AE 514 : Estimation Theory & KALMAN Filtering.

Chapter 1 : Least Squares (observers)

Ex: satellite : star locations → position, velocity, orientation, angular vel.

airplane : GPS, airspeed, gyroscopes → position, vel., orient., angular vel

self driving car : LIDAR, speedometer → position (lane position), velocity.

NOISE IN MEASUREMENTS → estimate

new meas → updated estimate → use for control (based on separation principle)

Notation:

x : true state (unknown)
$\tilde{x}$: measured state (known)
$\hat{x}$: estimated value (computed)

} state

↑ all vectors

v : measurement noise (unknown) — stat model (known) — Noise in the sensor

w : process noise (unknown) — stat model — noise in the dynamics

↓ "modeling noise as coming from a Gaussian distribution"

e : residual error

$e = \tilde{x} - \hat{x}$ } → how much our estimate differs from meas.

$(x) - \hat{x}$

$\tilde{x} = x + v$

$\tilde{x} = \hat{x} + e = \not{\tilde{x}} + \hat{x} - \not{\tilde{x}}$

Greatest Mathematician

↙ "Few, but ripe."

LEAST SQUARES: GAUSS (1820)

knew how to do FFT (1960's)

Model: $y(t) = \sum_{i=1}^{n} x_i h_i(t)$ → h_i: basis function → fit curves

↑ output; ↑ parameters

($\sum_{i=1}^{n} x_i a_{it}$ → fitting a line

least squares is linear in x_i's ($h_i(t)$ might not be linear)

m: measurements. $m \gg n$

n: parameters.

Model: conceptual

$$\begin{bmatrix} y(t_1) \\ \vdots \\ y(t_m) \end{bmatrix} = \begin{bmatrix} h_1(t_1) & \cdots & h_n(t_1) \\ & \vdots & \\ h_1(t_m) & \cdots & h_n(t_m) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

(scalars; n)

tall. → find closest parameters that fit the output data

Measurements

$$\begin{bmatrix} \tilde{y}(t_1) \\ \vdots \\ \tilde{y}(t_m) \end{bmatrix} = \underbrace{\begin{bmatrix} h_1(t_1) & \cdots & h_n(t_1) \\ & \vdots & \\ h_1(t_m) & \cdots & h_n(t_m) \end{bmatrix}}_{H} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} v(t_1) \\ \vdots \\ v(t_m) \end{bmatrix}$$

(m ↓ t; n)

given a state estimate:
compute what we would expect the output to be...

$$\underbrace{\begin{bmatrix} \hat{y}(t_1) \\ \vdots \\ \hat{y}(t_m) \end{bmatrix}}_{\text{estimate of output}} = \underbrace{\begin{bmatrix} & & \\ & H & \\ & & \end{bmatrix}}_{\text{data}} \underbrace{\begin{bmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_n \end{bmatrix}}_{\text{state estimate}}$$

$\hat{y}$, $\hat{x}$

$e = \tilde{y} - \hat{y}$ error in the output.

$$\hat{y} = H\hat{x} + \tilde{y} - \tilde{y}$$

$$\Rightarrow \tilde{y} = H\hat{x} + \tilde{y} - \hat{y}$$

$$\boxed{\tilde{y} = H\hat{x} + e}$$

$\tilde{y}$: meas. ; $\hat{x}$: pick $\hat{x}$ to minimize

$$e = \tilde{y} - H\hat{x}$$

$$\min \ \frac{1}{2}\|e\|_2^2$$

$$\min_{\hat{x}} J = \frac{1}{2}(\tilde{y} - H\hat{x})^T(\tilde{y} - H\hat{x})$$

minimize by setting $\frac{\partial J}{\partial \hat{x}} = 0$

Vector Derivatives

$x, \ f(x)$ $\quad \Delta f = \frac{\partial f}{\partial x}\Delta x$ (Δf: how much does f get perturbed; Δx: perturb x)

$\partial f = \frac{\partial f}{\cancel{\partial x}}\cancel{\partial x}$ Leibnitz

$f: \mathbb{R} \to \mathbb{R}$. $\quad \Delta f = \frac{\partial f}{\partial x}\Delta x$.

