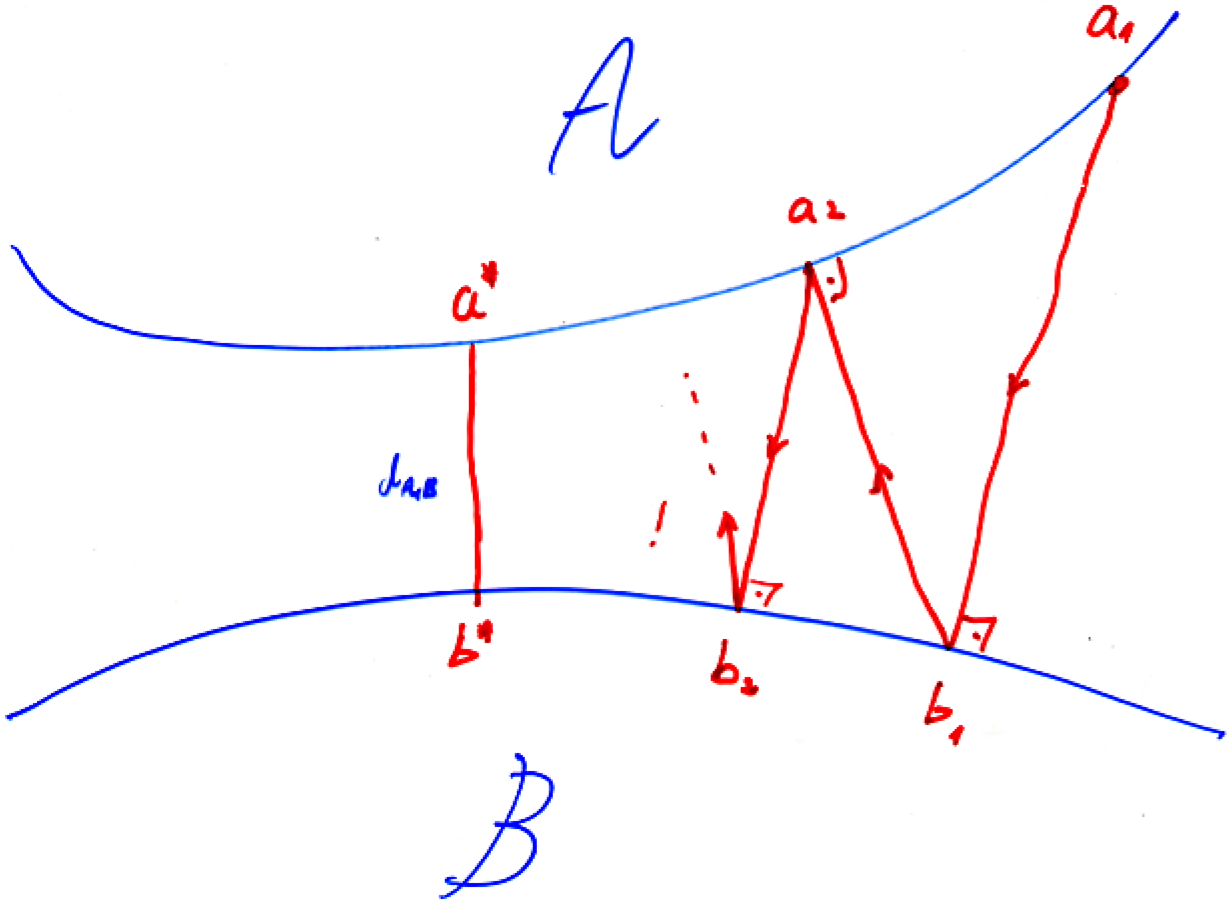
alternating Projections



$$a^*, b^* = \arg\min_{a \in A,\, b \in B} \|a - b\| \quad \text{achieve } d_{A,B}$$

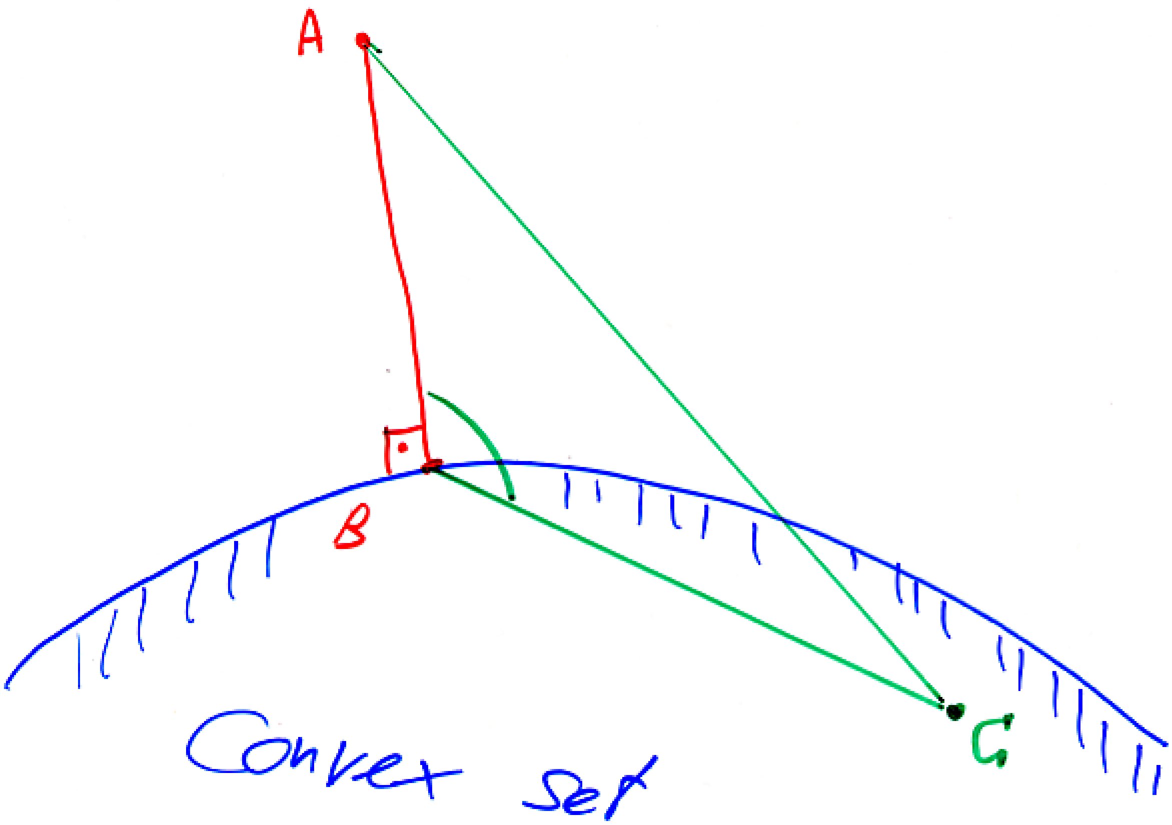
If A and B are convex, ~~then~~

$$a_1 \to b_1 \to a_2 \to b_2 \to \dots \qquad a_{i+1} = \arg\min_{a \in A} \|a - b_i\|$$

