A very complicated proof of the minimax theorem

Jonathan Borwein FRSC FAAS FAA FBAS

CARMA

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Centre for Computer Assisted Research Mathematics and its Applications
The University of Newcastle, Australia

EViMS

http://carma.newcastle.edu.au/meetings/evims/

http://www.carma.newcastle.edu.au/jon/minimax.pdf

For 2014 Presentations

Revised 15-06-14

Contents: It is always worthwhile revisiting ones garden

Abstract

The justly celebrated *von Neumann minimax theorem* has many proofs. I will briefly discuss four or five of these approaches.

Abstract

The justly celebrated *von Neumann minimax theorem* has many proofs. I will briefly discuss four or five of these approaches.
Then I shall reproduce the most complex one I am aware of.

Abstract

The justly celebrated *von Neumann minimax theorem* has many proofs. I will briefly discuss four or five of these approaches.
Then I shall reproduce the most complex one I am aware of.

This provides a fine didactic example for many courses in convex analysis or functional analysis.

Abstract

The justly celebrated *von Neumann minimax theorem* has many proofs. I will briefly discuss four or five of these approaches.
Then I shall reproduce the most complex one I am aware of.

This provides a fine didactic example for many courses in convex analysis or functional analysis.

This will also allow me to discuss some lovely basic tools in convex and nonlinear analysis.

- Companion paper to appear in new journal of *Minimax Theory and its Applications* and is available at `http://www.carma.newcastle.edu.au/jon/minimax.pdf`.

Contents

- We work in a real Banach space with norm dual X^* or indeed in Euclidean space, and adhere to notation in [1, 2]. We also mention general Hausdorff topological vector spaces [10].

- We work in a real Banach space with norm dual X^* or indeed in Euclidean space, and adhere to notation in [1, 2]. We also mention general Hausdorff topological vector spaces [10].

The classical von Neumann minimax theorem is:

Theorem (Concrete von Neumann minimax theorem (1928))

Let A be a linear mapping between Euclidean spaces E and F. Let $C \subset E$ and $D \subset F$ be nonempty compact and convex. Then

$$d := \max_{y\in D} \min_{x\in C} \langle Ax, y\rangle = \min_{x\in C} \max_{y\in D} \langle Ax, y\rangle =: p. \tag{1}$$

In particular, this holds in the economically meaningful case where both C and D are *mixed strategies* – simplices of the form

$$\Sigma := \{z\colon \sum_{i\in I} z_i = 1, z_i \geq 0, \ \forall\ i \in I\}$$

for finite sets of indices I.

Contents

More generally we have:

Theorem (Von Neumann-Fan minimax theorem)

Let X and Y be Banach spaces. Let $C \subset X$ be nonempty and convex, and let $D \subset Y$ be nonempty, weakly compact and convex. Let $g : X \times Y \to \mathbb{R}$ be convex with respect to $x \in C$ and concave and upper-semicontinuous with respect to $y \in D$, and weakly continuous in y when restricted to D. Then

$$d := \max_{y \in D} \inf_{x \in C} g(x,y) = \inf_{x \in C} \max_{y \in D} g(x,y) =: p. \quad (2)$$

To deduce the concrete Theorem from this theorem we simply consider

$$g(x,y) := \langle Ax, y \rangle.$$

Contents

Various proof techniques

In my books and papers I have reproduced a variety of proofs of the general and concrete Theorems. All have their benefits and additional features:

- The original proof via *Brouwer's fixed point theorem* [1, §8.3] and more refined subsequent algebraic-topological treatments such as the *KKM principle* [1, §8.1, Exer. 15].

Various proof techniques

In my books and papers I have reproduced a variety of proofs of the general and concrete Theorems. All have their benefits and additional features:

- The original proof via *Brouwer's fixed point theorem* [1, §8.3] and more refined subsequent algebraic-topological treatments such as the *KKM principle* [1, §8.1, Exer. 15].
- Tucker's proof of the concrete (simplex) Theorem via *schema and linear programming* [12].

