MAX PLANCK INSTITUTE
FOR MATHEMATICS
IN THE SCIENCES

Quadratically constrained polynomial optimisation in statistics

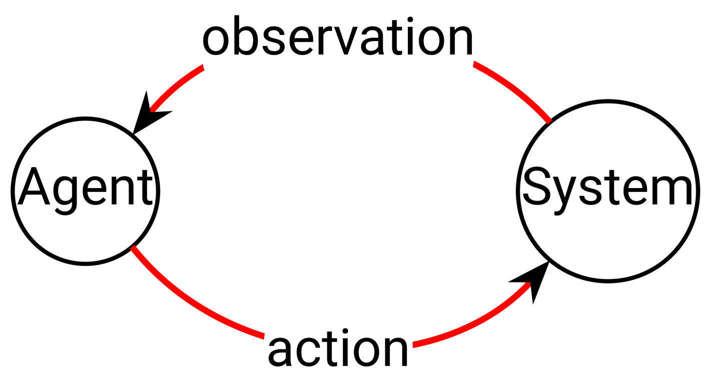
Kemal Rose

Workshop on Solving Polynomial Equations and Applications 2022 at CWI Amsterdam

MAX PLANCK INSTITUTE
FOR MATHEMATICS
IN THE SCIENCES

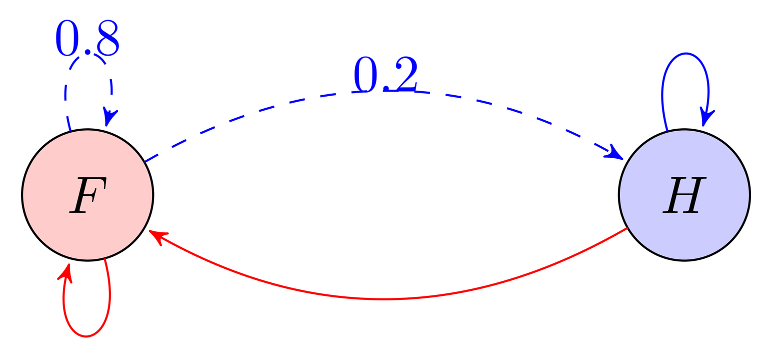
What is a POMDP

Partially observable Markov decision processes model decision processes in which an agent manipulates the state of a system in a sequence of events, having only partial information about the state of the system.



Crying baby example

The task is to feed a baby when it is hungry and not when it is full. The decision is based only on the information whether the baby cries. Feeding a baby ensures that it is no longer hungry, while an unfed baby might turn hungry:



The information whether a baby cries might not reveal the true state of the baby:

$$P(\text{cries} \mid \text{fed}) = 0.2, \; P(\text{cries} \mid \text{hungry}) = 0.9$$

The decision rule π of the agent is a stochastic map from observations to actions.

$$\pi = (P(\text{feed} \mid \text{cries}), P(\text{feed} \mid \text{doesn't cry}))$$

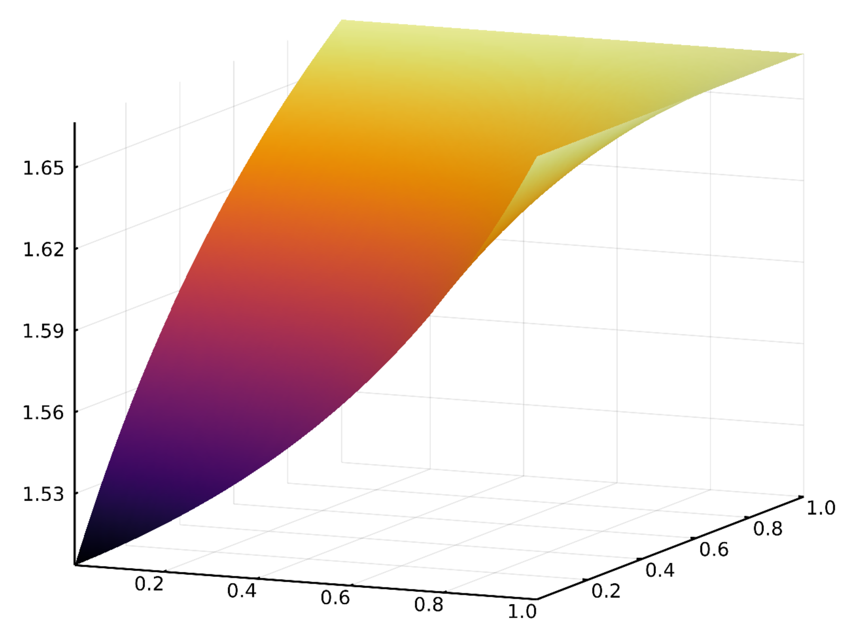
The optimisation problem

We optimise the *expected discounted reward*

$$R(\pi) = \mathbb{E}\left(\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t)\right).$$