$f: \underset{x}{\mathbb{R}^n} \to \mathbb{R}$ $\quad \Delta f = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{bmatrix} = \sum_i \frac{\partial f}{\partial x_i}\Delta x_i$ perturb ea. part separately and sum up

$f: \underset{x}{\mathbb{R}} \to \underset{f}{\mathbb{R}^m}$ $\quad \Delta f = \begin{bmatrix} \Delta f_1 \\ \vdots \\ \Delta f_m \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x} \\ \vdots \\ \frac{\partial f_m}{\partial x} \end{bmatrix} \Delta x$

$f: \mathbb{R}^n \to \mathbb{R}^m$ $\quad \Delta f = \begin{bmatrix} \Delta f_1 \\ \vdots \\ \Delta f_m \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}}_{m \times n} \begin{bmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{bmatrix}$ (output perturb ↕; input perturb)

$$f(x) = \sin(x^T x) + \ln(c^T x)$$
$$= \sin(\Delta x^T x) + \sin(x^T \Delta x) + \ln(c^T \Delta x)$$
$$\Delta f = \begin{bmatrix} \_ \end{bmatrix}\Delta x$$

Ex. • $f(x) = c^T x$ $\quad \frac{\partial f}{\partial x} = c^T$ $\quad \Delta f = c^T \Delta x$
$c \in \mathbb{R}^n$

• $f(x) = Ax$ $\quad \frac{\partial f}{\partial x} = A$ $\quad \Delta f = A\Delta x$
$A \in \mathbb{R}^{m \times n}$

• $f(x) = \frac{1}{2}x^T Q x$ ← perturb ea. x separately and sum } product rule

$$\Delta f = \frac{1}{2}\underbrace{\Delta x^T Q x}_{\text{scalar}} + \frac{1}{2}x^T Q \Delta x = \frac{1}{2}(x^T Q^T \Delta x + x^T Q \Delta x) = \frac{1}{2}x^T(Q^T + Q)\Delta x$$

(transpose)

$$\frac{\partial f}{\partial x} = \frac{1}{2} x^T(Q^T+Q)$$

$$\frac{\partial}{\partial x}\left(\frac{1}{2}x^T Q x\right) = x^T Q$$ if Q is sym — always assume

$A \in \mathbb{R}^{n \times n}$ not nec. sym

$$A = \frac{1}{2}(A+A^T) + \frac{1}{2}(A-A^T)$$ sym/skew decomp

$$x^T A x = \frac{1}{2}x^T(A+A^T)x + \frac{1}{2}x^T(A-A^T)x$$

space of matrices: skew (high dim), sym (high dim)

$$\frac{1}{2}x^T A x - \frac{1}{2}x^T A^T x = 0$$

$$x^T K x = 0$$ (K skew sym)

- $f(x) = \sin(x^T x) + \ln(c^T x)$

use chain rule.

$$\Delta f = \frac{\partial \sin}{\partial u}\bigg|_{x^T x}\frac{\partial u}{\partial x}\Delta x + \frac{\partial \ln}{\partial u'}\frac{\partial u'}{\partial x}\Delta x$$

$$= \cos(x^T x)\left[\Delta x^T x + x^T \Delta x\right] + \frac{1}{c^T x}c^T \Delta x$$

$$= \cos(x^T x)\left[2x^T \Delta x\right] + \frac{1}{c^T x}c^T \Delta x$$

$$\frac{\partial f}{\partial x} = \left[2\cos(x^T x)x^T + \frac{1}{c^T x}c^T\right]$$

$$f(x) = x^T Q x \qquad \frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial}{\partial x}\left(x^T(Q+Q^T)\frac{1}{2}\right)$$

$$\Delta \frac{\partial f}{\partial x} = \Delta x^T \underbrace{(Q+Q^T)\frac{1}{2}}_{\frac{\partial^2 f}{\partial x^2}} \qquad \frac{\partial^2 f}{\partial x^2} = \frac{1}{2}(Q+Q^T)$$

Taylor Exp of $f(x)$

$$\Delta \frac{\partial f}{\partial x} = \Delta x^T\left[\frac{1}{2}(Q+Q^T)\right]$$

$$f(x+\Delta x) = f(x) + \frac{\partial f}{\partial x}\Delta x + \frac{1}{2}\Delta x^T \frac{\partial^2 f}{\partial x^2}\Delta x + \sum_{ijk}\left(\frac{\partial^3 f}{\partial x^3}\right)_{ijk}\Delta x_i \Delta x_j \Delta x_k$$