$$b_i = \arg\min_{b \in B} \|a_i - b\|$$

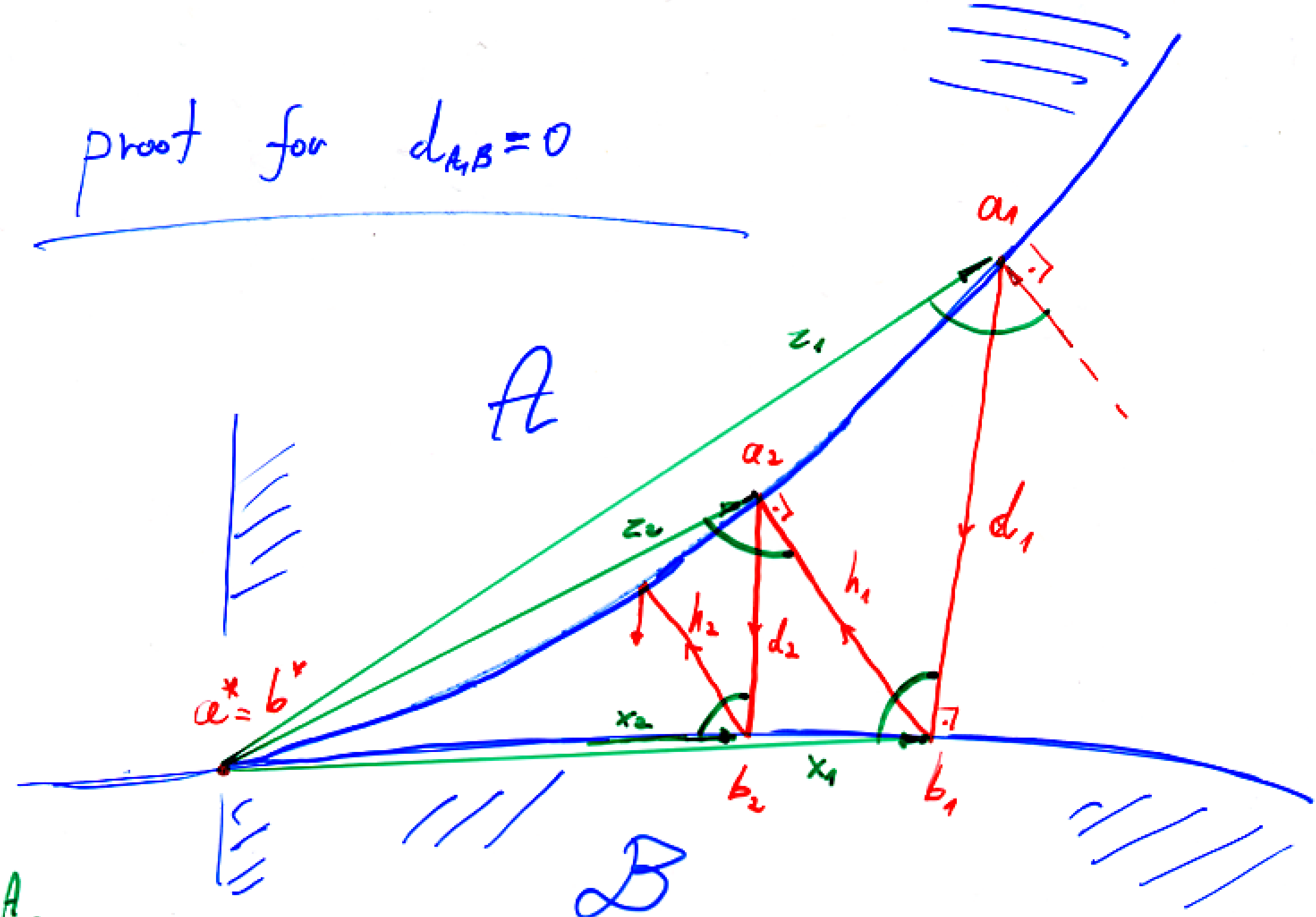
Then $\Rightarrow$ $a_i \to a^*$ $\quad b_i \to b^*$

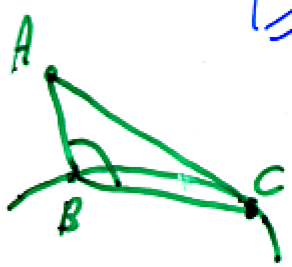
* Pythagoras Inequality *



$$\sphericalangle > 90^\circ \implies AB^2 + BC^2 \leq AC^2$$

proof for $d_{A,B}=0$





convexity of sets $\Rightarrow \measuredangle > 90°$

$\Rightarrow AC^2 > AB^2 + BC^2$

"Pythagoras Inequality"

$\therefore$

$$z_1^2 \geq d_1^2 + x_1^2$$
$$x_1^2 \geq h_1^2 + z_2^2$$
$$z_2^2 \geq d_2^2 + x_2^2$$
$$\vdots$$

$$\therefore \quad z_1^2 \geq z_{i+1}^2 + d_i^2 + h_i^2 + d_{i-1}^2 + \dots + h_1^2 + d_1^2 \quad \forall i$$

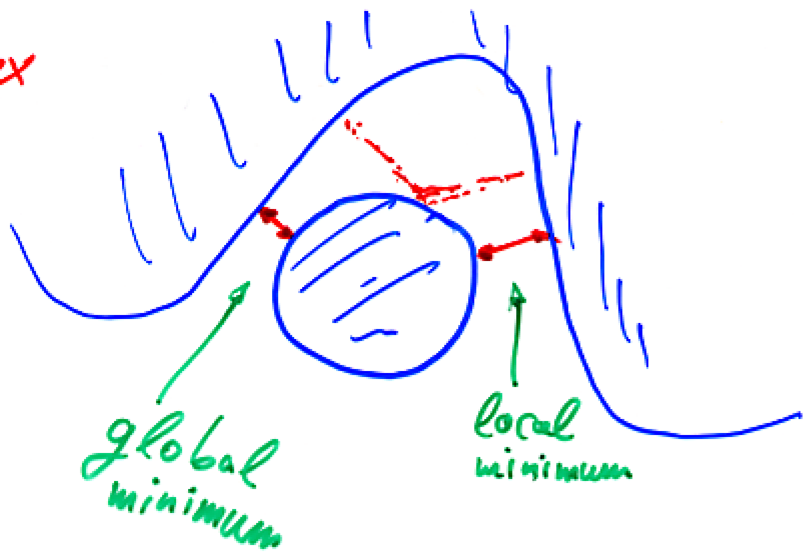
assume $\underline{\underline{z_1 < \infty}}$

