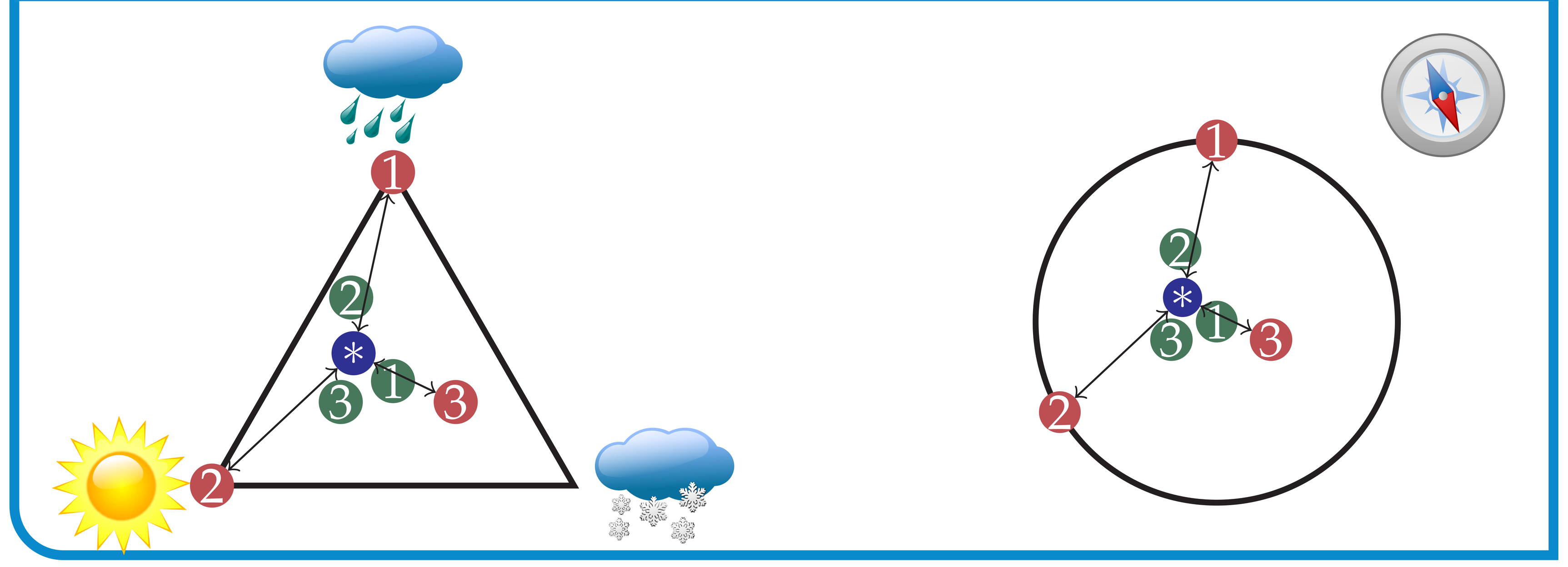
Efficient Minimax Strategies for Square Loss Games

Queensland University of Technology
Brisbane Australia

Wouter M. Koolen   Alan Malek   Peter Bartlett

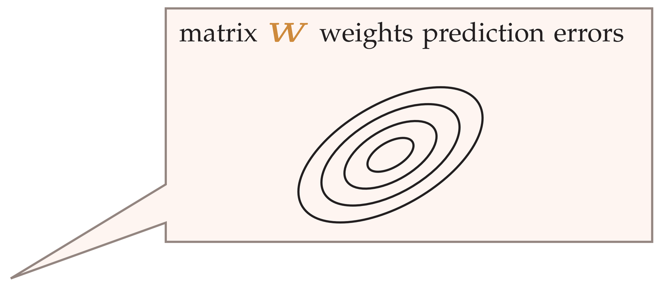
Online Prediction



Square Loss Games

Fix a convex set $\mathcal{C}$, positive definite matrix $\boldsymbol{W}$, length T.
For each round $t = 1, \ldots, T$,

- We play $\boldsymbol{a}_t \in \mathcal{C}$
- Nature reveals $\boldsymbol{x}_t \in \mathcal{C}$
- We incur loss



$$\ell(\boldsymbol{a}_t, \boldsymbol{x}_t) \;:=\; \|\boldsymbol{a}_t - \boldsymbol{x}_t\|^2_{\boldsymbol{W}} \;=\; (\boldsymbol{a}_t - \boldsymbol{x}_t)^\intercal \boldsymbol{W}^{-1} (\boldsymbol{a}_t - \boldsymbol{x}_t)$$