($f(x)$: scalar; $\frac{\partial f}{\partial x}$: row vec; Δx: col vec; Δx^T: row vec; $\frac{\partial^2 f}{\partial x^2}$: matrix; Δx: col vec)

$$\frac{\partial^3 f}{\partial x^3} = \frac{\partial}{\partial x}\left[\underbrace{\frac{\partial^2 f}{\partial x^2}}_{\text{matrix}}\right] \qquad \Delta x^T\left[\frac{\partial^3 f}{\partial x^3}\right]\Delta x$$

matrix cookbook

FROM ABOVE ... 3 sensor

$$\underbrace{\begin{bmatrix} \hat{y}(t_1) \\ \vdots \\ \hat{y}(t_m) \end{bmatrix}}_{y,\ p} = \underbrace{\begin{bmatrix} & H & \end{bmatrix}}_{\text{data}} \underbrace{\begin{bmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_n \end{bmatrix}}_{x}$$

estimate of output — data — state estimate

$e = \tilde{y} - \hat{y}$ error in the output.

minimize magnitude of residual error.

$\hat{y} = H\hat{x} + \tilde{y} - \tilde{y}$

$\Rightarrow \tilde{y} = H\hat{x} + \tilde{y} - \hat{y}$

$$\boxed{\tilde{y} = H\hat{x} + e}$$

meas. ($\tilde{y}$), pick $\hat{x}$ to minimize

$\frac{\partial J}{\partial \hat{x}} = 0$

$e = \tilde{y} - H\hat{x}$

$\min \frac{1}{2}\|e\|_2^2$

$\min_{\hat{x}} J = \frac{1}{2}(\tilde{y} - H\hat{x})^T(\tilde{y} - H\hat{x})$

$J = \frac{1}{2}(\tilde{y}^T\tilde{y} - 2\tilde{y}^T H\hat{x} + \hat{x}^T H^T H\hat{x})$

minimize by setting $\frac{\partial J}{\partial \hat{x}} = 0$

$$\frac{\partial J}{\partial \hat{x}} = -2\tilde{y}^T H + 2\hat{x}^T\left(\frac{1}{2}(H^TH)^T + \frac{1}{2}(H^TH)\right) = 0$$

$$= -2\tilde{y}^T H + 2\hat{x}^T(H^TH) = 0$$

$$\Rightarrow \hat{x}^T = \tilde{y}^T H(H^TH)^{-1} \Rightarrow \boxed{\hat{x} = (H^TH)^{-1}H^T\tilde{y}}$$

left inverse of H — should look familiar

$\tilde{y} \neq H\hat{x}$

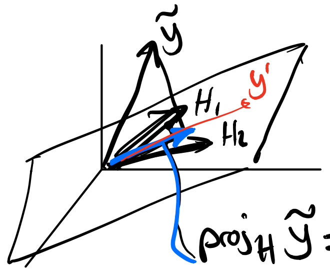
$= H(H^TH)^{-1}H^T\tilde{y}$

this is the projection of $\tilde{y}$ onto the range of H

$(H^TH)^{-1}H^T \times H = I$

$H(H^TH)^{-1}H^T\tilde{y}$ → is the identity restricted to the range of H

projection matrix



$= H = [H_1\ H_2]$

$\text{proj}_H \tilde{y} = H(H^TH)^{-1}H^T\tilde{y}$

Technically:

$\text{proj}_H(\tilde{y}) = \text{proj}_H(\text{proj}_H(\tilde{y}))$

left inverse: $(H^T H)^{-1} H^T$

right inverse: $H^T (H H^T)^{-1}$

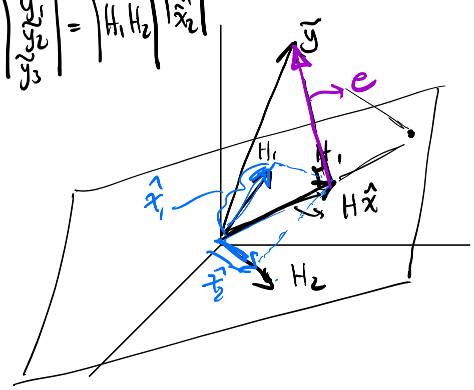
"a projection is a map that is the identity but only on a subspace"