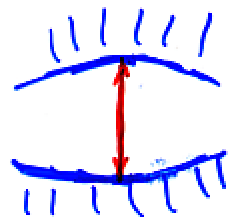
$$\therefore \quad \sum_{i=1}^{\infty} h_i^2 + d_i^2 \leq z_1^2$$

$\therefore$ (non-negative series) $h_i, d_i \to 0 \quad \xRightarrow[a_i \to a^*;\, b_i \to b^*]{\text{(unique minimum)}} z_i \to 0$

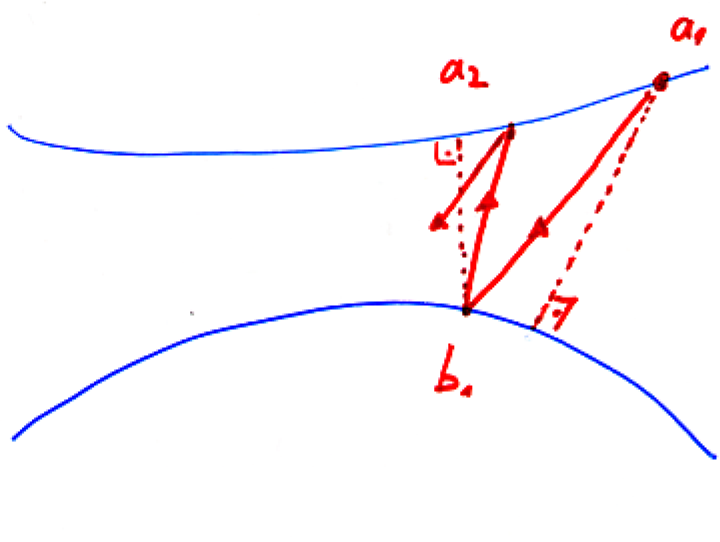
Bad Cases ...

1) non convex sets

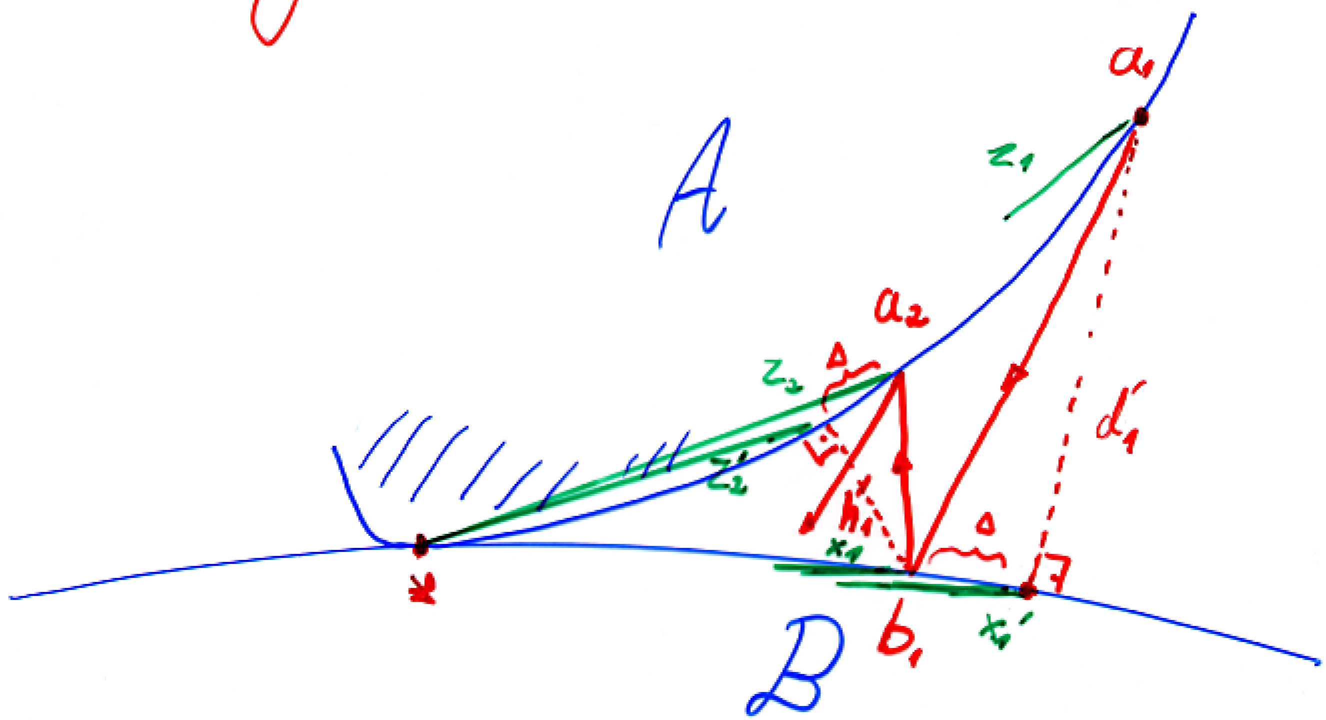




2) non accurate projections



Noisy (non accurate) Projections



d_i', h_i' = true (ideal) projection distance

z_i, x_i = actual distance from optimum $*$

z_i', x_i' = distance of true projection " " "

Suppose $|x_i^2 - x_i'^2| < \Delta^2$, $|z_i^2 - z_i'^2| < \Delta^2$

$\boxed{1}$ $\Rightarrow$ $$z_1^2 \geq \sum_{i=1}^{m} \left(d_i^2 + h_i^2 - \Delta^2\right) \quad \forall m$$

$\Rightarrow$ $$\liminf_i \left[d_i^2 + h_i^2 - \Delta^2\right] \leq 0$$

gradient descent in L.M.S.

$\boxed{2}$ "Black Hole : $A_\infty(\Delta)$: $a_i \in A_\infty(\Delta) \Rightarrow a_{i+1} \in A_\infty(\Delta)$

Aspects to think about....