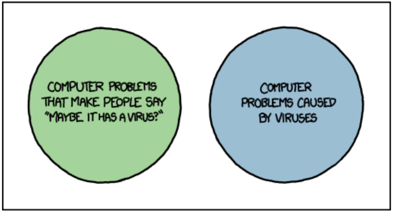
Various proof techniques

In my books and papers I have reproduced a variety of proofs of the general and concrete Theorems. All have their benefits and additional features:

- The original proof via *Brouwer's fixed point theorem* [1, §8.3] and more refined subsequent algebraic-topological treatments such as the *KKM principle* [1, §8.1, Exer. 15].
- Tucker's proof of the concrete (simplex) Theorem via *schema and linear programming* [12].
- From a *compactness and Hahn Banach separation—or subgradient—argument* [4], [2, §4.2, Exer. 14], [3, Thm 3.6.4].
  - This approach also yields *Sion's convex- concave-like minimax theorem*, see [2, Thm 2.3.7] and [11] which contains a nice early history of the minimax theorem.

- From *Fenchel's duality theorem* applied to indicator functions and their conjugate support functions see [1, §4.3, Exer. 16], [2, Exer. 2.4.25] in Euclidean space, and in generality [1, 2, 3]. Bauschke and Combettes discuss this in Hilbert space.

- From *Fenchel's duality theorem* applied to indicator functions and their conjugate support functions see [1, §4.3, Exer. 16], [2, Exer. 2.4.25] in Euclidean space, and in generality [1, 2, 3]. Bauschke and Combettes discuss this in Hilbert space.
  - In J.M. Borwein and C. Hamilton, "Symbolic Convex Analysis: Algorithms and Examples," *Math Programming*, **116** (2009), 17–35, we show that much of this theory can be implemented in a computer algebra system.



SCAT illustrated

http://carma.newcastle.edu.au/ConvexFunctions/links.html

e^{e^x} has conjugate $y(\log W(y) - W(y) - 1/W(y))$ (Lambert W)

```
> f11:=convert(exp(exp(x)),PWF);
```

$$f11 := \left\{ e^{e^x} \quad all(x) \right.$$

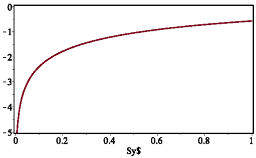
```
> g11:=Conj(f11,y);
```

$$g11 := \begin{cases} \infty & y < 0 \\ -1 & y = 0 \\ -\dfrac{y\left(\mathrm{LambertW}(y)^2 - \mathrm{LambertW}(y)\ln(y) + 1\right)}{\mathrm{LambertW}(y)} & 0 < y \end{cases}$$

```
> sdg11:=SubDiff(g11);
```

$$sdg11 := \begin{cases} \{\ \} & y < 0 \\ \{\ \} & y = 0 \\ \{-\mathrm{LambertW}(y) + \ln(y)\} & 0 < y \end{cases}$$

```
> Plot(sdg11,y=-1..1,view=[0..1,-5..0],axes=boxed,labels=["$y$",""]
);
```



- In the reflexive setting the role of C and D is entirely symmetric. More generally, we should need to introduce the weak* topology and choose not to do so here.

- In the reflexive setting the role of C and D is entirely symmetric. More generally, we should need to introduce the weak* topology and choose not to do so here.

About 35 years ago while first teaching convex analysis and conjugate duality theory, I derived the proof in Section 3, that seems still to be the most abstract and sophisticated I know.

- In the reflexive setting the role of C and D is entirely symmetric. More generally, we should need to introduce the weak* topology and choose not to do so here.

About 35 years ago while first teaching convex analysis and conjugate duality theory, I derived the proof in Section 3, that seems still to be the most abstract and sophisticated I know.

- I derived it in order to illustrate the power of functional-analytic convex analysis as a mode of argument.

- In the reflexive setting the role of C and D is entirely symmetric. More generally, we should need to introduce the weak* topology and choose not to do so here.

About 35 years ago while first teaching convex analysis and conjugate duality theory, I derived the proof in Section 3, that seems still to be the most abstract and sophisticated I know.

- I derived it in order to illustrate the power of functional-analytic convex analysis as a mode of argument.
- I really do not *now* know if it was original at that time . But I did *discover* it in Giaquinto's [6, p. 50] attractive encapsulation:

> *In short, discovering a truth is coming to believe it in an independent, reliable, and rational way.*

- Once a result is discovered, one may then look for a more direct proof.

