Max-plus algebra in the numerical solution of Hamilton-Jacobi and Isaacs equations

Marianne Akian

(INRIA Saclay - Île-de-France and CMAP, École Polytechnique)

BIRS Workshop 11w5086
Advancing numerical methods for viscisity solutions and Applications
Feb. 14–18, 2011

works with A. Lakhoua, S. Gaubert, A. Guterman, S. Detournay.

Dynamic programming equations of optimal control and zero-sum games problems

For instance if the Hamiltonian H is convex:

$$H(x,p,X) = \sup_{\alpha\in\mathcal{A}} \left[p\cdot f(x,\alpha) + \frac{1}{2}\operatorname{tr}(\sigma(x,\alpha)\sigma^T(x,\alpha)X) + r(x,\alpha) \right]$$

and under regularity conditions, v is the viscosity solution of

$$-\frac{\partial v}{\partial t} + H(x, \frac{\partial v}{\partial x}, \frac{\partial^2 v}{\partial x^2}) = 0,\ (x,t)\in X\times[0,T), \qquad v(x,T) = \phi(x),\ x\in X,$$

if and only if v is the value function of the finite horizon stochastic control problem:

$$\begin{aligned} v(x,t) = \quad & \sup \mathbb{E}[\textstyle\int_t^T r(\mathbf{x}(s),\mathbf{a}(s))\,ds + \phi(\mathbf{x}(T)) \mid \mathbf{x}(t) = x] \\ & d\mathbf{x}(s) = f(\mathbf{x}(s),\mathbf{a}(s)) + \sigma(\mathbf{x}(s),\mathbf{a}(s))dW(s), \quad \mathbf{x}(s)\in X \\ & \mathbf{a} \text{ strategy, } \mathbf{a}(s)\in\mathcal{A}. \end{aligned}$$

Max-plus or tropical algebra

- It is the idempotent semiring $\mathbb{R}_{\max} := (\mathbb{R} \cup \{-\infty\}, \oplus, \otimes)$, where $a \oplus b = \max(a, b)$ and $a \otimes b = a + b$. The neutral elements are $\mathbb{0} = -\infty$ and $\mathbb{1} = 0$.
- It is the limit of the logarithmic deformation of $\mathbb{R}_+$ semiring:

$$\begin{aligned} \max(a, b) = & \lim_{\varepsilon \to 0_+} \varepsilon \log(e^{a/\varepsilon} + e^{b/\varepsilon}) \\ a + b = & \varepsilon \log(e^{a/\varepsilon} e^{b/\varepsilon}) \end{aligned}$$

  and the usual order of $\mathbb{R}$ is a "natural" order on $\mathbb{R}_{\max}$, for which all elements are "positive" or "zero".
- The *complete max-plus algebra* $\overline{\mathbb{R}}_{\max}$ is obtained by completing $\mathbb{R}_{\max}$ with the $+\infty$ element with the convention $+\infty + (-\infty) = -\infty$.
- One can define on $\mathbb{R}_{\max}$ or $\overline{\mathbb{R}}_{\max}$ notions similar to those of usual algebra: matrices, scalar product, linear spaces, measures, integrals, cones,...

Part I: Max-plus discretizations

First order HJ equations, or dynamic programming equations of undiscounted deterministic optimal control problems are max-plus linear, that is the *Lax-Oleinik semigroup* $S^t : \phi \mapsto v(\cdot, T-t)$ is max-plus linear (Maslov, 87):

$$S^t(\sup(\lambda_1 + \phi_1, \lambda_2 + \phi_2)) = \sup(\lambda_1 + S^t(\phi_1), \lambda_2 + S^t(\phi_2)) \ ,$$

where $\lambda + \phi : x \mapsto \lambda + \phi(x)$.

Recall that the set of all functions $X \to \mathbb{R}_{\max}$ or $\overline{\mathbb{R}}_{\max}$ is a *max-plus semimodule* (that is a linear space over $\mathbb{R}_{\max}$), where

- the addition is the pointwise maximum, which is equivalent to the supremum,
- the multiplication by a scalar is the pointwise addition $\lambda \cdot \phi = \lambda + \phi$.

Max-plus analogue of linear PDEs

Usual algebra	Max-plus algebra
Parabolic PDE: $-\frac{\partial v}{\partial t} + Lv = 0$	Evolution HJ PDE: $-\frac{\partial v}{\partial t} + H(x, \frac{\partial v}{\partial x}) = 0$
Heat equation $Lv := \Delta v$	LQ problem: $H(x, p) := \frac{p^2}{2}$
Elliptic PDE: $Lv = 0$	Stationnary HJ: $H(x, \frac{\partial v}{\partial x}) = 0$
Eigenproblem: $Lv = \lambda v$	Ergodic HJ: $-\lambda + H(x, \frac{\partial v}{\partial x}) = 0$

with

$$Lv = \frac{1}{2} \sum_{i,j=1}^{n} a_{ij}(x) \frac{\partial^2 v}{\partial x_i \partial x_j} + \sum_{i=1}^{n} g_i(x) \frac{\partial v}{\partial x_i} - \delta(x) v + c(x).$$

Max-plus analogue of discretization schemes

Usual algebra	Max-plus algebra
Probabilist point of view: Discretize the Brownian process	Discretize the max-plus *Brownian process* (A., Quadrat, Viot, 98).
Variational point of view: Weak solution of $-\frac{1}{2}\Delta v = f$ on $\Omega,\ v = 0$ on $\partial\Omega$ $v \in \mathcal{V},\ \frac{1}{2}\int \nabla v \nabla \phi = \int f\phi\ \forall \phi \in \mathcal{V}$, where $\mathcal{V} = H_0^1(\Omega)$. *FEM*: replace V by finite dimensional subspaces	Generalized solution of HJ (Kolokoltzov and Maslov, 88): $v^t \in \mathcal{W},\ \langle v^{t+\delta}, \phi\rangle = \langle S^\delta v^t, \phi\rangle\ \forall \phi \in \mathcal{Z}$ $\mathcal{W},\ \mathcal{Z}$ are subsemimodules of $\mathbb{R}_{\max}^X$. *Max-plus FEM*: replace $\mathcal{W}$ and $\mathcal{Z}$ by finitely generated subsemimodules (A. Gaubert, Lakhoua, SICON 08).
	Replace S^δ by a finite dimensional max-plus linear operator (Fleming and McEneaney, 00).
Finite difference point of view: Error: use linearity and regularity, or monotonicity	impossible possible.

