Weakly chained matrices, policy iteration, and impulse control

Parsiad Azimzadeh, Peter A. Forsyth

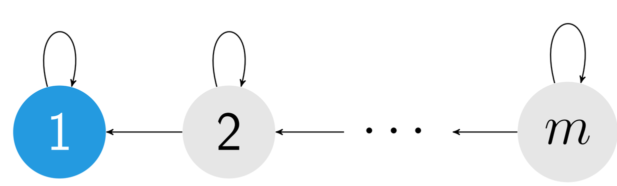
David R. Cheriton School of Computer Science, University of Waterloo

Weakly chained diagonally dominant (WCDD) matrices

$A := (a_{ij}) \in \mathbb{C}^{n \times n}$ is WCDD if for each i,

- $|a_{ii}| \geq \sum_{j \neq i} |a_{ij}|$ and
- there exists a path $i \to \cdots \to r$ in the graph of A such that $|a_{rr}| > \sum_{j \neq r} |a_{rj}|$.

$$\begin{pmatrix} 1 & & & & \\ -1 & 1 & & & \\ & \ddots & \ddots & & \\ & & -1 & 1 & \\ & & & -1 & 1 \end{pmatrix}$$



WCDD matrices are nonsingular

Proof. Let $A := (a_{ij})$ be WCDD. Suppose $Av = 0$ for some $v \neq 0$. W.l.o.g., let i be such that $|v_i| = \max_j |v_j| = 1$. Since A is WCDD, there is a path $i \to \cdots \to r$ such that $|a_{rr}| > \sum_{j \neq r} |a_{rj}|$. By the Gershgorin circle theorem,

$$|a_{ii}| \leq \sum_{j \neq i} |a_{ij}||v_j| \leq \sum_{j \neq i} |a_{ij}|.$$

Since A is WCDD, $|a_{ii}| \geq \sum_{j \neq i} |a_{ij}|$ so that the above holds with equality. Therefore, $|v_j| = 1$ whenever $|a_{ij}| \neq 0$. Repeating the argument inductively along the path $i \to \cdots \to r$, we arrive at $|a_{rr}| = \sum_{j \neq r} |a_{rj}|$, a contradiction.

Some well-known matrix families

- $A := (a_{ij})$ is WDD (resp. SDD) if for each i, $|a_{ii}| \geq \sum_{j \neq i} |a_{ij}|$ (resp. $>$).
- A matrix is an L-matrix if it has positive diagonal entries and nonpositive off-diagonal entries.
- An L-matrix whose inverse is elementwise nonnegative is an M-matrix.

[AF16, Thm. 3.5]

nonsingular WDD L-matrix

$\iff$ WCDD L-matrix

$\iff$ WDD M-matrix

The Bellman problem

Find $\vec{u} \in \mathbb{R}^m$ s.t. $\sup_{\pi \in \Pi} \{-A(\pi)\vec{u} + y(\pi)\} = 0$.

Assumptions

- $\Pi := \Pi_1 \times \cdots \times \Pi_m$ is nonempty.
- $A : \Pi \to \mathbb{R}^{m \times m}$ and $y : \Pi \to \mathbb{R}^m$ are decoupled in the sense that the i-th rows of $A(\pi)$ and $y(\pi)$ depend only on $\pi_i := \operatorname{proj}_i(\pi) \in \Pi_i$.
- The order $\geq$ and sup on $\mathbb{R}^m$ are elementwise.

Application to MDPs

Let $(X_n^\pi)_{n \geq 0}$ be a Markov chain with states $\{1, \ldots, m\}$ where $\pi \in \Pi$ is a stationary policy and Π_i is assumed to be finite. Suppose

$$\mathbb{P}(X_{n+1}^\pi = j \mid X_n^\pi = i) = [P(\pi)]_{ij},$$

where $P : \Pi \to \mathbb{R}^{m \times m}$ is decoupled. Letting $0 \leq \rho < 1$, the value $v \in \mathbb{R}^m$ of a discounted Markov decision process (MDP) is given by

$$v_i := \sup_{\pi \in \Pi} \mathbb{E}\left[\sum_{n \geq 0} \rho^n [y(\pi)]_{X_n^\pi} \,\middle|\, X_0 = i\right],$$

where $y : \Pi \to \mathbb{R}^m$ is decoupled.