* desired optimum $\stackrel{?}{=}$ fixed point
  $\stackrel{?}{=}$ stable point (stability region)
  $\stackrel{?}{=}$ unique fixed point

* projection rule (metric) vrs. true optimality criterion

* physical meaning of intermediate quantites:

$$a_1 \rightarrow b_1 \rightarrow a_2 \rightarrow b_2 = ?$$

$$z_i,\ x_i,\ h_i,\ d_i = ?$$

Divergence "Pythagoras Theorem"

If

E = closed and convex set of probability distributions

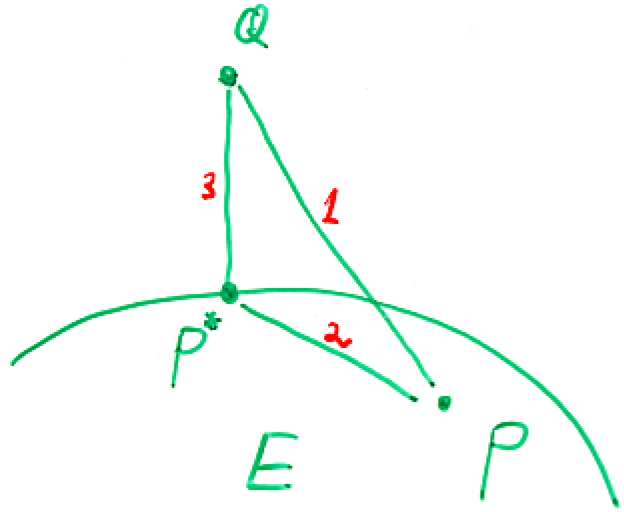
$Q \notin E$

$P^* \in E$ is closest to Q in divergence sense

$$D(P^* \| Q) = \min_{P \in E} D(P \| Q)$$

Then for any $P \in E$

$$\underset{1}{D(P\|Q)} \geq \underset{2}{D(P\|P^*)} + \underset{3}{D(P^*\|Q)}$$



Recall that for $P \geq 0$, $Q \geq 0$

$$D(P\|Q) \triangleq \sum_x P(x) \log \frac{P(x)}{Q(x)}$$

— If P, Q are prob. measures* ($\sum_x P = \sum_x Q = 1$) $\Rightarrow$
$D(P\|Q) \geq 0$ with eq iff $P \equiv Q$

— not a metric!

* $D(P\|Q) \neq D(Q\|P)$ not symmetric

* no triangle inequality

— $D(P\|Q) \leq \log 1/Q_{min}$ $\Rightarrow$ bounded if $Q(x) > 0$ $\forall x$

* if $\sum P \neq 1$, $\sum Q \neq 1$ Then "modified pythagoras" hold.

Csiszar-Tusnady Theorem

$\mathcal{P}, \mathcal{Q}$ convex closed sets of measures (not necessarily of unit sum)

$Q_0(x) > 0 \quad \forall x \in \mathcal{X} \qquad |\mathcal{X}| < \infty$

$$P_n = \arg\min_{P \in \mathcal{P}} D(P \| Q_{n-1}) \qquad n = 1, 2, \ldots$$

$$Q_n = \arg\min_{Q \in \mathcal{Q}} D(P_n \| Q) \qquad n = 1, 2, \ldots$$

i.e., $Q_0 \to P_1 \to Q_1 \to P_2 \to \ldots$

sequence of alternating minimizations

Then

$$\lim_{n \to \infty} D(P_n \| Q_n) = D(\mathcal{P} \| \mathcal{Q}) \triangleq \min_{P \in \mathcal{P}, \, Q \in \mathcal{Q}} D(P \| Q)$$

$$\triangleq D(P^* \| Q^*)$$

a minimizing pair

$P_n \to P^*, \quad Q_n \to Q^*$

E.M.

$$\max_{\theta} f(\underline{y} \mid \theta) \quad \text{given } \underline{y} = y_1 \ldots y_n$$

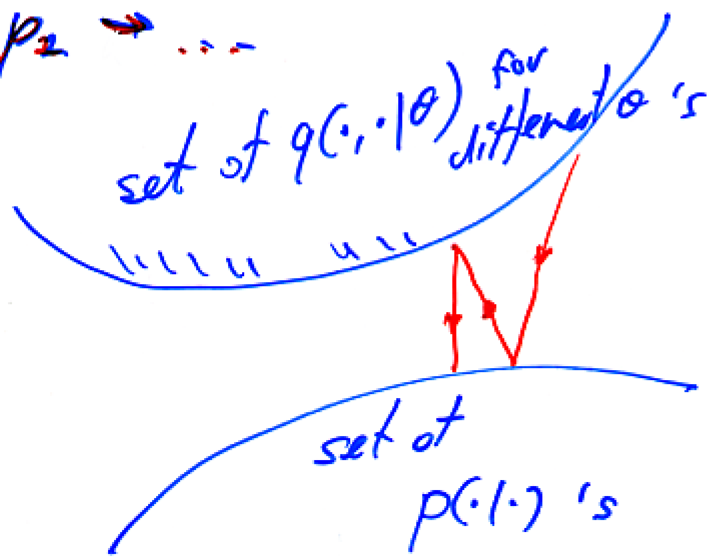
$f(\underline{y} \mid \theta) = \int q(\underline{x}, \underline{y} \mid \theta)\, d\underline{x}$

$$\Longleftrightarrow \max_{p(\underline{x})} \max_{\theta} \int p(\underline{x}) \log \frac{q(\underline{x}, \underline{y} \mid \theta)}{p(\underline{x})}\, dx$$

marginalizing $p(\cdot)$ can only increase likelihood

$$\Longleftrightarrow \max_{p(x \mid y)} \max_{\theta} -E_{\underbrace{T_{\underline{y}}(y)}_{\text{empirical measure of } \underline{y}}} D\big(p(\cdot \mid y) \,\|\, q(\cdot, y \mid \theta)\big)$$

$$\theta_0 \to p_1 \to \theta_1 \to p_2 \to \ldots$$



converges if $q(\cdot,\cdot \mid \theta)$ is convex set w.r.t. θ.

R-D-Computation

$$R(D) = \min_{W:\ E_{p,W}\, d(X,Y) \le D} I(p,W) \qquad p = p(x), \quad W = W(y|x)$$

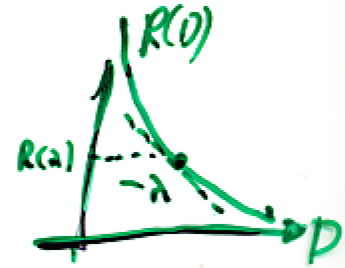
= the minimum rate in compressing a source $X \sim p$ with distortion $\le D$ w.r.t. distortion measure $d(\cdot,\cdot)$

$$R(D) = I(p,W^*) \ : \ Q^* = [p \circ W^*]_y \text{ is the optimum reproduction}$$

= codebook distribution

unconstrained version

$$R(\lambda) \triangleq \min_W \underbrace{\Big[I(p,W) + \lambda \cdot \underbrace{d(p,W)}_{E_{p,W}\, d(X,Y)} \Big]}_{D\left(p\circ W \,\|\, p \times \underbrace{[p\circ W]_y}_{Q(y)} \cdot e^{-\lambda d(x,y)}\right)}$$



$$\underbrace{[p\circ W]_y \cdot e^{-\lambda d(x,y)}}_{\tilde{Q}(x,y)}$$

Blahut A. for $R(\lambda)$

$$\Longleftrightarrow \quad \min_W \min_Q D\left(p \circ W_{(x,y)} \,\|\, p_{(x)} \times Q_{(y)} \cdot e^{-\lambda d(x,y)}\right)$$