The max-plus finite element method

- The max-plus scalar product is given by:

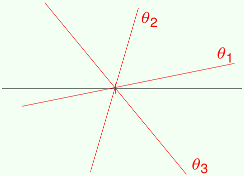
$$\langle u, v \rangle = \sup_{x \in X} u(x) + v(x) \ .$$

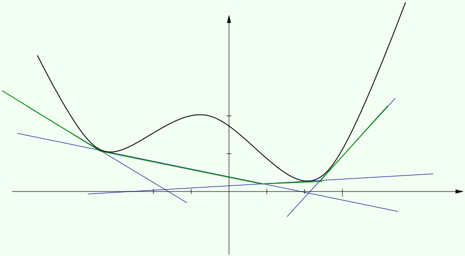
- We fix the max-plus semimodules $\mathcal{W}$ and $\mathcal{Z}$ for solutions and test functions, together with some approximation of them by finitely genererated subsemimodules $\mathcal{W}_h$ and $\mathcal{Z}_h$ (here and in the sequel h refers to discretized objects):

$$\mathcal{W}_h = \text{span}\{w_1, \ldots, w_p\} \quad \textit{finite elements}$$
$$\mathcal{Z}_h = \text{span}\{z_1, \ldots, z_q\} \quad \textit{test functions}$$

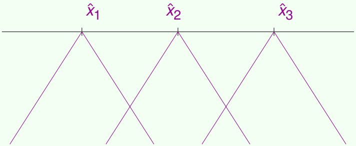
Examples of semimodule and their discretizations:

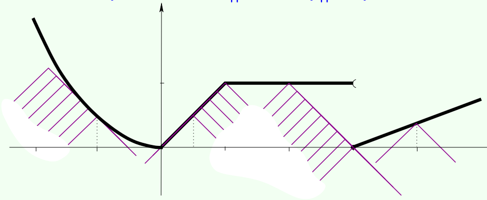
- $\mathcal{W}$ is the space of l.s.c. convex functions and $w_i : x \mapsto \theta_i \cdot x$, $\theta_i \in \mathbb{R}^n$.





- $\mathcal{W}$ is the space of l.s.c. c-semiconvex functions and $w_i : x \mapsto -\frac{c}{2}\|x - \hat{x}_i\|^2$, $x_i \in \mathbb{R}^n$.
- $\mathcal{W}$ is the space of 1-Lip functions and $w_i : x \mapsto -\|x - \hat{x}_i\|$, $x_i \in \mathbb{R}^n$.





The max-plus FEM (continued)

- The approximation v_h^t of the generalized solution v^t of HJ equation must satisfy

$$v_h^t \in \mathcal{W}_h, \quad \langle v_h^{t+\delta}, \phi \rangle = \langle S^\delta v_h^t, \phi \rangle \; \forall \phi \in \mathcal{Z}_h, \; t = \delta, 2\delta, \dots,$$

- This is equivalent to

$$v_h^t = \sup_{1 \le j \le p} \lambda_j^t + w_j$$

and

$$\sup_{1 \le j \le p} (\lambda_j^{t+\delta} + \langle w_j, z_i \rangle) = \sup_{1 \le j \le p} (\lambda_j^t + \langle S^\delta w_j, z_i \rangle) \quad \forall 1 \le i \le q \;\; .$$

- This equation is of the form $M\lambda^{t+\delta} = K\lambda^t$, where M and K are analogues of the mass and stiffness matrices, respectively.

- To compute $\lambda^{t+\delta}$ as a function of λ^t, one need to solve a max-plus linear system of the form $M\mu = \nu$, which may not have a solution.
- But it has always a greatest subsolution ($M\mu \leq \nu$), $M^\sharp \nu$, where $M^\sharp$ is a the adjoint of M, it is a min-plus linear operator:

$$(M^\sharp \nu)_j = \min_{1 \leq i \leq q} -M_{ij} + \nu_i \ .$$

- So we take for max-plus FEM iteration:

$$\lambda^{t+\delta} = M^\sharp K \lambda^t \ .$$

Summary of max-plus FEM:

- Approach v^t by $v_h^t := \sup_{1\leq j\leq p} \lambda_j^t + w_j$ where the λ_j^0 are given, and

$$\lambda_j^{t+\delta} = \min_{1\leq i\leq q}\left(-\langle w_j, z_i\rangle + \max_{1\leq k\leq p}\left(\langle S^\delta w_k, z_i\rangle + \lambda_k^t\right)\right), \ t = \delta, 2\delta, \ldots, \ 1$$

- This is a zero-sum two player (deterministic) game dynamic programming equation !
- The states and actions are in $[p] : \{1, \ldots, p\}$ and $[q]$, Min plays in states $j \in [p]$, choose a state $i \in [q]$ and receive M_{ij} from Max, Max plays in states $i \in [q]$, chooses a state $k \in [p]$ and receive K_{ik} from Min. $\lambda_j^{N\delta}$ is the value of the game after N turns (Min + Max) starting in state j.