- Once a result is discovered, one may then look for a more direct proof.
- When first hunting for certainty it is reasonable to use whatever tools one possess.

- Once a result is discovered, one may then look for a more direct proof.
- When first hunting for certainty it is reasonable to use whatever tools one possess.
  - For example, I have often used the *Pontryagin maximum principle* [7] of optimal control theory to discover an inequality for which I subsequently find a direct proof, say from Jensen-like inequalities [2].

- Once a result is discovered, one may then look for a more direct proof.
- When first hunting for certainty it is reasonable to use whatever tools one possess.
  - For example, I have often used the *Pontryagin maximum principle* [7] of optimal control theory to discover an inequality for which I subsequently find a direct proof, say from Jensen-like inequalities [2].

So it seemed fitting to write the proof down for the first issue of the new journal *Minimax Theory and its Applications* dedicated to all matters minimax.

Contents

Needed Tools

I enumerate the prerequisite tools, sketching only the final two as they are less universally treated.

1. Hahn-Banach separation If $C \subset X$ is closed and convex in a Banach space and $x \in X \setminus C$ there exists $\varphi \neq 0$ in X^* such that

$$\boxed{\varphi(x) > \sup_{x \in C} \varphi(x)}$$

as I learned from multiple sources including [7, 8].

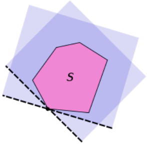
Needed Tools

I enumerate the prerequisite tools, sketching only the final two as they are less universally treated.

1. Hahn-Banach separation If $C \subset X$ is closed and convex in a Banach space and $x \in X \setminus C$ there exists $\varphi \neq 0$ in X^* such that

$$\boxed{\varphi(x) > \sup_{x \in C} \varphi(x)}$$

as I learned from multiple sources including [7, 8].



Support and separation

We need only the Euclidean case which follows from *existence and characterisation of the best approximation* of a point to a closed convex set [1, §2.1, Exer. 8].

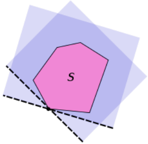
Needed Tools

I enumerate the prerequisite tools, sketching only the final two as they are less universally treated.

1. Hahn-Banach separation If $C \subset X$ is closed and convex in a Banach space and $x \in X \setminus C$ there exists $\varphi \neq 0$ in X^* such that

$$\boxed{\varphi(x) > \sup_{x \in C} \varphi(x)}$$

as I learned from multiple sources including [7, 8].



Support and separation

We need only the Euclidean case which follows from *existence and characterisation of the best approximation* of a point to a closed convex set [1, §2.1, Exer. 8].

2. Lagrangian duality for the *abstract convex programme*, see [1, 2, 3], and [5, 8] for the standard formulation, that I learned first from Luenberger [7].

Contents

Theorem (Lagrange Multipliers)

Suppose that $C \subset X$ is convex, $f : X \to R$, is convex and $G : X \to Y$ ordered by a closed convex cone S with nonempty norm interior is S-convex. Suppose that *Slater's condition* holds:

$$\exists\, \widehat{x} \in X \text{ with } G(\widehat{x}) \in -\operatorname{int} S.$$

Then, the programme

$$p := \inf\{f(x) : G(x) \leq_S 0, x \in C\} \tag{3}$$

has a *Lagrange multiplier* $\lambda \in S^+ := \{\mu : \mu(s) \geq 0,\ \forall s \in S\}$ so that

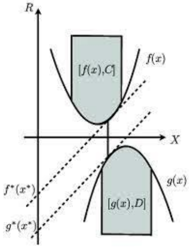
$$p := \inf_{x \in C} f(x) + \lambda(G(x)). \tag{4}$$

If, moreover, $p = G(x_0)$ for a feasible x_0 then *complementary slackness* obtains: $\lambda(G(x_0)) = 0$, while $G(x_0) \leq_S 0$ and $\lambda \geq_{S^+} 0$.

- In Luenberger this result is derived directly from the Separation theorem.

- In Luenberger this result is derived directly from the Separation theorem.
- In [1, 2] it is derived from the nonemptyness of of the subdifferential of a continuous convex function, from Fenchel duality, and otherwise (all being equivalent).