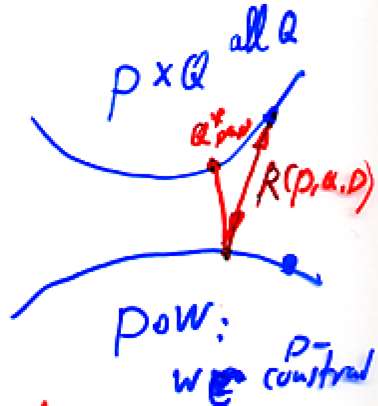
$$Q_0 \to \underbrace{W_1}_{\text{easy}} \to Q_1 \to \underbrace{W_2}_{\text{easy}} \to Q_2 \to \dots$$

physical meaning

constrained problem :

$$R(D) = \min_{W:\, d(p,W) \le D} \; \min_{Q} \; D(p \circ W \,\|\, p \times Q)$$

Define $\quad R(p,Q,D) = \min_{W:\, d(p,W) \le D} D(p \circ W \,\|\, p \times Q)$



$$Q^*(p,Q,D) = [p \circ W^*]_y$$

$R(p,Q,D)$ = rate of mismatched codebook

$Q^*(p,Q,D)$ = limiting $\overset{\text{w.p. 1}}{\vee}$ emprical distribution of D-matching codeword.

$$\boxed{\begin{matrix}\text{code book}\\ Q_0\end{matrix}} \rightarrow \underset{\downarrow \underline{x}}{\underline{y}} \rightarrow \underbrace{Q^*(p,Q_0,D)}_{= Q_1} \rightarrow \boxed{\begin{matrix}\text{code}\\ Q_1\end{matrix}} \rightarrow \underset{\downarrow \underline{x}'}{\underline{y}'} \rightarrow Q^*(p,Q_1,D) \cdots$$

$$\Longrightarrow Q^*_{opt}$$

* practical implementation using parametric codebook...

C - computation

$$C = \max_{P} I(p, W)$$

= maximum rate in transmission over channel in W.

$C = I(p^*, W)$: p^* is the optimum input= codebook distribution.

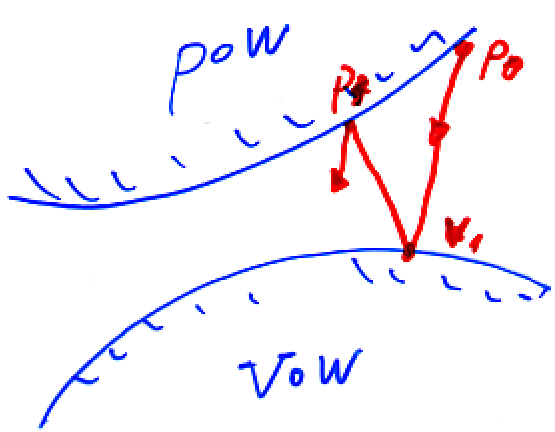
Arimoto-Blahut for C:

$$\Leftrightarrow \quad C = \max_{P} D\big(p \circ W \,\|\, p \times \underbrace{[p \circ W]_y}_{Q(y)}\big)$$

$$= \max_{P} \underbrace{\max_{V} - D(p \circ W \,\|\, V \circ W)}_{\max = I(p,W)}$$

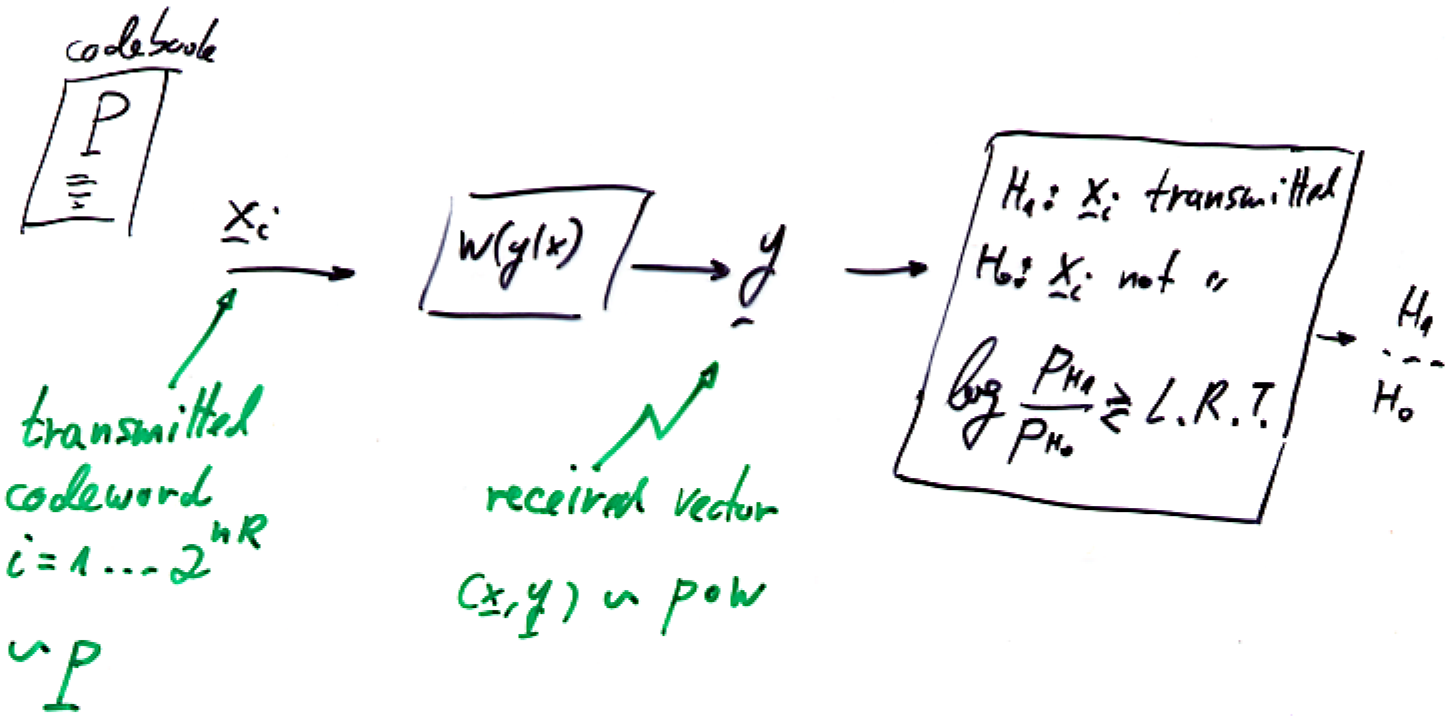
$V(x|y)$

$$\underbrace{P_0 \to V_1}_{\substack{\text{easy} \\ \text{(induced backward)}}} \to \underbrace{P_1 \to V_2}_{\text{easy}} \to \ldots$$



Physical Meaning ?