Theorem. If $\vec{u}$ solves the Bellman problem with $A(\pi) := I - \rho P(\pi)$ and y as above, $\vec{u} = v$.

Policy iteration

1 **for** $\ell = 1, 2, \ldots$

2 Pick π^ℓ such that $-A(\pi^\ell)u^{\ell-1} + y(\pi^\ell)$
$= \max_{\pi \in \Pi}\{-A(\pi)u^{\ell-1} + y(\pi)\}$

3 Solve $A(\pi^\ell)u^\ell = y(\pi^\ell)$ for u^ℓ

Convergence

Theorem. If Π is compact, A and y are continuous, and each $A(\pi)$ is an M-matrix, $(u^\ell)_\ell$ defined by Policy-Iteration converges to the solution of the Bellman problem for any initial guess u^0. For a more general result, see [AF16, Prop. 2.2].

That's great, but...

- The Bellman problem arises naturally in MDPs and monotone numerical schemes for PDEs.
- In these applications, the resulting input matrices $\{A(\pi)\}$ are WDD L-matrices.
- However, they are not always WCDD! (e.g., $\rho = 0$ in MDP) Here, policy iteration may fail due to a singular iterate $A(\pi^\ell)$! [AF16, Example 4.9]

Monotone scheme for HJBQVI

For the remainder, we consider a parameterization of the Bellman problem corresponding to a monotone numerical scheme for the Hamilton-Jacobi-Bellman quasi-variational inequality (HJBQVI)

$$-\max\left\{\sup_{b \in B}\left\{\left(\partial_t + L^b\right)u + f^b\right\}, Mu - u\right\} = 0$$

where

$$(Mu)(t, x) := \sup_{z \in Z(t,x)} \{u(t, \Gamma^z(t, x)) + k^z(t, x)\}.$$

Corresponding Bellman problem

A numerical scheme for the HJBQVI involves discretizing the time horizon $[0, T]$ into $t_1 < \cdots < t_N$ and solving a Bellman problem at each point t_n.

Boldface is used to denote objects arising from the HJBQVI discretization (e.g., on a grid with nodes $x_1, \ldots, x_m$, $\mathbf{B}_i$ is a discretization of the control set B at x_i and $\mathbf{L}(\mathbf{b})$ is a matrix associated with L^b).

Let

$$\Pi := \Pi_1 \times \cdots \times \Pi_m \text{ where } \Pi_i := \mathbf{B}_i \times \mathbf{Z}_i \times D_i$$

and $D_i \subset \{0, 1\}$. For each $\pi \in \Pi$, we write $\pi \equiv (\mathbf{b}, \mathbf{z}, d)$ with the obvious interpretation. Let

$$I - A(\pi) := (I - \operatorname{diag} d)\,\mathbf{L}(\mathbf{b}) + (\operatorname{diag} d)\,\mathbf{\Gamma}(\mathbf{z})$$
$$\text{and } y(\pi) := (I - \operatorname{diag} d)\,\mathbf{f}(\mathbf{b}) + (\operatorname{diag} d)\,\mathbf{k}(\mathbf{z})$$

where $I - \mathbf{L}(\mathbf{b})$ and $I - \mathbf{\Gamma}(\mathbf{z})$ are SDD and WDD L-matrices, respectively. The Bellman problem with A, y, and Π as given above and solution $\vec{u}$ has the following correspondence with the HJBQVI:

$$\mathbf{L}(\mathbf{b})\vec{u} + \mathbf{f}(\mathbf{b}) \approx L^b u(t, \cdot) + f^b(t, \cdot)$$
$$\text{and } \mathbf{\Gamma}(\mathbf{z})\vec{u} + \mathbf{k}(\mathbf{z}) \approx u(t, \Gamma^z(t, \cdot)) + k^z(t, \cdot)$$

Convergence

While the conditions on $I - \mathbf{L}(\mathbf{b})$ and $I - \mathbf{\Gamma}(\mathbf{z})$ ensure that $A(\pi)$ is a WDD L-matrix, it is not necessarily the case that $A(\pi)$ is WCDD!