- In Luenberger this result is derived directly from the Separation theorem.
- In [1, 2] it is derived from the nonemptyness of of the subdifferential of a continuous convex function, from Fenchel duality, and otherwise (all being equivalent).



convex-concave Fenchel duality

To handle equality constraints, one needs to use cones with empty interior and to relax Slater's condition, via Fenchel duality as in [1, §4.3], [2, §4.4] or [3, Thm. 4.4.3].

Contents

Riesz-Markov-Kakutani representation theorem

3. (1909-1938-41) For a (locally) compact Hausdorff space Ω the continuous function space, also Banach algebra and Banach lattice:

> *$C(\Omega)$, in the maximum norm, has dual $M(\Omega)$ consisting of all signed regular Borel measures on Ω.*

as I learned from Jameson, Luenberger [7] for $\Omega := [a,b]$, Rudin [10] and Royden [9].

- Moreover, the positive dual functionals correspond to positive measures, as follows from the lattice structure.

Contents

Vector integration

4. The concept of a *weak vector integral*, as I learned from Rudin [10, Ch. 3]. Given a measure space (Q,μ) and a Hausdorff topological vector space Y, and a weakly integrable function[1] $F : Q \to Y$ the integral $y := \int_Q F(x)\mu(dx)$ is said to exist weakly if for each $\varphi \in Y^*$ we have

$$\varphi(y) = \int_Q \varphi(F(x))\,\mu(\mathrm{dx}), \tag{5}$$

and the necessarily unique value of $y = \int_Q F(x)\,\mu(\mathrm{dx})$ defines the *weak integral* of F.

[1] That is, for each dual functional φ, the function $x \mapsto \varphi(F(x))$ is integrable with respect to μ.

In [10, Thm. 3.27], Rudin establishes existence of the weak integral for a Borel measure on a compact Hausdorff space Q, when F is continuous and $D := \overline{\mathrm{conv}}\, F(Q)$ is compact. Moreover, when μ is a probability measure $\int_Q F(x)\,\mu(\mathrm{dx}) \in \overline{\mathrm{conv}}\,\mathrm{F(Q)}$.

Proof

To show existence of y it is sufficient, since D is compact, to show that, for a probability measure μ, (5) can be solved simultaneously in D for any finite set of linear functionals $\{\varphi_1, \varphi_2, \ldots, \varphi_n\}$.

In [10, Thm. 3.27], Rudin establishes existence of the weak integral for a Borel measure on a compact Hausdorff space Q, when F is continuous and $D := \overline{\mathrm{conv}}\, F(Q)$ is compact. Moreover, when μ is a probability measure $\int_Q F(x)\,\mu(\mathrm{dx}) \in \overline{\mathrm{conv}}\,\mathrm{F(Q)}$.

Proof

To show existence of y it is sufficient, since D is compact, to show that, for a probability measure μ, (5) can be solved simultaneously in D for any finite set of linear functionals $\{\varphi_1, \varphi_2, \ldots, \varphi_n\}$. We do this by considering $T := (\varphi_1, \varphi_2, \ldots, \varphi_n)$ as a linear mapping from Y into $\mathbb{R}^n$. Consider

$$\mathbf{m} := \left(\int_Q \varphi_1(F(x))\,\mu(\mathrm{dx}), \ldots, \int_{\mathrm{Q}} \varphi_{\mathrm{n}}(\mathrm{F(x)})\,\mu(\mathrm{dx}) \right)$$

and use the Euclidean space version of the Separation theorem to deduce a contradiction if $\mathbf{m} \notin \overline{\mathrm{conv}}\, T(F(Q))$.

In [10, Thm. 3.27], Rudin establishes existence of the weak integral for a Borel measure on a compact Hausdorff space Q, when F is continuous and $D := \overline{\mathrm{conv}}\, F(Q)$ is compact. Moreover, when μ is a probability measure $\int_Q F(x)\,\mu(\mathrm{dx}) \in \overline{\mathrm{conv}}\,\mathrm{F(Q)}$.