A geometric rewritting of the max-plus FEM :

- The FEM iterations can also be written as:

$$v_h^{t+\delta} = \Pi_h \circ S^\delta(v_h^t) \quad \text{and} \quad v_h^0 = P_{\mathcal{W}_h} v^0$$

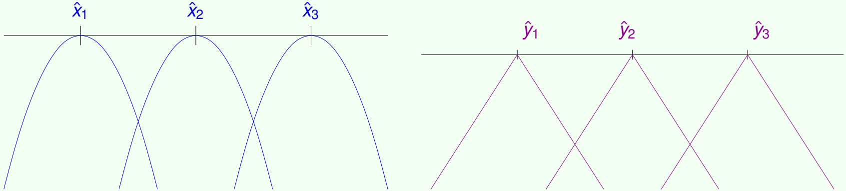
where

$$\begin{aligned} \Pi_h = & \ P_{\mathcal{W}_h} \circ P^{-\mathcal{Z}_h} \\ P_{\mathcal{W}_h} v = & \ \max\{w \in \mathcal{W}_h \mid \ w \leq v\} \\ P^{-\mathcal{Z}_h} v = & \ \min\{w \in -\mathcal{Z}_h \mid \ w \geq v\} \ . \end{aligned}$$

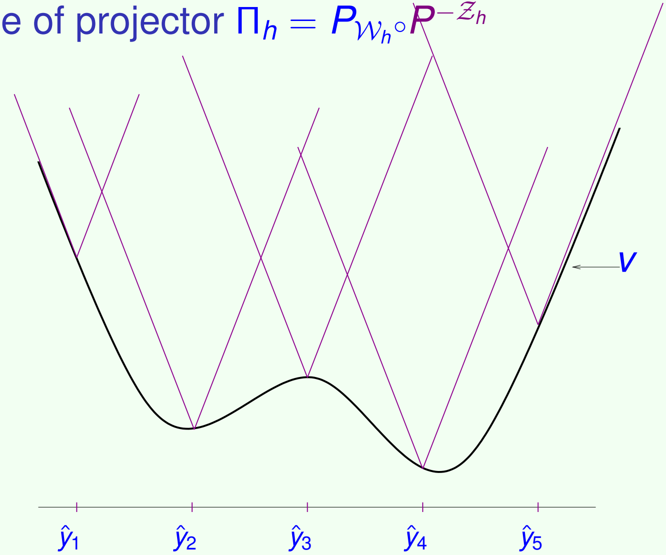
- These max-plus projectors were studied by Cohen, Gaubert, Quadrat, they are nonexpansive in the sup-norm.

Example of projector $\Pi_h = P_{\mathcal{W}_h} \circ P^{-\mathcal{Z}_h}$

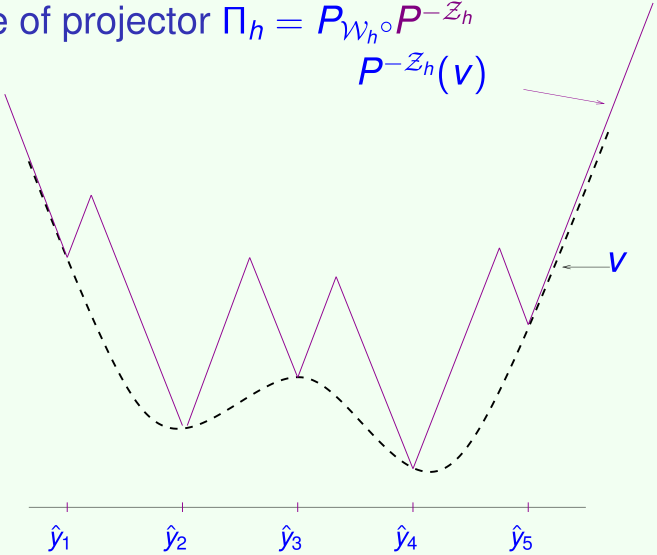
We choose $P2$ finite elements and $P1$ test functions



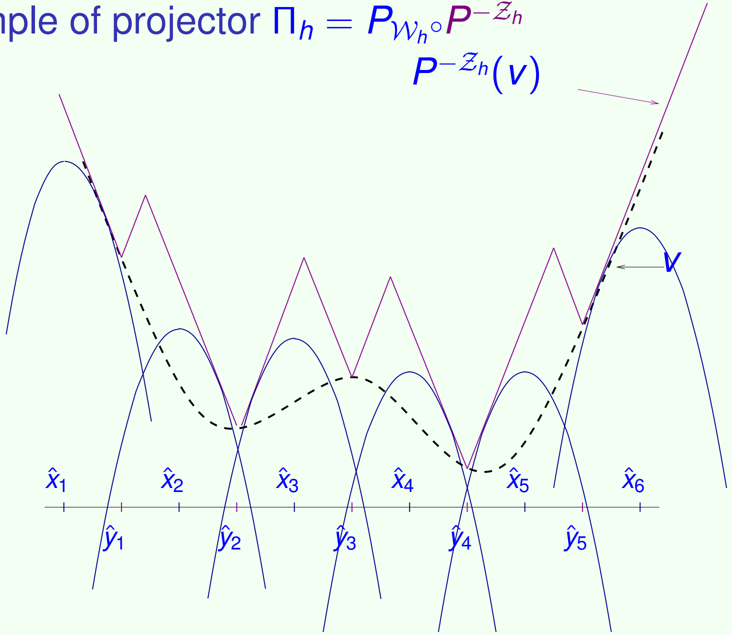
Example of projector $\Pi_h = P_{\mathcal{W}_h} \circ P^{-\mathcal{Z}_h}$



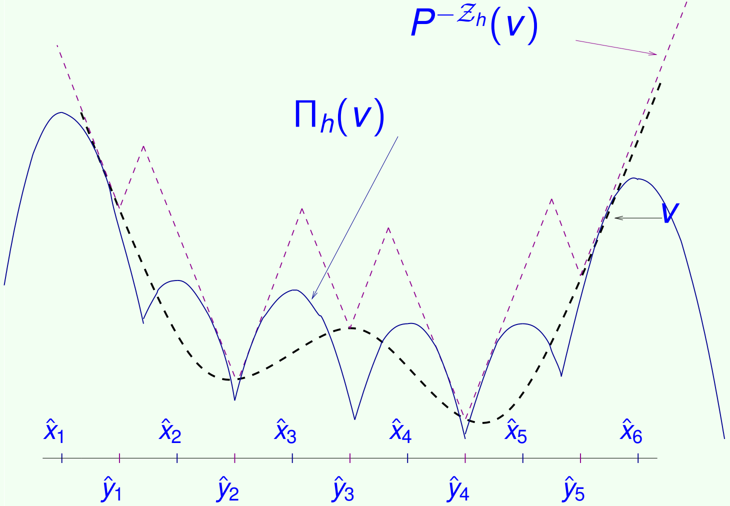
Example of projector $\Pi_h = P_{\mathcal{W}_h} \circ P^{-\mathcal{Z}_h}$