Q: Does $\underbrace{-D(p\circ W \,\|\, V\circ W)}_{\sum_x \sum_y p(x)W(y|x)\log \frac{V(x|y)}{p(x)}}$ represent an empirically measurable quantity?



* suppose receiver estimates backward channel as $V(x|y)$

$$\Longrightarrow \quad L.R. = \frac{1}{n}\sum_{j=1}^{n} \log \frac{V(x_{ij}\,|\,y_j)}{p(x_j)} \simeq \sum_x \sum_y p(x)W(y|x)\log\frac{V(x|y)}{p(x)}$$

with high prob.

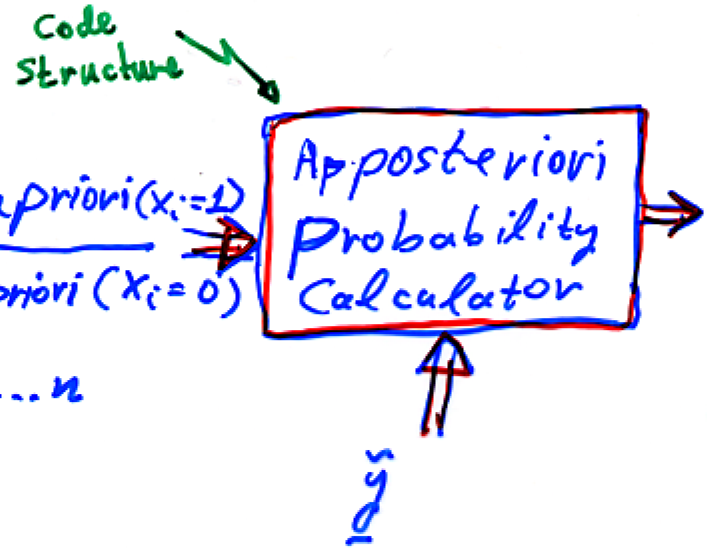
Can we realize / simulate $p_0 \to V_1 \to p_1 \to V_2 \to \ldots.$?

alternating optimizations :

geometric meaning	physical meaning	
✓	✓	EM
✓	✓	R-D
✓	?	C
?	✓	turbo-decoding

* Turbo Decoding *

$$\ell x_i = \frac{\text{apriori}(x_i=1)}{\text{apriori}(x_i=0)} \qquad i=1\ldots n$$



$$\underbrace{\log \frac{\sum_{\underline{x}:x_i=1} p(\underline{\tilde{y}}|\underline{x})\cdot \text{apri}(\underline{x})}{\sum_{\underline{x}:x_i=0} p(\underline{\tilde{y}}|\underline{x})\cdot \text{apri}(\underline{x})}}_{LLR_i}$$

Define : $\underline{LLR} = D(\ell\underline{x}, \underline{\tilde{y}})$

Iterative decoding (recursive LLR computation):

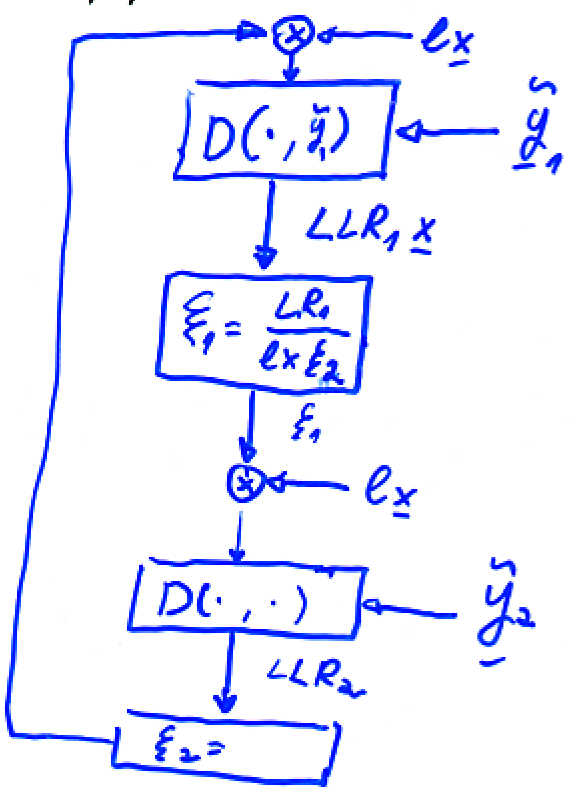
$$LLR_{1i}^{(k)} = D(\ell\underline{x}\,\xi_{2i}^{(k-1)}, \underline{\tilde{y}}_1) \triangleq \ell x_i \cdot \xi_{1i}^{(k)} \cdot \xi_{2i}^{(k-1)}$$

$$LLR_{2i}^{(k)} = D(\ell\underline{x}\,\xi_{1i}^{(k)}, \underline{\tilde{y}}_2) \triangleq \ell x_i \cdot \xi_{1i}^{(k)} \cdot \xi_{2i}^{(k)}$$

$i=1\ldots n \qquad k=1,2,3,\ldots$

extrinsic info for next iteration ("LLR amplification")

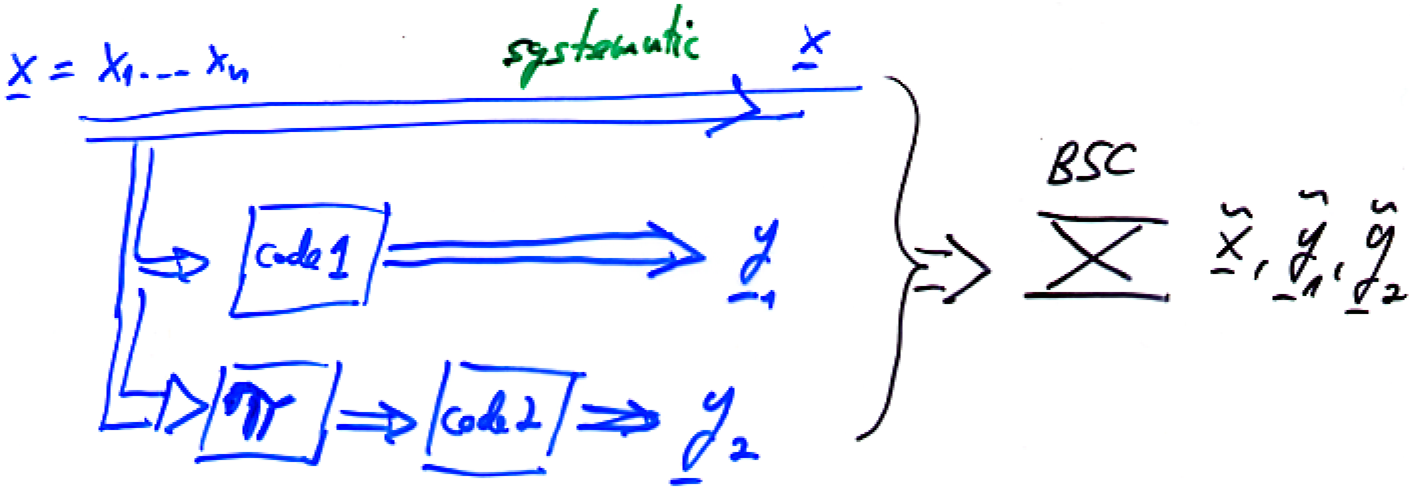
$\xi_2^{(0)} \equiv 1$



steady state (fixed point)