$y' \in$ Range of H $\exists z$ s.t. $y' = Hz$ } writing y' in the basis of the cols of H

project y' onto range of H...

$$H(H^T H)^{-1} H^T y' = H(H^T H)^{-1} H^T H z$$
$$= Hz = y'$$

$$\begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \tilde{y}_3 \end{bmatrix} = \begin{bmatrix} H_1 & H_2 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}$$



$H = [H_1 \; H_2]$

want $H\hat{x}$ ← projection

$H\hat{x}$: projection of $\tilde{y}$ onto range of H

$$H(H^T H)^{-1} H^T \tilde{y}$$

$$\hat{x} = (H^T H)^{-1} H^T \tilde{y}$$

$H^T H$: needs to be invertible.
need cols of H to be lin ind.

if cols of H are lin dep $\Rightarrow \exists x \neq 0$ s.t. $Hx = 0$

intuition: H has a nontrivial nullspace

$Hx = 0$

$H = [H_1 \cdots H_n]$

$H_1 x_1 + \cdots + H_n x_n = 0$

two diff sets of parameters could give you the same meas.

$\Rightarrow$ parameters are not unique

$H(\hat{x} + \hat{n})$ $\quad \hat{n} \in \text{Null}(H)$

$H\hat{x} + H\hat{n}$ ($H\hat{n} = 0$)

won't be able to distinguish between $\hat{x}$ & $\hat{x} + \hat{n}$

sym ↓

$H^T H$ needs to be positive definite..

$x^T H^T H x > 0$

$\forall x > 0$

$\|Hx\|_2^2$

$Hx \neq 0$

if $H^T H$ is invertible then $H^T H$ is PD and all eigenvalues of $H^T H$ are > 0

Ex.

$$\begin{bmatrix} y_1 \\ \vdots \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} \sin(t_1) & 2\sin(t_1) & 3\sin(t_1) \\ \vdots & \vdots & \vdots \\ \sin(t_m) & 2\sin(t_m) & 3\sin(t_m) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \Longrightarrow H^T H \text{ not invertible}$$

(2nd and 3rd columns: scalar multiple of first col.)

NEWS FLASH: ROGER PENROSE NOBEL PRIZE IN PHYSICS

WORK ON BLACK HOLES

- Moore Penrose Pseudo Inverses
- Hawking-Penrose Singularity Theorems
- Penrose Tiling - aperiodic tiling
- Conformal Cyclic Cosmology
  - most nuts scifi cosmology idea.

full col rank $\Rightarrow$ left inverse $(H^T H)^{-1} H^T$

full row rank $\Rightarrow$ right inverse $H^T (H H^T)^{-1}$

neither are full rank $\Rightarrow$ Moore penrose pseudo inverse

$$H = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^*$$

$$H^+ = V \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^*$$

general pseudo inverse

Gauss: $\hat{x} = (H^T H)^{-1} H^T \tilde{y}$

Gaussian Elimination: row reduced this to an upper triangular

row reduction operations

$$E_i \times [\, H \mid I \,] \Rightarrow [\, E_i H \mid E_i \,]$$

$$\Rightarrow [\, E_k \cdots E_1 H \mid E_k \cdots E_1 \,]$$

✓

$$\underbrace{E_k \cdots E_1}_{H^{-1}} H = I$$

$$\boxed{H^{-1} = E_k \cdots E_1}$$

Ex. $\tilde{y}(t_i) = 0.3\sin(t_i) + 0.5\cos(t_i) + 0.1 t_i + b + v_i \quad i = 1, \ldots, m$ (m → 1000; v_i zero-mean)

$$H = \begin{bmatrix} \sin(t_1) & \cos(t_1) & t_1 & \cos(t_1)\sin(t_1) & t_1^2 & 1 \\ \sin(t_2) & \cos(t_2) & t_2 & \cos(t_2)\sin(t_2) & t_2^2 & 1 \\ \vdots & \vdots & & & & \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \hat{x}_3 \\ \hat{x}_4 \\ \hat{x}_5 \\ \hat{x}_6 \end{bmatrix}$$

1000 (rows)

$\hat{x}_1 \to \sim 0.3$
$\hat{x}_2 \to \sim 0.5$
$\hat{x}_3 \to \sim 0.1$
$\hat{x}_4 \to \sim 0.001$
$\hat{x}_5 \to \sim 0.002$
$\hat{x}_6 \to \sim b$

pretty common

$y = mz + b$

$$\begin{matrix} y_1 \\ \vdots \\ y_m \end{matrix} = \begin{bmatrix} z_1 & 1 \\ \vdots & \vdots \\ z_m & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix}$$

Ex: scaler dynamical system: $\dot{y} = Ay + Bu$

$\dot{y} = ay + bu \;\to$ meas of y over time.