Example of projector $\Pi_h = P_{\mathcal{W}_h} \circ P^{-\mathcal{Z}_h}$



Example of projector $\Pi_h = P_{\mathcal{W}_h} \circ P^{-\mathcal{Z}_h}$



- As in the usual FEM, the error can be estimated from projection errors:

$$\begin{aligned} \|v_h^T - v^T\|_\infty &\leq (1+\tfrac{T}{\delta})\,\text{Projection error} \\ \text{Projection error} &= \sup_{t=0,\delta,\dots,T} \|P_{\mathcal{W}_h} \circ P^{-\mathcal{Z}_h} v^t - v^t\|_\infty \\ &\leq \sup_{t=0,\delta,\dots,T} (\|P^{-\mathcal{Z}_h} v^t - v^t\|_\infty + \|P_{\mathcal{W}_h} v^t - v^t\|_\infty). \end{aligned}$$

- By convexity techniques, we obtain

$$\text{Projection error} \leq C(\Delta x)^k/\delta$$

  with $k = 1$ or 2 depending on the "degree" of the finite elements and on the regularity of the solution v^t, and Δx equal to the "diameter" of the space discretization (Voronoi cells or Delaunay triangulation).
- The max-plus approximation theory seems limited to $k = 2$?

However, this was an ideal FEM method. One need to compute:

$$M_{ij} = \langle w_j, z_i \rangle = \sup_{x \in X} w_j(x) + z_i(x)$$

$$K_{ik} = \langle z_i, S^\delta w_k \rangle = \sup_{x \in X, \mathbf{u}(\cdot)} \; z_i(x) + \int_0^\delta \ell(\mathbf{x}(s), \mathbf{u}(s))\, ds + w_k(x)$$

- For good choices of w_j and z_i, M_{ij} can be computed analytically.
- Computing K_{ik} is a usual optimal control problem, but horizon δ may be small, and the final and terminal rewards w_j and z_i may be chosen to be nice, so that K_{ik} may be well approximated.
- Then

$$\|v_h^T - v^T\|_\infty \leq (1 + \frac{T}{\delta})(\text{Projection error} + \text{Approximation error})$$

- For instance, using r-order one-step approximations of $S^\delta w_i(x)$, Approximation error$= O(\delta^{r+1})$.
- So the total max-plus FEM error is in the order of

$$(\Delta x)^k/\delta + \delta^r \quad ,$$

with $r \geq 1$, and $k = 1$ or 2.

Remarks

- These error estimates are similar to those of some semi-lagrangian schemes.
- They need some regularity of l and f and do not work for Dirichlet limit conditions, or variational inequalities (stopping time problems).
- Hence it is not clear that they are less diffusive than usual finite difference methods.
- δ need to be small and $\Delta x \simeq \delta^{\frac{r+1}{k}}$.
- The matrices are full, then the complexity ($O(\epsilon^{-(1+2n)})$ when $k = 2$ and $r = 1$) is too large to be able to handle problems with dimension > 2.
- It is comparable with the complexity of the finite difference method, if we consider the usual estimation of this method that is in $O(\delta^{1/2})$.

Perspectives

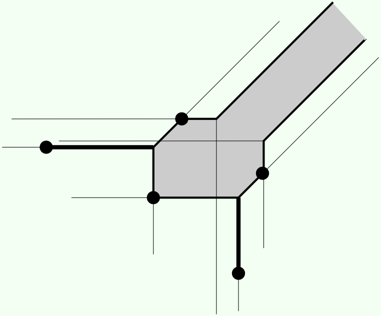
- Take higher order methods to approximate K or $S^\delta w_i$, for instance a direct or Pontryagin method with step $\Delta t << \delta$ and order r.
- Then the error is in the order of

$$(\Delta t)^r + (\Delta x)^k/\delta \ ,$$

  as soon as δ is small enough (but of order 1) to ensure that convexity propagate and that the global optimum of the control problem related to the computation of K_{ij} is accessible by Pontryagin method.
- The complexity would then be in $O(\epsilon^{-(1+n)})$ when $r = 1$ and $k = 2$, thus comparable to that of the finite difference method, if the error of this method were in $O(\Delta t)$.
- But it should be able to handle less regular value functions, and also less regular lagrangians and drifts, so Dirichlet boundary conditions or variational inequalities.
- It has some similarity with the point of view of McEneaney combining Riccati solutions with max-plus linearity.
- However, the problem of *Curse of dimensionality* is still there.

Part II: Tropical convex sets

$C \subset \mathbb{R}^n_{\max}$ is a tropical convex set if
$f, g \in C \implies [f, g] := \{(\lambda + f) \vee (\mu + g) \mid \lambda, \mu \in \mathbb{R}_{\max},\ \lambda \vee \mu = 0\} \in C$



Tropical convex cones $\Leftrightarrow$ subsemimodules over $\mathbb{R}_{\max}$.

Theorem

Every closed tropical convex cone of $\mathbb{R}_{\max}^n$ is the intersection of tropical half-spaces, which means:

$$C = \{u \in \mathbb{R}_{\max}^n \mid Au \leq Bu\}$$

with $A, B \in \mathbb{R}_{\max}^{I\times[n]}$, and I possibly infinite.

This comes from the *max-plus separation theorem*, see for instance Zimmermann 77, Cohen, Gaubert, Quadrat 01 and LAA04.

