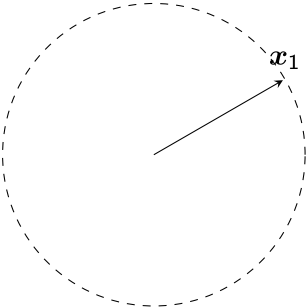
Minimax Algorithm for Learning Rotations

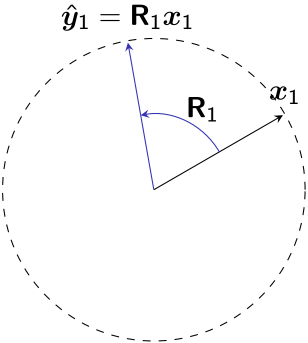
Wojciech Kotłowski
CWI
Manfred K. Warmuth
UCSC

Open problem at COLT 2011
July 9, 2011

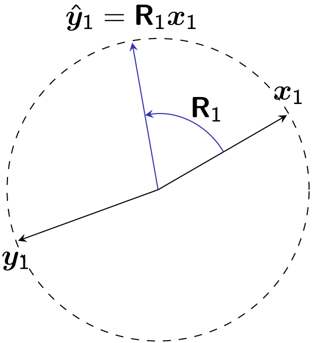
On-line Protocol



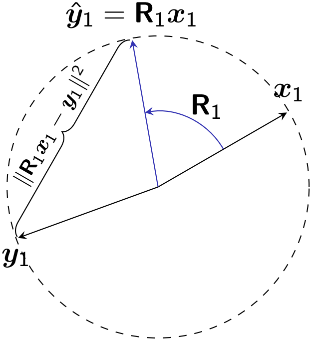
On-line Protocol



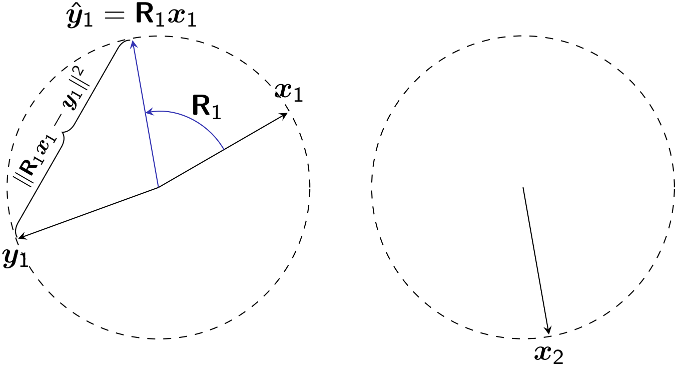
On-line Protocol



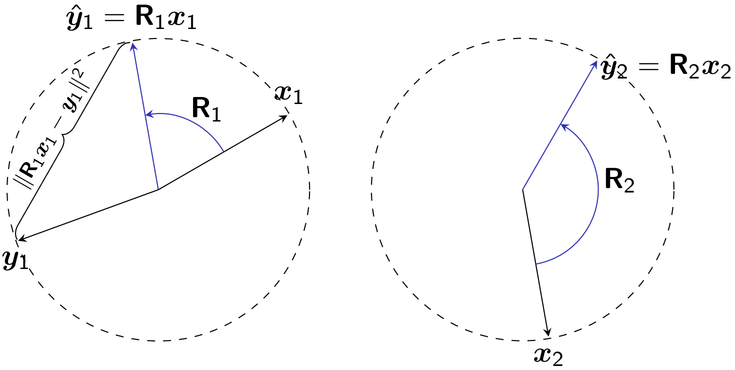
On-line Protocol



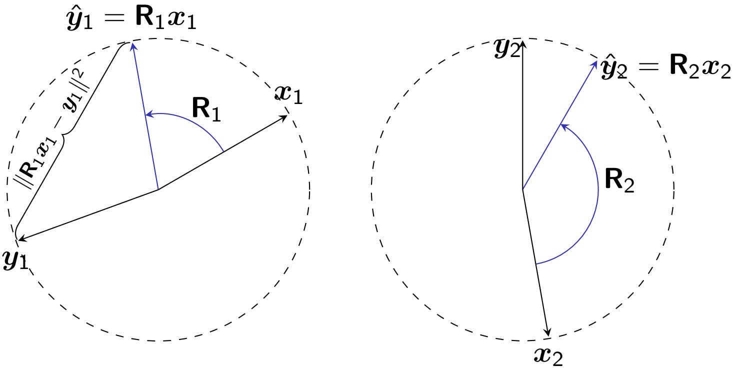
On-line Protocol



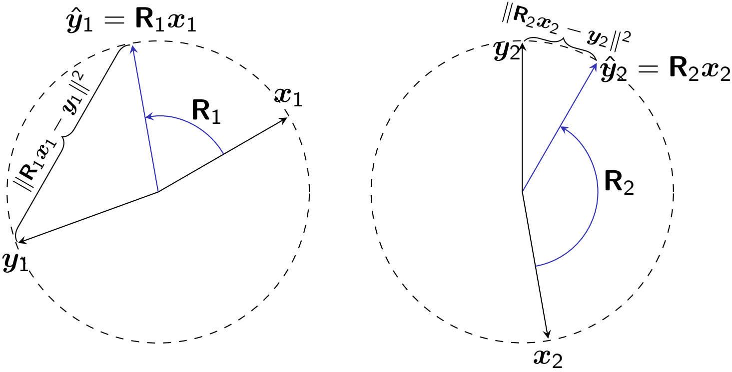
On-line Protocol



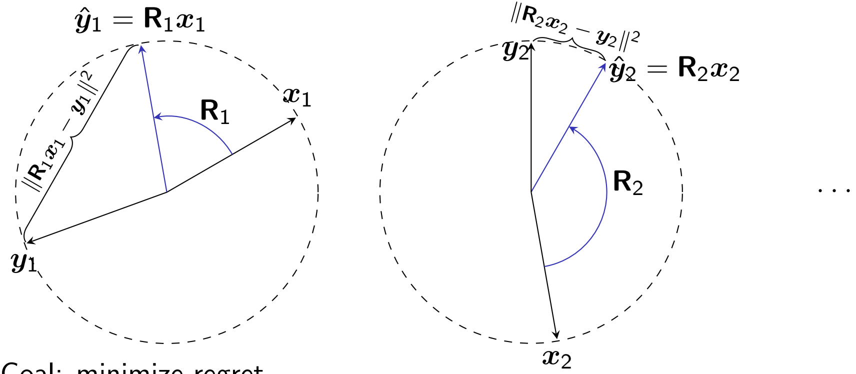
On-line Protocol



On-line Protocol



On-line Protocol



Goal: minimize regret

$$\sum_{t=1}^{T} L_t(\mathbf{R}_t) - \min_{\mathbf{R} \in \mathcal{SO}(n)} \sum_{t=1}^{T} L_t(\mathbf{R})$$

Applications in vision, robotics, matrix completion, subspace tracking, quantum physics, etc. [Arora 2009, Hazan et al. 2010]

Problem Setting

- Set of rotation matrices $\mathcal{SO}(n)$ is *non-convex*, but is a *Lie group* equipped with *Lie algebra*
- Loss is linear:

$$\begin{aligned} L_t(\mathbf{R}_t) &= \frac{1}{2}\mathbb{E}\left[\|\mathbf{R}_t \boldsymbol{x}_t - \boldsymbol{y}_t\|^2\right] \\ &= \frac{1}{2}\mathbb{E}[\underbrace{\|\mathbf{R}_t \boldsymbol{x}_t\|^2}_{1} + \underbrace{\|\boldsymbol{y}_t\|^2}_{1} - 2(\mathbf{R}_t \boldsymbol{x}_t) \cdot \boldsymbol{y}_t] \\ &= 1 - (\mathbb{E}[\mathbf{R}_t]\boldsymbol{x}_t) \cdot \boldsymbol{y}_t \end{aligned}$$