↳ discrete time

$y_{k+1} = \Phi y_k + \Gamma u_k$

$\Phi = e^{a\Delta t}$

$\Gamma = \int_0^{\Delta t} b e^{at} dt = \frac{b}{a}(e^{a\Delta t} - 1)$

Find Φ, Γ

$$\begin{bmatrix} \tilde{y}_2 \\ \tilde{y}_3 \\ \vdots \\ \tilde{y}_m \end{bmatrix} = \underbrace{\begin{bmatrix} \tilde{y}_1 & u_1 \\ \tilde{y}_2 & u_2 \\ \vdots & \vdots \\ \tilde{y}_{m-1} & u_{m-1} \end{bmatrix}}_{H} \begin{bmatrix} \Phi \\ \Gamma \end{bmatrix} + \begin{bmatrix} e_2 \\ e_3 \\ \vdots \\ e_m \end{bmatrix} \Rightarrow \begin{bmatrix} \hat{\Phi} \\ \hat{\Gamma} \end{bmatrix} = (H^T H)^{-1} H^T \begin{bmatrix} \tilde{y}_2 \\ \vdots \\ \tilde{y}_m \end{bmatrix}$$

↑ y's shifted forward by 1

System Identification

Weighted Least Squares

$\Rightarrow$ trust some measurements more than others

before: $J = \frac{1}{2} e^T e$

now: $J = \frac{1}{2} e^T W e$ where W is a sym PD weighting matrix (positive definite)

modify w_i to trust some measurements more than others.

$\Rightarrow$ in general... pick W to be diagonal

$$W = \begin{bmatrix} w_1 & & 0 \\ & \ddots & \\ 0 & & w_m \end{bmatrix}$$

trust meas i : $w_i > 0 \uparrow$ large

don't trust meas j : $w_j > 0 \downarrow$ small

ex. $J = \frac{1}{2} e_1 w_1 e_1 + \frac{1}{2} e_2 w_2 e_2$ (w_1 large, w_2 small) } less penalty for larger errors in e_2

$$J = \frac{1}{2} e^T W e = \frac{1}{2} (\tilde{y} - H\hat{x})^T W (\tilde{y} - H\hat{x})$$

$$= \frac{1}{2} \left(\tilde{y}^T W \tilde{y} - 2 \tilde{y}^T W H \hat{x} + \hat{x}^T H^T W H \hat{x} \right)$$

$\Rightarrow$ WLS: $\hat{x} = (H^T W H)^{-1} H^T W \tilde{y}$

for perfect meas: w_i set really large...

Constrained Least Squares

$m_1 \,|\, \tilde{y}_1 = H_1 \hat{x} + e_1 \quad \leftarrow$ uncertain meas.

$m_2 \,|\, \tilde{y}_2 = H_2 \hat{x} \quad \leftarrow$ certain meas (n)

Note: knowing that x satisfys some linear constraints

$m_2 \left[Ax = b \right. \Rightarrow A\hat{x} = b \quad H_2 = A \quad \tilde{y}_2 = b$ ↖ robot arm

$m_1 + m_2 > n$

$\boxed{m_2 < n}$

if $m_2 = n$ $\hat{x}$ determined

$m_2 > n$ $\hat{x}$ overdetermined

$\boxed{m_2 < n}$ - if $n = 10$, then $m_2 = 5$ maybe

$\boxed{m_1 > n}$ → $m_1 = 100$ maybe

$$\min_{\hat{x}} \; J = \frac{1}{2} e_1^T W_1 e_1 = \frac{1}{2}(\tilde{y}_1 - H_1\hat{x})^T W_1 (\tilde{y}_1 - H_1\hat{x})$$

↑ objective

s.t. $\tilde{y}_2 = H_2 \hat{x}$ ⟸ constraints.

HOW DO WE SOLVE THIS: ..