Tropical polyhedral cones are the intersection of finitely many tropical half-spaces ($I = [m]$), or equivalently, the convex hull of finitely many rays.
See the works of Gaubert, Katz, Butkovič, Sergeev, Schneider, Allamigeon,....
See also the tropical geometry point of view Sturmfels, Develin, Joswig, Yu,....

Recall: $Au \leq Bu \Leftrightarrow u \leq f(u)$ with $f(u) = A^{\sharp}Bu$,

$$(f(u))_j = \inf_{i \in I}(-A_{ij} + \max_{k \in [n]}(B_{ik} + u_k)) \enspace .$$

f is a min-max function (Olsder 91) when I is finite. In that case, $f : \mathbb{R}^n \to \mathbb{R}^n$ when the columns of A and the rows of B are not $\equiv -\infty$.

But the following are equivalent for $f : \mathbb{R}^n \to \mathbb{R}^n$:

1. f can be written as $f(u) = A^\sharp Bu$ with $A, B \in \mathbb{R}_{\max}^{I\times[n]}$;
2. f is the *dynamic programming operator* of a zero-sum two player deterministic game:
$$[f(u)]_j = \inf_{i\in I} \max_{k\in[n]}(r_{jik} + u_k)$$
3. f is order preserving ($u \leq y \Rightarrow f(u) \leq f(y)$) and additively homogeneous ($f(\lambda + u) = \lambda + f(u)$).
4. f is the *dynamic programming operator* of a zero-sum two player stochastic game:
$$[f(u)]_j = \inf_{\alpha\in\mathcal{A}} \sup_{\beta\in\mathcal{B}}(r_j^{\alpha,\beta} + \sum_{k\in[n]}(P_{jk}^{\alpha,\beta} u_k))$$

Then $C := \{u \in (\mathbb{R} \cup \{-\infty\})^n \mid u \leq f(u)\}$ is a tropical convex cone.
See Kolokoltsov; Gunawardena, Sparrow; Rubinov, Singer for $3 \Rightarrow 2$ or 4, take $I = \mathbb{R}^n$ and $r_{jyk} = f(y)_j - y_k$.

Proposition ((A., Gaubert, Guterman 09), uses (Nussbaum, LAA 86))

Let f be a continuous, order preserving and additively homogeneous self-map of $(\mathbb{R}\cup\{-\infty\})^n$, then the following limit exists and is independent of the choice of u:

$$\bar{\chi}(f) := \lim_{N\to\infty} \max_{j\in[n]} f_j^N(u)/N \ ,$$

and equals the following numbers:

$$\begin{aligned}
\rho(f) &:= \max\{\lambda\in\mathbb{R}_{\max} \mid \exists u\in\mathbb{R}^n_{\max}\setminus\{-\infty\},\ f(u)=\lambda+u\} \ ,\\
\mathrm{cw}(f) &:= \inf\{\mu\in\mathbb{R} \mid \exists w\in\mathbb{R}^n, f(w)\leq\mu+w\} \ ,\\
\mathrm{cw}'(f) &:= \sup\{\lambda\in\mathbb{R}_{\max} \mid \exists u\in\mathbb{R}^n_{\max}\setminus\{-\infty\},\ f(u)\geq\lambda+u\} \ .
\end{aligned}$$

Moreover, there is at least one coordinate $j\in[n]$ such that $\chi_j(f) := \lim_{N\to\infty} f_j^N(u)/N$ exists and is equal to $\bar{\chi}(f)$.

$\chi_j(f)$ is the *mean payoff* of the game starting in state j.
See also Vincent 97, Gunawardena, Keane 95, Gaubert, Gunawardena 04.

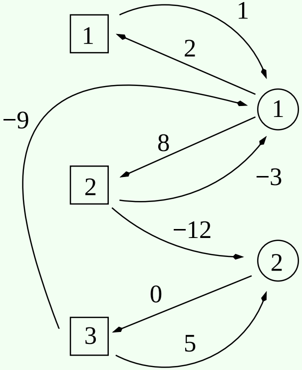
Theorem

Let

$$C = \{u \in \mathbb{R}^n_{\max} \mid Au \leq Bu\}$$

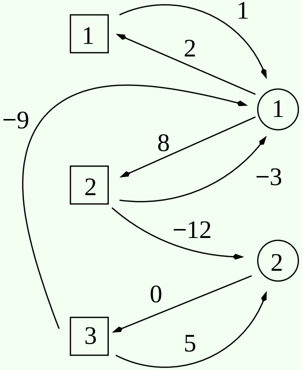
$\exists u \in C \setminus \{-\infty\}$ *iff Max has at least one winning position in the mean payoff game with dynamic programming operator* $f(u) = A^\sharp Bu$, *i.e.,* $\exists j \in [n], \chi_j(f) \geq 0$.

$$A = \begin{pmatrix} 2 & -\infty \\ 8 & -\infty \\ -\infty & 0 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & -\infty \\ -3 & -12 \\ -9 & 5 \end{pmatrix}$$

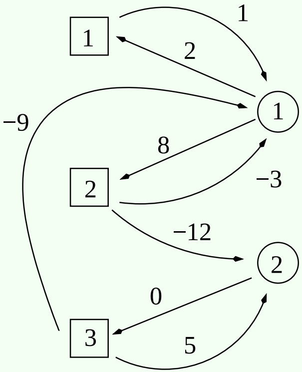


players receive the weight of the arc

$$\begin{aligned} 2 + u_1 &\leq 1 + u_1 \\ 8 + u_1 &\leq \max(-3 + u_1, -12 + u_2) \\ u_2 &\leq \max(-9 + u_1, 5 + u_2) \end{aligned}$$



$$
\begin{aligned}
2 + u_1 &\le 1 + u_1 \\
8 + u_1 &\le \max(-3 + u_1, -12 + u_2) \\
u_2 &\le \max(-9 + u_1, 5 + u_2)
\end{aligned}
$$



$\chi(f) = (-1, 5), u = (-\infty, 0)$ solution

Theorem ((A., Gaubert, Guterman 09))

Whether an (affine) tropical polyhedron

$$\{u \in \mathbb{R}^n_{\max} \mid \max(\max_{j\in[n]}(A_{ij}+u_j), c_i) \leq \max(\max_{j\in[n]}(B_{ij}+u_j), d_i),\ i \in [m]\}$$

is non-empty reduces to whether a specific state of a mean payoff game is winning.