$LLR_1 = LLR_2 = \text{constant}$

Turbo Code



$$p(\tilde{\underline{x}}, \tilde{\underline{y}}_1, \tilde{\underline{y}}_2 \,|\, \underline{x}) = p(\tilde{\underline{x}}|\underline{x}) \cdot p(\tilde{\underline{y}}_1|\underline{x}) \cdot p(\tilde{\underline{y}}_2|\underline{x})$$

memoryless $\Rightarrow \tilde{\underline{y}}_1 \leftrightarrow \underline{x} \leftrightarrow \tilde{\underline{y}}_2$

$\underline{y}_i = \text{func}(\underline{x})$

optimal decoding:

1) sequence-wise: $\hat{\underline{x}}_{M.L.} = \arg\max_{\underline{x}} p(\tilde{\underline{x}}, \tilde{\underline{y}}_1, \tilde{\underline{y}}_2 \,|\, \underline{x})$

2) bit-wise: $\hat{x}_{i\,M.L.} = \arg\max_{x_i} p(\tilde{\underline{x}}, \tilde{\underline{y}}_1, \tilde{\underline{y}}_2 \,|\, x_i)$ for $i = 1 \dots n$

$$\Longleftrightarrow \underbrace{\log \frac{\sum_{\underline{x}: x_i = 1} p(\tilde{\underline{x}}, \tilde{\underline{y}}_1, \tilde{\underline{y}}_2 \,|\, \underline{x})}{\sum_{\underline{x}: x_i = 0} p(\; -''- \;)}}_{L.L.R._i} \;\underset{\hat{x}_i = 0}{\overset{\hat{x}_i = 1}{\gtrless}}\; 0$$

Q: when $\hat{x}_{i\,M.L.} \neq (\hat{\underline{x}}_{M.L.})_i$? ...

Abstraction: Projections on Product Measures

n-length real vector $\rightarrow$ $p_0(\underline{x}) \triangleq \sum_{i=1}^{n} \log p(\tilde{x}_i \mid x_i)$ = "product measure"

2^n-1 length real vectors $\rightarrow$ $p_1(\underline{x}) \triangleq \log p(\tilde{\underline{y}}_1 \mid \underline{x})$ (non product)

$p_2(\underline{x}) \triangleq \log p(\tilde{\underline{y}}_2 \mid \underline{x})$ (non product)

*marginalization:

$$\pi(p(\underline{x})) \triangleq \Big[\log \underbrace{\sum_{x_1 \ldots x_{i-1} x_{i+1} \ldots x_n} \exp p(\underline{x})}_{p(x_i)} \Big]_{i=1}^{n}$$

= the induced product ("memoryless") measure

M.L:

$$\Rightarrow \quad \hat{x}_{ML \cdot i} = \arg\max_{x \in \{0,1\}} \pi(p_0 + p_1 + p_2)_i(x)$$

Turbo:

$\underbrace{Q_1(\underline{x}), \; Q_2(\underline{x})}_{\text{product measures}}$ = "log-likelihood amplification" measures $\left(\xi_{1i} = \dfrac{\exp Q_{1i}(1)}{\exp Q_{1i}(0)} \right)$

$$Q_2^{(0)} \equiv 1 \qquad \boxed{\begin{aligned} Q_1^{(k)} &= \pi\left(p_0 + Q_2^{(k-1)} + p_1\right) - \left(p_0 + Q_2^{(k-1)}\right) \\ Q_2^{(k)} &= \pi\left(p_0 + Q_1^{(k)} + p_2\right) - \left(p_0 + Q_1^{(k)}\right) \end{aligned}}$$

$k = 1, 2, 3, \ldots$

$$LLR^{(k)} = \pi\left(p_0 + Q_2^{(k+1)} + p_1\right) \quad \text{or} \quad \pi\left(p_0 + Q_1^{(k)} + p_2\right)$$

Fixed point of Turbo iteration

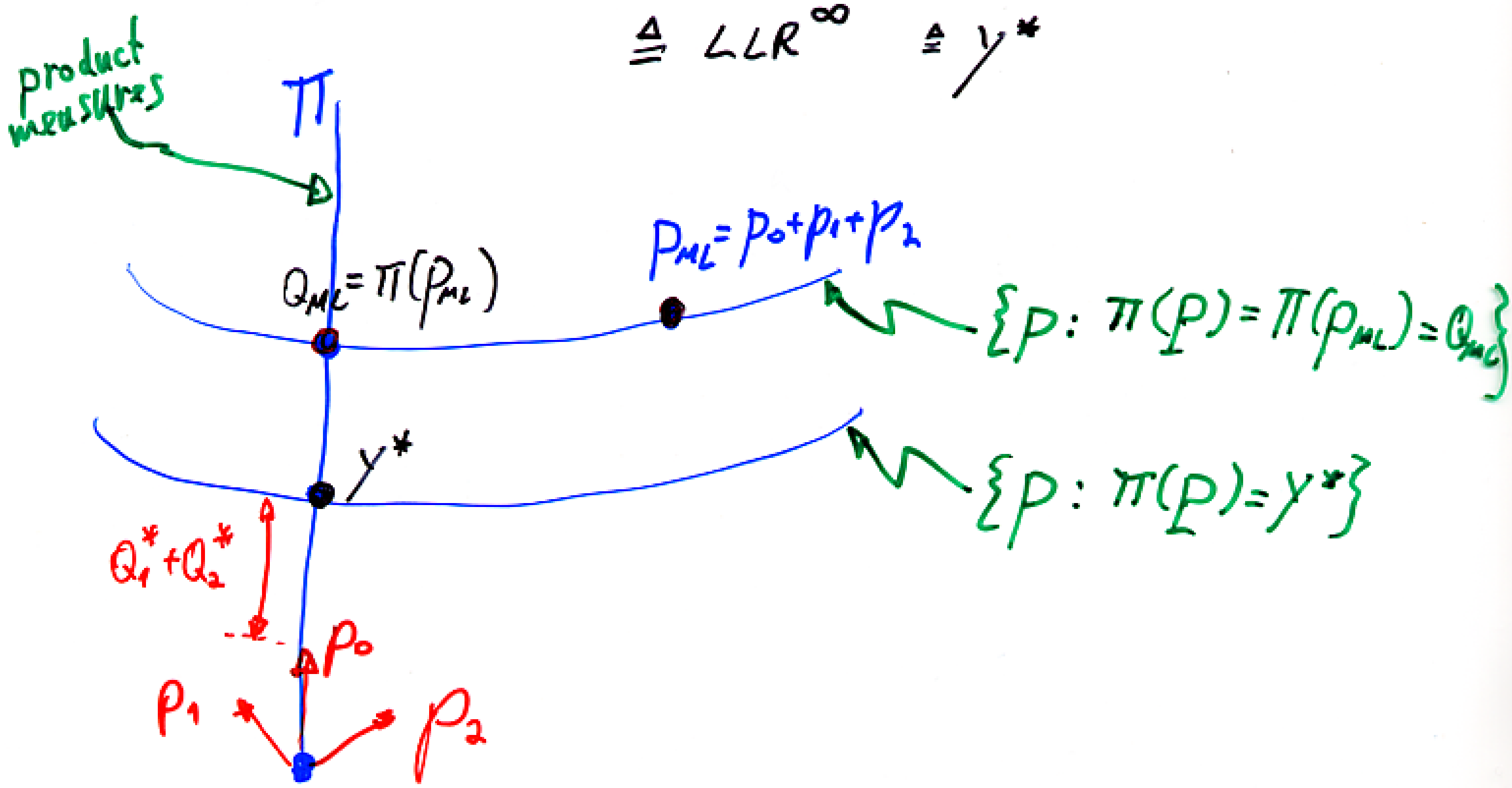
Given p_0, p_1, p_2 (log likelihood measures of $\underline{x}$)