Lagrange Multipliers:

Ex. $\min_x f(x)$ s.t. $g(x) = 0$

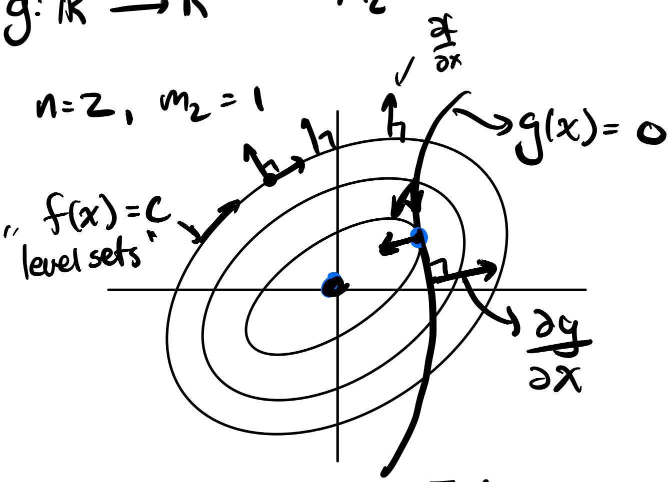
$f: \mathbb{R}^n \to \mathbb{R}$

$g: \mathbb{R}^n \to \mathbb{R}^{m_2}$ $\quad m_2 < n$

$$\mathcal{L} = f(x) - \lambda^T g(x)$$

λ^T: price you pay to violate a constraint

$n = 2$, $m_2 = 1$



Optimality conds:

before: $\frac{\partial f}{\partial x} = 0$

NOW:

$\frac{\partial \mathcal{L}}{\partial x} = 0$: stationarity condition

$\frac{\partial \mathcal{L}}{\partial \lambda} = 0$: feasibility condition

"pay no price for violating constraints

ie. don't violate constraints

Fact: $\frac{\partial f}{\partial x} \perp$ to level sets

$$\Delta f = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = 0$$

$0 = \frac{\partial f}{\partial x} \Delta x$

step along level set → $\Delta f = 0$

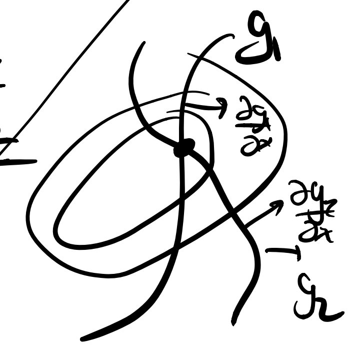
$\frac{\partial \mathcal{L}}{\partial \lambda} = 0$: $g(x) = 0$ } → make sure constraints are satisfied stationarity:

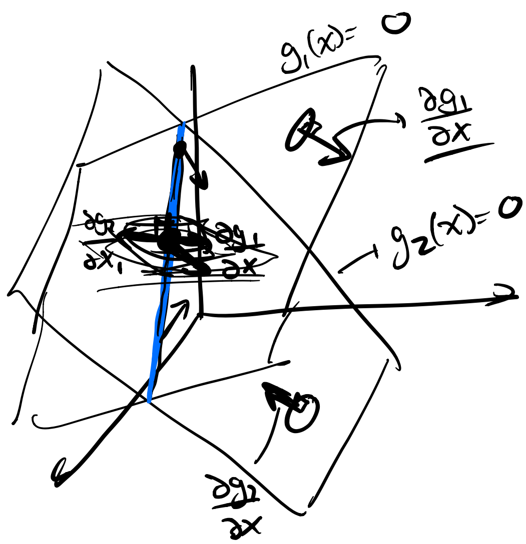
$$\frac{\partial \mathcal{L}}{\partial x} = 0 : \quad \frac{\partial f}{\partial x} - \lambda^T \frac{\partial g}{\partial x} = 0 \Rightarrow \boxed{\frac{\partial f}{\partial x} = \lambda^T \frac{\partial g}{\partial x}}$$

at minimum $\frac{\partial f}{\partial x}$ has to be a linear combination of derivative of constraints

if $g_1(x) = 0$
$g_2(x) = 0$
$\lambda^T = [\lambda_1, \lambda_2] \rightarrow$

$$\frac{\partial f}{\partial x} = \lambda_1 \frac{\partial g_1}{\partial x} + \lambda_2 \frac{\partial g_2}{\partial x_2}$$





$$\frac{\partial f}{\partial x} = \lambda^T \frac{\partial g}{\partial x}$$

"going down hill needs to take you ⊥ to the constraints"