The proof relies on Kohlberg's theorem (80) on the existence of invariant half-lines $f(u+t\eta) = u+(t+1)\eta$ for t large.

Corollary

Each of the following problems:

1. *Is an (affine) tropical polyhedron empty?*
2. *Is a prescribed initial state in a mean payoff game winning?*

can be transformed in linear time to the other one.

- Hence, algorithms (value iteration, policy iteration) and complexity results for mean-payoff games can be used in tropical convexity.
- Conversely one can compute $\chi(f)$ by dichotomy solving emptyness problems for convex polyhedra, so tropical linear programs.
- Can we find new algorithms for mean payoff games using this correspondance?
- Can we find polynomial algorithms for all these problems?

Part III: Policy iterations for stationnary zero-sum games

Consider the stationnary Isaacs equation:

$$-\rho + H(x, \frac{\partial v}{\partial x}, \frac{\partial^2 v}{\partial x^2}) = 0,\ x \in X$$

where we look for the *mean payoff* ρ and the *bias function* v.
Using a monotone discretization, one obtains the additive spectral problem:

$$\rho + u = f(u),\ u \in \mathbb{R}^N \quad ,$$

where f is the *dynamic programming operator* of a zero-sum two player undiscounted stochastic game.

We want to construct a fast algorithm that works even when the Markov matrices associated to fixed strategies may not be irreducible and for a large $N \Longrightarrow$ policy iterations with algebraic multigrid methods.

Policy iterations for optimal control problems

- For a discounted infinite horizon problem, one need to solve:

$$u = f(u), \quad \text{where } [f(u)]_j = \sup_{\alpha \in \mathcal{A}} f(u; j, \alpha) := r_j^\alpha + \sum_{k \in [N]} (P_{jk}^\alpha u_k) \ .$$

  Here, for each strategy $\bar{\alpha} : [N] \to \mathcal{A}$, the matrix $(P_{ij}^{\bar{\alpha}(j)})_{jk}$ is strictly submarkovian.
- The policy iteration (Howard 60): starts with $\bar{\alpha}_0$, and iterates:
  - v^{n+1} is the value with fixed strategy $\bar{\alpha}_n$:

$$v_j^{n+1} = f(v^{n+1}; j, \bar{\alpha}_n(j)), \ j \in [N] \ .$$

  - find $\bar{\alpha}_{n+1}$ optimal for v^{n+1}:

$$\bar{\alpha}_{n+1}(j) \in \text{Argmax}_{\alpha \in \mathcal{A}} f(v^{n+1}; j, \alpha) \ .$$

- It generalizes Newton algorithm
- v^n is nonincreasing.
- If $\mathcal{A}$ is finite, it converges in finite time to the solution.

Policy iterations for games

Now

$$[f(u)]_j = \sup_{\alpha\in\mathcal{A}} f(u;j,\alpha) := \inf_{\beta\in\mathcal{B}} (r_j^{\alpha,\beta} + \sum_{k\in[n]} (P_{jk}^{\alpha,\beta} u_k)) \ .$$

and $u \to f(u;j,\alpha)$ is non linear.

- Assume the non linear system

  $$v = f^{\bar{\alpha}}(v), \quad \text{with } f^{\bar{\alpha}}(v)_j := f(v;j,\bar{\alpha}(j)),\ j \in [N]$$

  has a unique solution for any strategy $\bar{\alpha}$ of Max, then solving it with Policy Iteration, one obtains the policy iteration of (Hoffman and Karp 66, indeed introduced in the ergodic case).
- Assume they have a possibly non unique solution, then the nested and the global policy iterations may cycle. To avoid this, one need to use a method similar to that of (Denardo& Fox 68) in the one-player ergodic case.

Accurate policy iterations for games

In the previous case:

- It suffices to fix the values of $\bar{\alpha}_n(j)$ as much as possible (that is when they are already optimal)
- and to choose for v^{n+1} the nondecreasing limit:

$$v^{n+1} = \lim_{k\to\infty} (f^{\bar{\alpha}_n})^k(v^n) \ .$$

- This limit is the unique solution of the restricted system:

$$v_j = (v^n)_j, j \in C, \quad v_j = (f^{\bar{\alpha}_n}(v))_j, j \notin C$$

  where C is the set of critical nodes of the concave map $f^{\bar{\alpha}_n}$ defined as in (Akian, Gaubert 2003). This system can be solved again by a policy iteration for one-player.
- When the game is deterministic, $f^{\bar{\alpha}_n}$ is min-plus linear, and the set of critical nodes is the usual one defined in max-plus spectral theory. It is the analogue of the Aubry or Mather sets. See in that case (Cochet, Gaubert, Gunawardena 99).
- See (Cochet-Terrasson, Gaubert 2006) for general mean-payoff games.

Numerical results of Policy iteration for mean-payoff stochastic games with algebraic multigrid methods (A., Detournay)