A rational function on a product of simplices. A possible reward function is

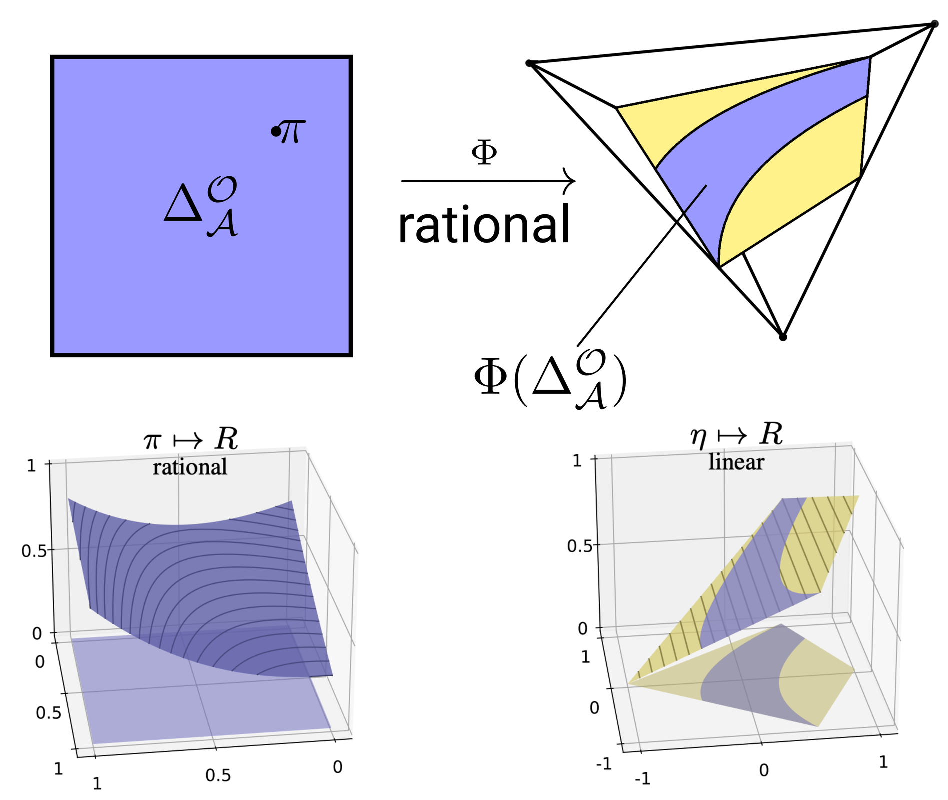
$$R(\pi) = \frac{-108.1 + 160.0\pi_1 + 40.55\pi_2 - 90.0\pi_1\pi_2 - 90.0\pi_1^2}{1.9 + 8.5\pi_1 - 0.45\pi_2}$$



The set of state-action frquencies

We denote by $\Phi(\pi)$ the state-action frequency of π. Have a linear function r with

$$R = r \circ \Phi.$$



What is new?

We recast the optimisation problem by factorising the reward function. $\Phi(\Delta_{\mathcal{A}}^{\mathcal{O}})$ is quadratically constrained! In fact, a join of Segre varieties intersected with the simplex and a linear space.

- Reward optimisation is equivalent to optimising the linear function r over the quadratically constrained set $\Phi(\Delta_{\mathcal{A}}^{\mathcal{O}})$.
- Investigate critical equations coming from the KKT approach.
- Have better numerical stability for discount factors γ close to 1.
- New approaches possible that leverage the geometry of $\Phi(\Delta_{\mathcal{A}}^{\mathcal{O}})$, I.e. Riemannian optimisation.

Semidefinite programming

We apply SDP relaxation to the quadratically constrained optimisation problem.

$$\begin{array}{ll} \text{minimise} & \textbf{trace}\,(QX) \\ \text{subject to} & X \succcurlyeq 0, \\ & X_{i,i} = 1 \;\forall i. \end{array}$$

- The resulting SDP problem can be solved efficiently.
- Computational experiments suggest: the objective value of the relaxed problem does not change.
- Approach: investigate the faces of the polyhedral cone of convex Lagrange multipliers.

References

[MM21] J. Müller and G. Montúfar. The Geometry of Memoryless Stochastic Policy Optimization in Infinite-Horizon POMDPs, 2021.

[MR17] Guido Montúfar and Johannes Rauh. Geometry of policy improvement. In *International Conference on Geometric Science of Information*, pages 282–290. Springer, 2017.

Code can be found at
`https://github.com/marinagarrote/Algebraic-Optimization-of-Sequential-Decision-Rules`

Ongoing & Future Research

→ Apply methods from Riemannian optimization
→ Investigate the multi-agent case.
→ Determine polar/ED degrees of $\Phi(\Delta_{\mathcal{A}}^{\mathcal{O}})$
→ Show objective value exactness for SDP relaxation.