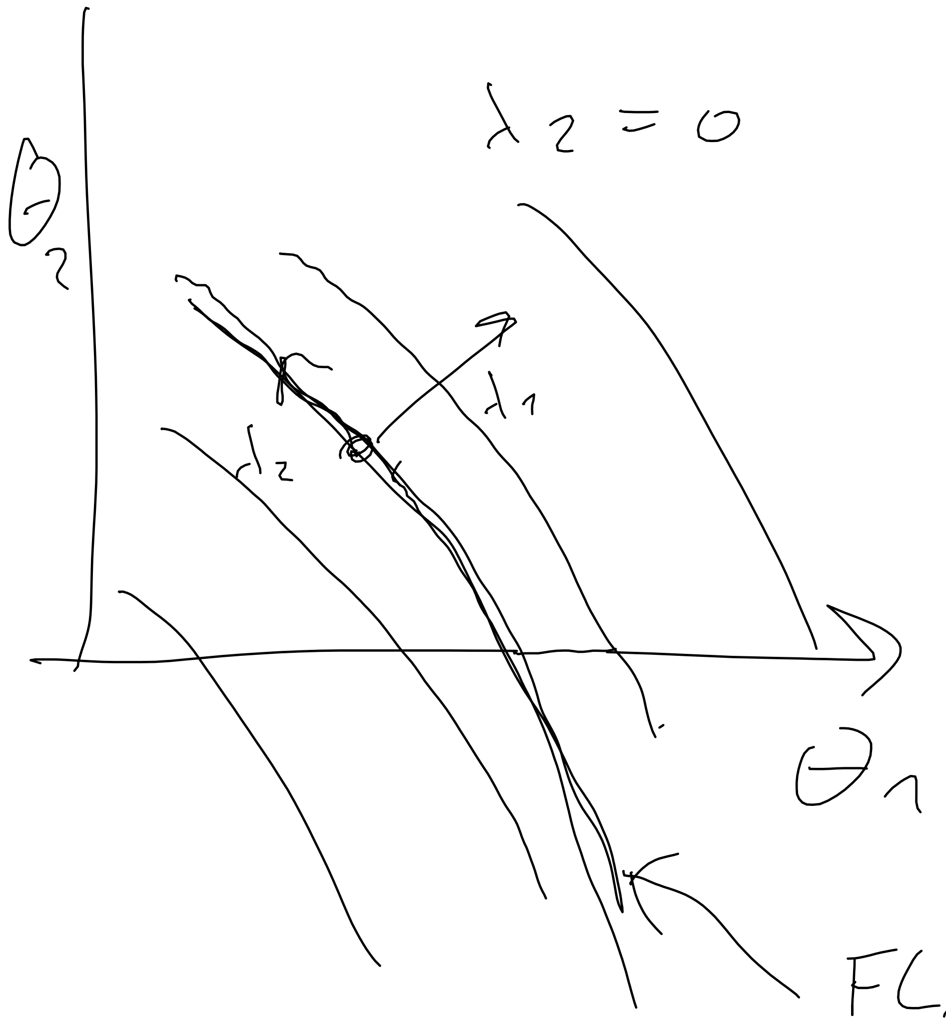
2-D



MORE OF 2-D:

$\lambda_1 > 0$

$\lambda_2 = 0$



FLAT VALLEY

NON UNIQUENESS