Solve the stationnary Isaacs equation:

$$-\rho + \varepsilon \Delta v + \max_{\alpha \in \mathcal{A}} (\alpha \cdot \nabla v) + \min_{\beta \in \mathcal{B}} (\beta \cdot \nabla v) + \|x\|_2^2 = 0 \text{ on } (-1/2, 1/2)^2$$

with Neuman boundary conditions.
Take

$$\mathcal{A} := B_\infty(0,1)$$

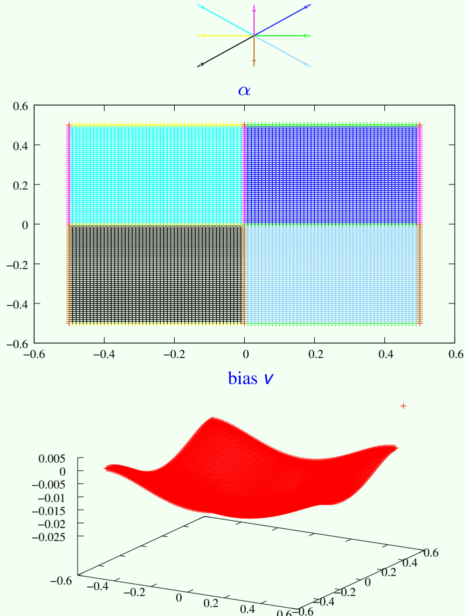
and

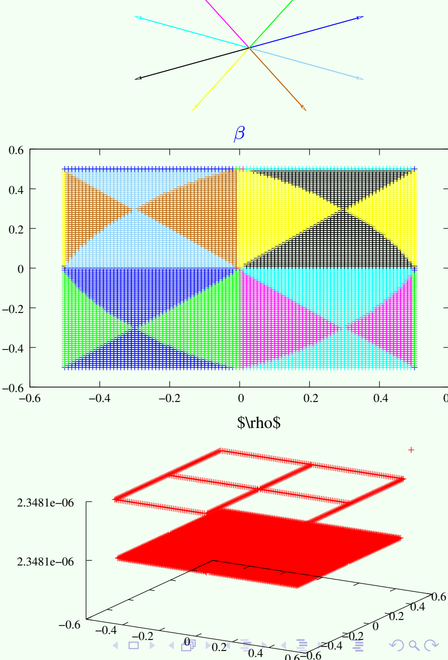
$$\mathcal{B} := \{(0,0), (\pm 1, \pm 2), (\pm 2, \pm 1)\}$$

or

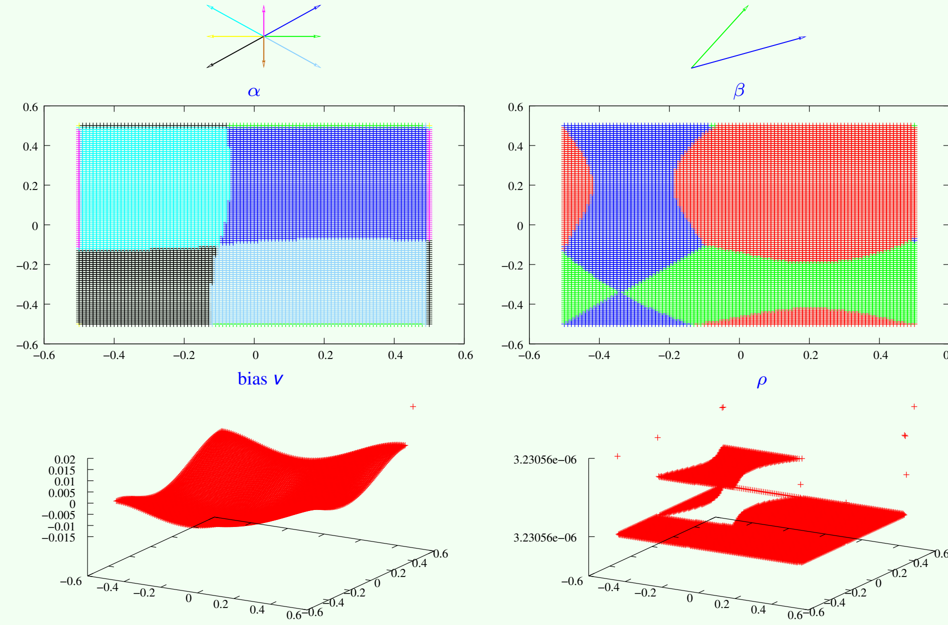
$$\mathcal{B} := \{(0,0), (1,2), (2,1)\} \quad .$$

Case $\mathcal{B} := \{(0,0),(\pm 1,\pm 2),(\pm 2,\pm 1)\}$





Case $\mathcal{B} := \{(0,0),(1,2),(2,1)\}$

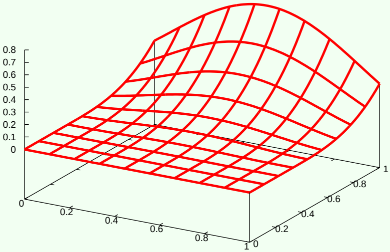


Variational inequalities problem (VI)

Optimal stopping time for first player

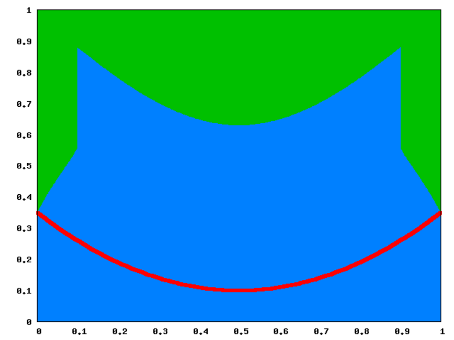
$$\begin{cases} max\left[\Delta v - 0.5\,\|\nabla v\|_2^2 + f, \phi - v\right] = 0 \text{ in } \Omega \\ v = \phi \text{ on } \partial\Omega \end{cases}$$

with solution on $\Omega = [0,1] \times [0,1]$ given by



Max chooses between play or stop ($\sharp\mathcal{A} = 2$) and receives ϕ when he stops
Min leads to $\|\nabla v\|_2^2$

