# When is There a Free Matrix Lunch

Manfred K. Warmuth

University of California - Santa Cruz

June 14 - COLT 2007

Help from Dima Kuzmin

# Expert setting

- for $t = 1, 2, \ldots, T$
  - Propose probability vector $\mathbf{w}$ of $n$ experts
  - Choose expert $i$ with probability $w_i$
  - Receive a loss vector $\boldsymbol{\ell} \in [0, 1]^n$
  - Expert $i$ incurs loss $\ell_i$
  - Alg. incurs expected loss $\mathbf{w} \cdot \boldsymbol{\ell}$
  - Update weights

$$w_i := \frac{w_i \; e^{-\eta \ell_i}}{\text{normalization}}$$

Bound

$$\text{total expected loss} - \ell^* \;\leq\; \sqrt{2\ell^* \ln n} + \ln n$$

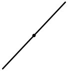
when $\eta$ is tuned as a function of $n$ and the best loss

$$\ell^* = \inf_i \sum_{t=1}^{T} \ell_i^t$$

# Lifting it to the matrix setting

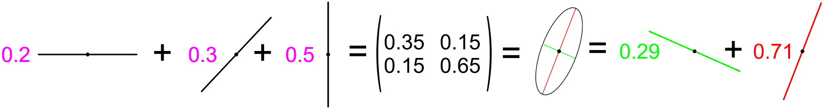
| | states | mixture states | loss spec | state loss | mix loss |
|---|---|---|---|---|---|
| old | $\begin{pmatrix}0\\1\\0\end{pmatrix}$ | $\mathbf{w} = \begin{pmatrix}0.1\\0.3\\0.6\end{pmatrix}$ | $\boldsymbol{\ell} = \begin{pmatrix}1\\1\\0\end{pmatrix}$ | $\ell_i$ | $\mathbf{w} \cdot \boldsymbol{\ell}$ |
| new | $\mathbf{u}\mathbf{u}^\top$ | $\mathbf{W} = \mathbf{U} \begin{pmatrix}0.1 & 0 & 0\\0 & 0.3 & 0\\0 & 0 & 0.6\end{pmatrix} \mathbf{U}^\top$ | $\mathbf{L} = \mathbf{V} \begin{pmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 0\end{pmatrix} \mathbf{V}^\top$ | $\mathbf{u}\mathbf{L}\mathbf{u}^\top$ | $\mathbf{W} \bullet \mathbf{L}$ |

# Mixtures of dyads

- Dyads $\mathbf{u}\mathbf{u}^\top$ are degenerate ellipses: 
- Mixtures of dyads are density matrices

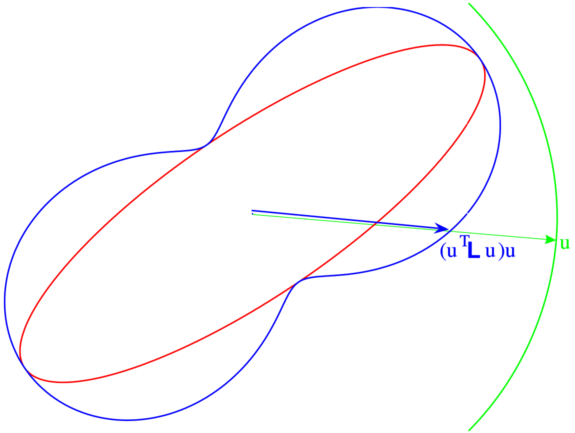
$$\mathbf{W} = \mathbf{U} \begin{pmatrix} 0.1 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & 0.6 \end{pmatrix} \mathbf{U}^\top = 0.1\ \mathbf{u}_1\mathbf{u}_1^\top + 0.3\ \mathbf{u}_2\mathbf{u}_2^\top + 0.6\ \mathbf{u}_3\mathbf{u}_3^\top$$

- Many mixtures lead to same density matrix




- There always exists a decomposition into $n$ dyads that correspond to eigenvectors