Find product measures Q_1, Q_2 ($\Leftrightarrow$ find two vectors in $\mathbb{R}^n$)
such that iteration does not change their values:

$$Q_1 = \pi(p_0 + Q_2 + p_1) - (p_0 + Q_2)$$
$$Q_2 = \pi(p_0 + Q_1 + p_2) - (p_0 + Q_1)$$

2n equations with 2n unknowns... $\Rightarrow Q_1^*, Q_2^*$

$$\Rightarrow \quad \pi(p_0 + Q_2^* + p_1) = \pi(p_0 + Q_1^* + p_2) = p_0 + Q_1^* + Q_2^*$$
$$\triangleq LLR^\infty \triangleq y^*$$

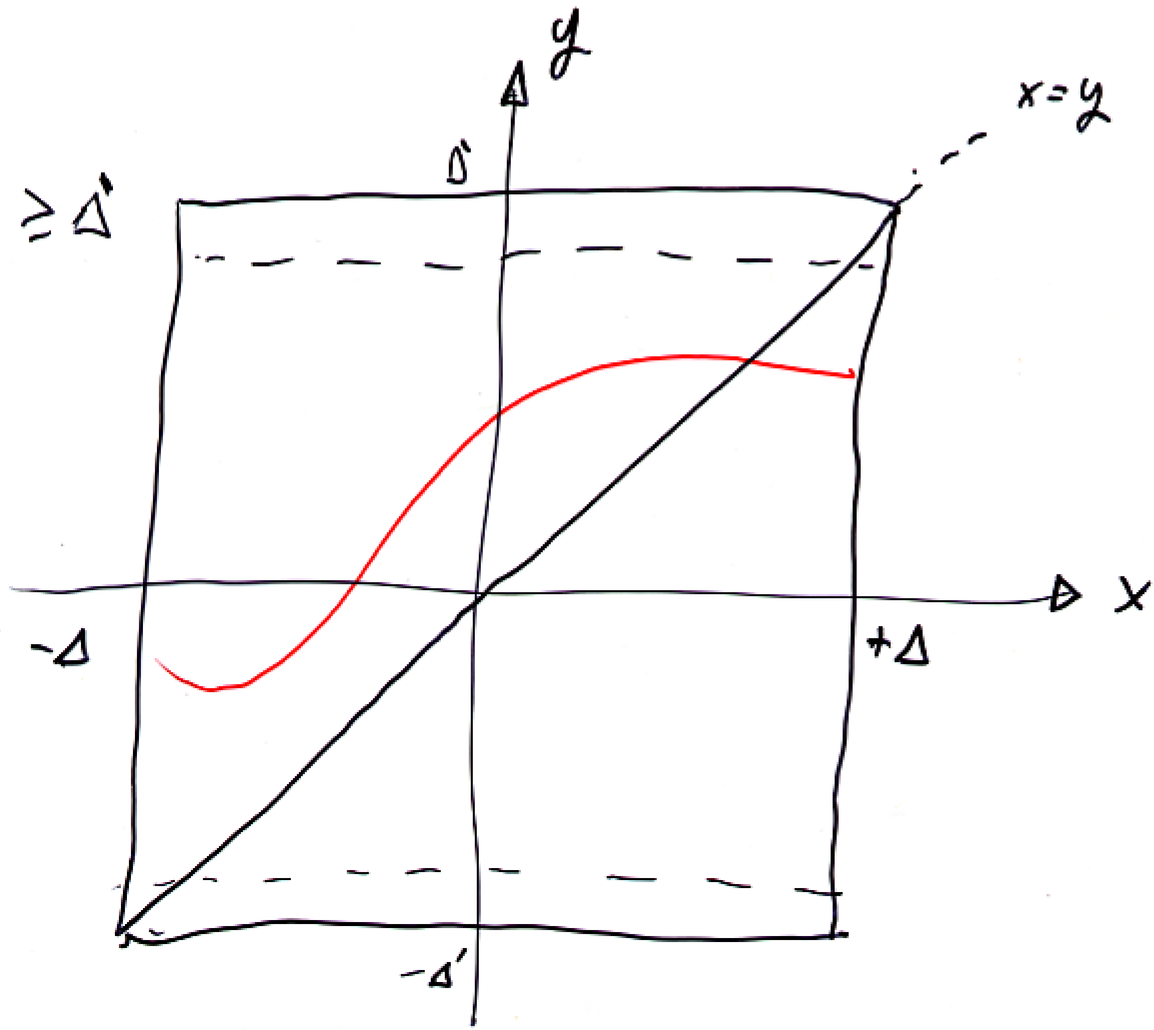


How do we know that Q_1^*, Q_2^* exist?

$F : S \to S'$ continuous function

S set (ball) in $\mathbb{R}^n$; $S' \subseteq S$

$\Delta \geq \Delta'$



Brouwer fixed point theorem : $\exists\, x \in S'$

s.t. $F(x) = x$