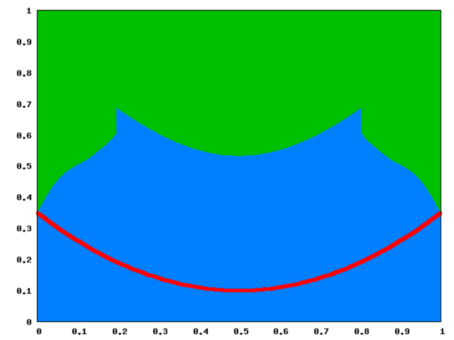
VI with 129×129 points grid



iterations = 100

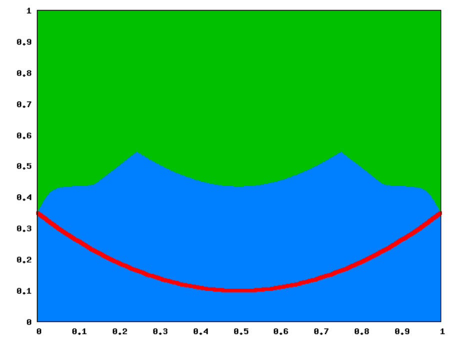
VI with 129×129 points grid

iterations = 200



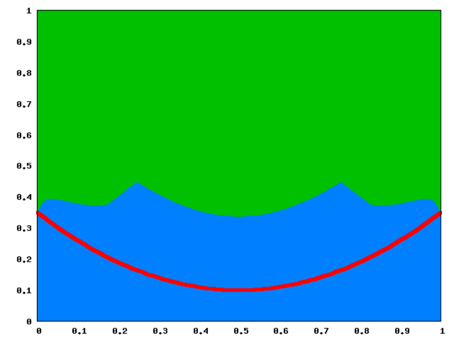
VI with 129 × 129 points grid

iterations = 300



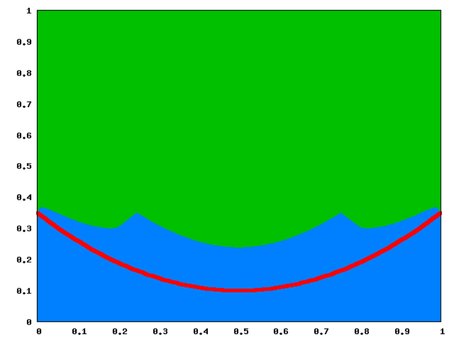
VI with 129×129 points grid

iterations = 400



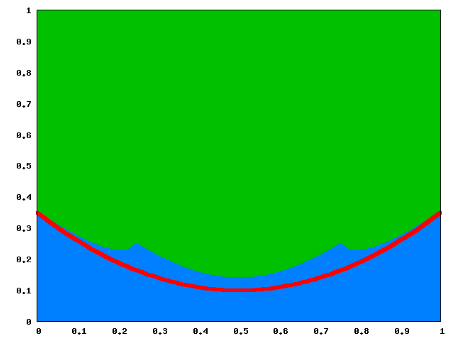
VI with 129×129 points grid

iterations = 500

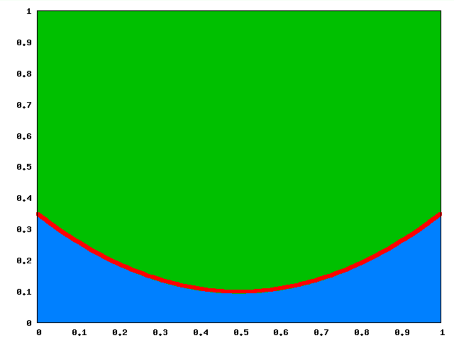


VI with 129 × 129 points grid

iterations = 600



VI with 129×129 points grid



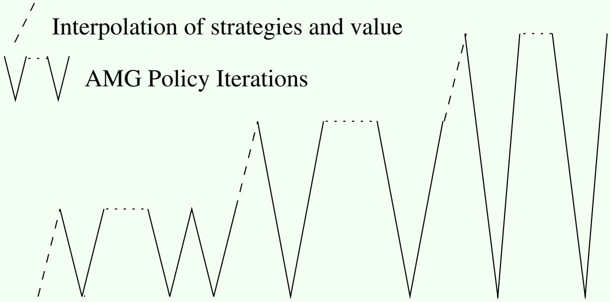
iteration 700!
in $\approx$ 8148 seconds
slow convergence

Policy iterations bounded by $\sharp\{$possible policies$\}$
$\rightarrow$ can be exponential in N

like Newton $\rightarrow$ improve with good initial guess? $\rightarrow$ *FMG*

Full Multilevel AMGπ

define the problem on each coarse grid Ω^H



interpolation of value v and strategies α, β
stopping criterion for AMGπ $\|r\|_{L^2} < cH^2$ (with $c = 0.1$)

Full multilevel AMGπ

$\Omega = [0,1] \times [0,1]$, 1025 nodes in each direction
Ω^H coarse grids (number of nodes in each direction)
$n =$current iteration from Max, $k =$ number of iterations from Min

Ω^H	n	k	$\|r\|_\infty$	$\|r\|_{L_2}$	$\|e\|_\infty$	$\|e\|_{L_2}$	cpu time s
3	1	1	$2.17e-1$	$2.17e-1$	$1.53e-1$	$1.53e-1$	$<< 1$
3	2	1	$1.14e-2$	$1.14e-2$	$3.30e-2$	$3.30e-2$	$<< 1$
5	1	2	$2.17e-4$	$8.26e-5$	$3.02e-2$	$1.71e-2$	$<< 1$
9	1	2	$4.99e-3$	$1.06e-3$	$1.65e-2$	$7.99e-3$	$<< 1$
9	2	1	$2.68e-3$	$5.41e-4$	$1.66e-2$	$8.15e-3$	$<< 1$
9	3	1	$2.72e-4$	$5.49e-5$	$1.68e-2$	$8.30e-3$	$<< 1$
513	1	1	$2.57e-7$	$4.04e-9$	$3.15e-4$	$1.33e-4$	2.62
1025	1	1	$1.31e-7$	$1.90e-9$	$1.57e-4$	$6.63e-5$	$1.17e+1$
1025	2	1	$6.77e-8$	$5.83e-10$	$1.57e-4$	$6.62e-5$	$2.11e+1$

Again max-plus algebra:

- Full multilevel scheme can make policy iteration faster and efficient!
- Can we generalize it for stochastic games with finite state space?
- Mean of game operators leads to an exponential number of actions at lower levels, so need to reduce the number of elements in a max-plus linear combination, this is a max-plus projection.
- Recall: policy iteration for games is exponential (O. Friedmann 09), and finding a polynomial time algorithm for zero-sum game is an open problem.