# Loss of dyads

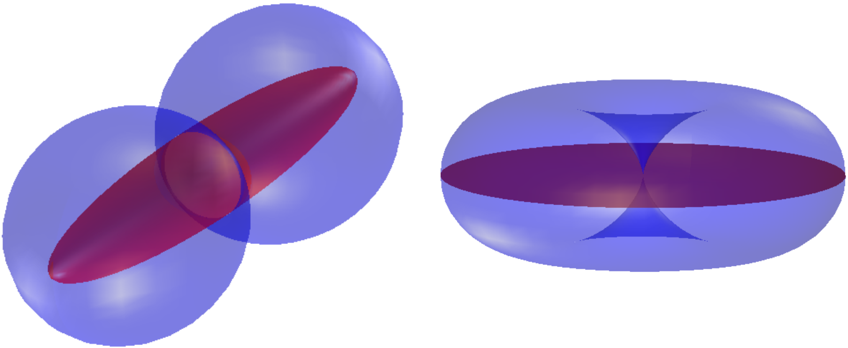



The ellipse is plot of vector **Lu** for unit vector **u**

The outer figure eight is loss $\mathbf{u}^\top \mathbf{L}\mathbf{u}$ times direction **u**

Loss is variance in direction **u** of covariance matrix **L**

# 3 dimensional cliffs



# Matrix WMR

- for $t = 1, 2, \ldots, T$
  - Propose $n$-dimensional density matrix $\mathbf{W} = \sum_i w_i \, \mathbf{u}_i \mathbf{u}_i^T$
  - Choose dyad $\mathbf{u}_i \mathbf{u}_i^\top$ with probability $w_i$
  - Receive a loss matrix $L$ with eigenvals in $[0, 1]$
  - Expert $\mathbf{u}_i \mathbf{u}_i^\top$ incurs loss $\mathbf{u}_i^\top \mathbf{L} \mathbf{u}_i$
  - Alg. incurs expected loss $\sum_i w_i \, \mathbf{u}_i^T \mathbf{L} \mathbf{u}_i = \mathbf{W} \bullet \mathbf{L}$
  - Update weights

$$\mathbf{W} := \frac{\exp(\log \mathbf{W} - \eta \mathbf{L})}{\text{trace normalization}}$$

Bound

$$\text{total expected loss} - L^* \;\leq\; \sqrt{2L^* \ln n} + \ln n$$

when $\eta$ is tuned as a function of $n$ and the best loss

$$\mathbf{L}^* = \inf_{\mathbf{u}} \sum_{t=1}^{T} \mathbf{u} \mathbf{L}^t \mathbf{u}^\top$$

# Punch line

Bounds are the same

- No additional cost for lifting the expert case to matrix case
- $n^2$ parameters as cheap as $n$ parameters
- Free matrix lunch

# The plot thickens

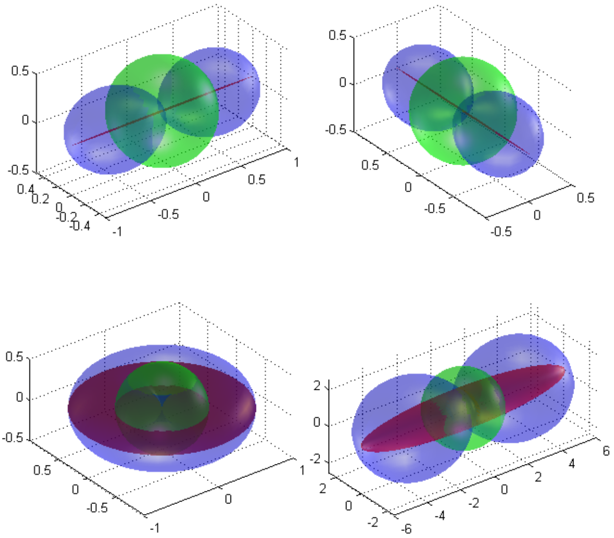
- Lift Winnow to matrix case
  Same bound as original Winnow
- Lift exponentiated gradient alg. with square loss to matrix case
  Same bound as original
- Bayes rule for density matrices
  Same cancellations
- At the core is the quantum relative entropy

$$\Delta(\widetilde{\mathbf{W}}, \mathbf{W}) := \widetilde{\mathbf{W}} \bullet (\mathbf{log}\, \widetilde{\mathbf{W}} - \mathbf{log}\, \mathbf{W})$$

Applications

- Online algorithm for PCA
- ???

# Winnowing subspaces



# Open problem

Problem class

- Online algorithm for solving a problem with symmetric matrix instances

**When is there a free matrix lunch?**

What properties are needes so that

- Worst case regret is attained when instances are diagonal

In that case matrix case for free