- Batch problem is known as *Wahba's problem*
  Solvable with SVD

Previous Work

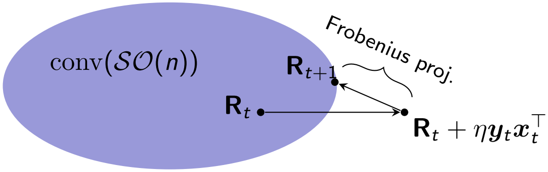
- [Arora, NIPS '09]: *Matrix exponentiated gradient algorithm*
  - exploits Lie group/algebra structure
  - deterministic
  - problems with the motivation

  [Hazan et al., COLT '10]:
  - Any deterministic algorithm $\Rightarrow$ can be forced to have regret $\Omega(T)$

Previous Work

- [Arora, NIPS '09]: *Matrix exponentiated gradient algorithm*
  - exploits Lie group/algebra structure
  - deterministic
  - problems with the motivation

  [Hazan et al., COLT '10]:
  - Any deterministic algorithm $\Rightarrow$ can be forced to have regret $\Omega(T)$
  - Randomized gradient descent:
    - Frobenius norm regularization with projection.
    - $2\sqrt{nT}$ regret bound.



Problem Statement

Is the problem already solved?

- Is Frobenious norm proper way to regularized rotations?
- Is there a notion of entropy defined over $\mathcal{SO}(n)$?
- Elegant Lie group and Lie algebra structure not exploited.
- For gradient descent, the constant in front of the $O(\sqrt{T})$ is not optimal (for $n = 2$, $2\sqrt{2T}$ instead of $\sqrt{T}$).

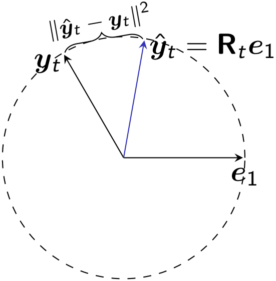
We hope to shed some light on these issues by finding the minimax algorithm for learning rotations

Partial Results

- If the instances $\boldsymbol{x}_t$ are restricted to be a fixed unit vector, say $\boldsymbol{e}_1$,
  then we can give the minimax algorithm:
  - Enough to predict with $\hat{\boldsymbol{y}}_t$, s.t. $\|\hat{\boldsymbol{y}}_t\| \leq 1$
  - The minimax prediction:

$$\hat{\boldsymbol{y}}_t = \frac{\boldsymbol{s}_{t-1}}{\sqrt{\|\boldsymbol{s}_{t-1}\|^2 + T - t + 1}},$$

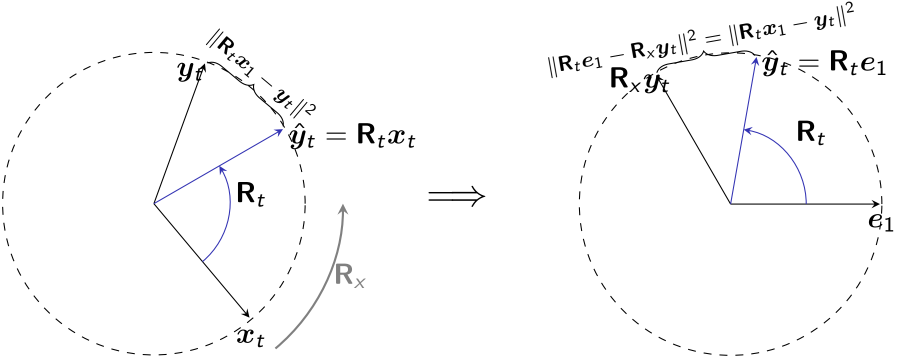
  where $\boldsymbol{s}_{t-1} = \sum_{q=1}^{t-1} \boldsymbol{y}_q$



Minimax regret: $\sqrt{T}$.

Partial Results

For $n = 2$, a simpler problem is equivalent to learning rotations.



$$\|\mathbf{R}_t e_1 - \mathbf{R}_x y_t\|^2 = \|\mathbf{R}_t \mathbf{R}_x x_t - \mathbf{R}_x y_t\|^2 \overset{(n=2)}{=} \|\mathbf{R}_x \mathbf{R}_t x_t - \mathbf{R}_x y_t\|^2 = \|\mathbf{R}_t x_t - y_t\|^2$$

For $n > 2$, the transformation does not work anymore.

Open problem

Minimax algorithm
for learning rotations
when $n \geq 2$