We refer to [AF16, Thm. 4.8 and Thm. 4.10] for convergence results in this more delicate case, and move to a numerical example employing these results.

Optimal consumption

Let $(S_t)_{t \geq 0}$ and $(R_t)_{t \geq 0}$ denote the amount of money in an investment and riskless bank account, respectively. An investor can, at any point in time t:

- consume continuously at rate b_t;
- transfer money to ($z > 0$) or from ($z < 0$) the investment incurring a transaction fee of $\lambda|z| + c$.

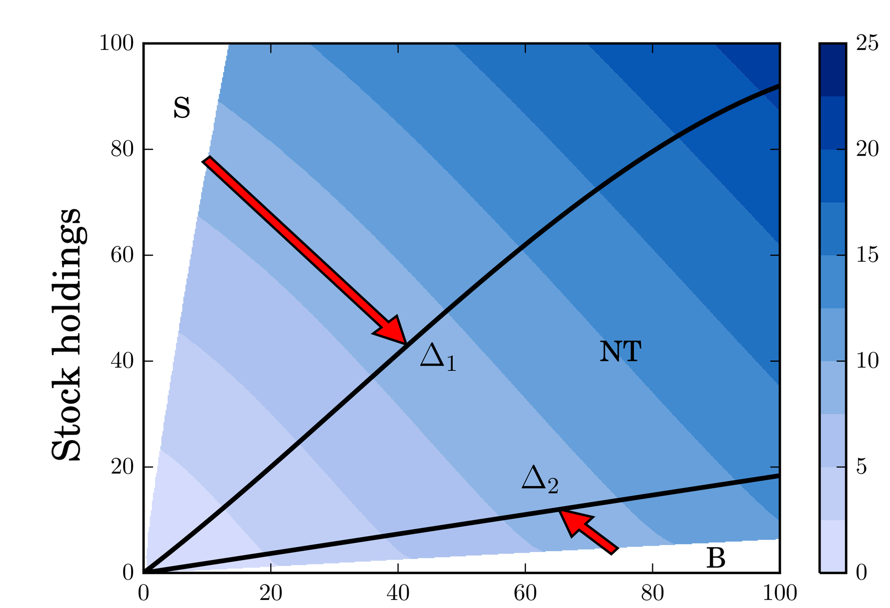
This is captured by a system of SDEs with impulse:

$$\begin{aligned} dS_t &= \mu S_t dt + \sigma S_t dW_t & \text{if } \tau_j < t < \tau_{j+1} \\ dR_t &= (rR_t - b_t)dt & \text{if } \tau_j < t < \tau_{j+1} \\ S_{\tau_j} &= S_{\tau_j -} + z_{\tau_j} & \\ R_{\tau_j} &= R_{\tau_j -} - z_{\tau_j} - \lambda\left|z_{\tau_j}\right| - c & \end{aligned}$$

The investor's expected utility at (t, S_t, R_t) is

$$v(t, S_t, R_t) := \sup \mathbb{E}\Big[\int_t^T e^{-\rho(s-t)} b_s^\gamma / \gamma ds + e^{-\rho(T-t)} \max\left\{R_T + (1-\lambda)S_T - c, 0\right\}^\gamma / \gamma\Big]$$

where $0 \leq 1 - \gamma < 1$ is the investor's relative risk-aversion and the sup is taken over all controls $((b_s)_{s \geq t}; \tau_j, z_j)$ such that $(S_s)_{s \geq t}$ and $(R_s)_{s \geq t}$ are nonnegative. It turns out that v solves an appropriately parameterized HJBQVI, for which an optimal control (obtained numerically) under reasonable parameters is shown below (details in [AF16, Section 6.2]).



[AF16] P. Azimzadeh and P. A. Forsyth.
Weakly chained matrices, policy iteration, and impulse control.
SIAM J. Numer. Anal., 54(3):1341–1364, 2016.