Our goal is to minimize regret w.r.t. best fixed action $\boldsymbol{a}$ in hindsight

$$\text{Regret} \;:=\; \sum_{t=1}^T \ell(\boldsymbol{a}_t, \boldsymbol{x}_t) - \min_{\boldsymbol{a}} \sum_{t=1}^T \ell(\boldsymbol{a}, \boldsymbol{x}_t)$$

Minimax Regret

If we assume a perfect adversary, how well can we do?

$$V \;:=\; \min_{\boldsymbol{a}_1} \max_{\boldsymbol{x}_1} \ldots \min_{\boldsymbol{a}_T} \max_{\boldsymbol{x}_T} \text{Regret}$$

We play to minimize the worst-case regret. We can solve for the **value-to-go** V in any history using the recurrence:

$$V(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T) \;:=\; -\min_{\boldsymbol{a}} \sum_{t=1}^T \ell(\boldsymbol{a}, \boldsymbol{x}_t) \qquad (1)$$

$$V(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}) \;:=\; \min_{\boldsymbol{a}_t} \max_{\boldsymbol{x}_t} \ell(\boldsymbol{a}_t, \boldsymbol{x}_t) + V(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t) \qquad (2)$$

The minimax regret V equals value-to-go $V(\epsilon)$ from empty history.
To play the minimax strategy: after seeing $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}$,

- Compute $V(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t)$
- Choose $\boldsymbol{a}_t$ as the minimizer of Equation (2)

Problem: this is expensive. Are there examples where V is a simple function of some statistics of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t$ that is simple to *precompute*?

Solving the Games

Using sufficient statistics $\boldsymbol{s} = \sum_{\tau=1}^t \boldsymbol{x}_\tau$ and $\sigma^2 = \sum_{\tau=1}^t \boldsymbol{x}_\tau^\intercal \boldsymbol{W}^{-1} \boldsymbol{x}_\tau$,

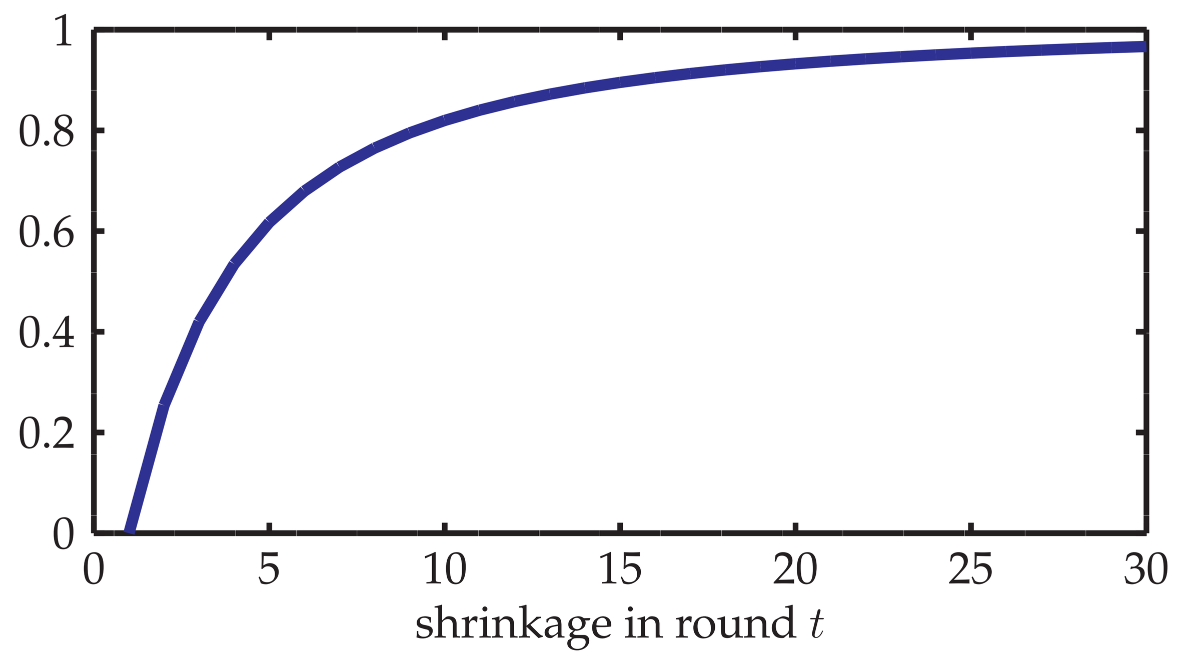
Theorem 1 (Brier Game) *Let $\mathcal{C} = \triangle$. For $\boldsymbol{W}$ satisfying an alignment condition, the value-to-go is*