$\frac{\partial f}{\partial x} = \lambda^T \frac{\partial g}{\partial x}$ } the down hill direction is ⊥ to the constraints

Full optimality conds:

$$\frac{\partial \mathcal{L}}{\partial x} = \frac{\partial f}{\partial x} - \lambda^T \frac{\partial g}{\partial x} = 0 \qquad \frac{\partial \mathcal{L}}{\partial \lambda} = g(x) = 0$$

λ: how much the constraints are pushing back against the objective function

$\frac{\partial f}{\partial x} = \lambda^T \frac{\partial g}{\partial x}$ if $\left|\frac{\partial f}{\partial x}\right|_2 \uparrow \rightarrow |\lambda| \uparrow$

FROM ABOVE:

$$\min_{\hat{x}} J = \frac{1}{2} e_1^T W_1 e_1 = \frac{1}{2}(\tilde{y}_1 - H_1\hat{x})^T W_1 (\tilde{y}_1 - H_1\hat{x}) \leftarrow$$

↑ objective

$$\text{s.t.} \rightarrow \tilde{y}_2 = H_2 \hat{x} \Leftarrow \text{constraints.}$$

HOW DO WE SOLVE THIS: ..

$$\mathcal{L} = \frac{1}{2}(\tilde{y}_1 - H_1\hat{x})^T W_1 (\tilde{y}_1 - H_1\hat{x}) - \lambda^T (H_2\hat{x} - \tilde{y}_2)$$

Solve for $\hat{x}$ & λ:

$$\frac{\partial \mathcal{L}}{\partial \hat{x}} = -\tilde{y}_1^T W_1 H_1 + \hat{x}^T H_1^T W_1 H_1 - \lambda^T H_2 = 0$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = H_2\hat{x} - \tilde{y}_2 = 0 \quad \} \rightarrow \text{constraint.}$$

→ isn't enough to solve for $\hat{x}$

$$\begin{bmatrix} H_2 \end{bmatrix} \hat{x} = \tilde{y}_2$$

$$\hat{x}^T = (\tilde{y}_1^T W_1 H_1 + \lambda^T H_2)(H_1^T W_1 H_1)^{-1}$$

needs to be invertible

$$\hat{x} = (H_1^T W_1 H_1)^{-1}(H_1^T W_1 \tilde{y}_1 + H_2^T \lambda) \Leftarrow (*)$$

$$H_2\hat{x} = H_2(H_1^T W_1 H_1)^{-1}(H_1^T W_1 \tilde{y}_1 + H_2^T \lambda) = \tilde{y}_2$$

$$H_2(H_1^T W_1 H_1)^{-1} H_2^T \lambda = \tilde{y}_2 - H_2(H_1^T W_1 H_1)^{-1} H_1^T W_1 \tilde{y}_1$$

want to invert

|

should be invertible...
if not → redundant constraints.

$$\underline{\lambda} = \left(H_2 (H_1^T W_1 H_1)^{-1} H_2^T \right)^{-1} \left(\tilde{y}_2 - H_2 (H_1^T W_1 H_1)^{-1} H_1^T W_1 \tilde{y}_1 \right)$$

plug into (★) and get finally...

page 17 book

$$\hat{x} = \bar{x} + K(\tilde{y}_2 - H_2 \bar{x})$$

where → $K = (H_1^T W_1 H_1)^{-1} H_2^T \left[H_2 (H_1^T W_1 H_1)^{-1} H_2^T \right]^{-1}$

→ $\underline{\bar{x}} = (H_1^T W_1 H_1)^{-1} H_1^T W_1 \tilde{y}_1$

$\bar{x}$: unconstrained least squares solution

K: "gain matrix"

multiply by → $\underline{K} (\tilde{y}_2 - H_2 \bar{x})$ ← how much you violate constraints

K → gain matrix

$(\tilde{y}_2 - H_2 \bar{x})$: how much does the unconstrained soln violate the constraints

$$\hat{x} = \bar{x} + K(\tilde{y}_2 - H_2 \bar{x})$$

(Woodbury Matrix Identity)

up thru 1.2 book up to page 19.