$$\alpha_t \boldsymbol{s}^\intercal \boldsymbol{W}^{-1} \boldsymbol{s} - \sigma^2 + (1 - t\alpha_t) \operatorname{diag}(\boldsymbol{W}^{-1})^\intercal \boldsymbol{s} + \mathit{const},$$

and the minimax and maximin strategies for round $t+1$ are given by

$$\boldsymbol{a}^* = \boldsymbol{p}^* = \frac{\boldsymbol{s}}{t} t\alpha_{t+1} + \boldsymbol{c}(1 - t\alpha_{t+1})$$

which is data mean $\frac{\boldsymbol{s}}{t}$ shrunk towards center $\boldsymbol{c}$.



Theorem 2 (Ball Game) *Let $\mathcal{C} = \bigcirc$. For any positive definite $\boldsymbol{W}$ the value-to-go is*

$$\boldsymbol{s}^\intercal \boldsymbol{A}_t \boldsymbol{s} - \sigma^2 + \mathit{const}.$$

For round $t+1$, the minimax strategy plays

$$\boldsymbol{a}^* = \left(\lambda_{\max} \boldsymbol{I} - (\boldsymbol{A}_{t+1} - \boldsymbol{W}^{-1})\right)^{-1} \boldsymbol{A}_{t+1} \boldsymbol{s}$$

and the maximin strategy plays two unit length vectors with

$$\Pr\left(\boldsymbol{x} = \boldsymbol{a}_\perp \pm \sqrt{1 - \boldsymbol{a}_\perp^\intercal \boldsymbol{a}_\perp} \boldsymbol{v}_{\max}\right) = \frac{1}{2} \pm \frac{\boldsymbol{a}_\parallel^\intercal \boldsymbol{v}_{\max}}{2\sqrt{1 - \boldsymbol{a}_\perp^\intercal \boldsymbol{a}_\perp}},$$

where $\lambda_{\max}$ and $\boldsymbol{v}_{\max}$ correspond to the largest eigenvalue of $\boldsymbol{A}_{t+1}$ and $\boldsymbol{a}_\perp$ and $\boldsymbol{a}_\parallel$ are the components of $\boldsymbol{a}^*$ perpendicular and parallel to $\boldsymbol{v}_{\max}$.

Quadratic

Linear

All coefficients (α_t, $\boldsymbol{A}_t$) are efficiently precomputable!

Regret Bounds

- $\text{Regret}_{\text{Brier}} \propto \sum_{t=1}^T \alpha_t$.
- $\text{Regret}_{\text{Ball}} = \lambda_{\max}(\boldsymbol{W}^{-1}) \sum_{t=1}^T \alpha_t$.
- [1] show that $\sum_{t=1}^T \alpha_t = O(\log(T) - \log\log(T))$.
- Compare with $O(\log(T))$ of Follow the Leader.

[1] E. Takimoto, M. Warmuth The minimax strategy for Gaussian density estimation In *COLT '00*

Coefficient Recursions

Brier Game: $\alpha_T = \frac{1}{T}$ and

$$\alpha_t = \alpha_{t+1}^2 + \alpha_{t+1}.$$

Ball Game: $\boldsymbol{A}_T = \frac{1}{T} \boldsymbol{W}^{-1}$ and

$$\boldsymbol{A}_t = \boldsymbol{A}_{t+1} \left(\boldsymbol{W}^{-1} + \lambda_{\max} \boldsymbol{I} - \boldsymbol{A}_{t+1}\right)^{-1} \boldsymbol{A}_{t+1} + \boldsymbol{A}_{t+1},$$

which maintains the eigenvectors of $\boldsymbol{W}$ and updates the eigenvalues.

Conclusion

- Both games have tractable value functions and optimal strategies.
- All coefficients can be *precomputed* (need to know horizon T).
- Our strategies are subgame perfect: we are minimax optimal *from every history*.
- Follow-up We now know the minimax strategy for arbitrary outcome space $\mathcal{C}$.