Proof

To show existence of y it is sufficient, since D is compact, to show that, for a probability measure μ, (5) can be solved simultaneously in D for any finite set of linear functionals $\{\varphi_1, \varphi_2, \ldots, \varphi_n\}$. We do this by considering $T := (\varphi_1, \varphi_2, \ldots, \varphi_n)$ as a linear mapping from Y into $\mathbb{R}^n$. Consider

$$\mathbf{m} := \left(\int_Q \varphi_1(F(x))\,\mu(\mathrm{dx}), \ldots, \int_{\mathrm{Q}} \varphi_{\mathrm{n}}(\mathrm{F(x)})\,\mu(\mathrm{dx}) \right)$$

and use the Euclidean space version of the Separation theorem to deduce a contradiction if $\mathbf{m} \notin \overline{\mathrm{conv}}\, T(F(Q))$. Since $\overline{\mathrm{conv}}\, T(F(Q)) = T(D)$ we are done. ■

Contents

Existence and properties of a barycentre

5. We need also the concept of the *barycentre* of a non-empty weakly compact convex set D in a Banach space, with respect to a Borel probability measure μ. As I learned from Choquet and Rudin [10, Ch. 3]), the barycentre (centre of mass)

$$b_D(\mu) := \int_D y\,\mu(\mathrm{dy})$$

exists and lies in D.
This is a special case of the discussion in part 4.

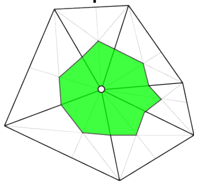
Existence and properties of a barycentre

5. We need also the concept of the *barycentre* of a non-empty weakly compact convex set D in a Banach space, with respect to a Borel probability measure μ. As I learned from Choquet and Rudin [10, Ch. 3]), the barycentre (centre of mass)

$$b_D(\mu) := \int_D y\,\mu(\mathrm{dy})$$

exists and lies in D.

This is a special case of the discussion in part 4.



Barycentre and Voronoi regions

For a polyhedron P with equal masses of $1/n$ at each of the n extreme points $\{e_i\}_{i=1}^n$ this is just

$$b_P = \frac{1}{n}\sum_{i=1}^n e_i.$$

Contents

Proof of the minimax theorem

We now provide the promised complicated proof.

Proof. We first note that always $p \geq d$, this is *weak duality*. We proceed to show $d \geq p$.

1. We observe that, on adding a dummy variable,

$$\boxed{p = \inf_{x \in C}\{r\colon g(x,y) \leq r, \text{ for all } y \in D, r \in \mathbb{R}\}.}$$

Proof of the minimax theorem

We now provide the promised complicated proof.

Proof. We first note that always $p \geq d$, this is *weak duality*. We proceed to show $d \geq p$.

1. We observe that, on adding a dummy variable,

$$\boxed{p = \inf_{x \in C}\{r\colon g(x,y) \leq r, \text{ for all } y \in D, r \in \mathbb{R}\}.}$$

2. Define a vector function $G\colon X \times \mathbb{R} \to C(D)$ by

$$G(x,r)(y) := g(x,y) - r.$$

- This is legitimate because g is continuous in the y variable.

Proof of the minimax theorem

We now provide the promised complicated proof.

Proof. We first note that always $p \geq d$, this is *weak duality*. We proceed to show $d \geq p$.

1. We observe that, on adding a dummy variable,

$$p = \inf_{x \in C}\{r\colon g(x,y) \leq r, \text{ for all } y \in D, r \in \mathbb{R}\}.$$

2. Define a vector function $G\colon X \times \mathbb{R} \to C(D)$ by

$$G(x,r)(y) := g(x,y) - r.$$

- This is legitimate because g is continuous in the y variable.
- We take the cone S to be the non-negative continuous functions on D and check that G is S-convex because g is convex in x for each $y \in D$.

An abstract convex programme

We now have an abstract convex programme

$$p = \inf\{r\colon G(x,r) \leq_S 0, x \in C\}, \tag{6}$$

where the objective is the linear function $f(x,r) = r$.

Fix $0 < \varepsilon < 1$. Then there is some $\widehat{x} \in C$ with $g(\widehat{x}, y) \leq p + \varepsilon$ for all $y \in D$. We deduce that

$$G(\widehat{x}, p-2) \leq -\mathbf{1} \in -\operatorname{int} S$$

where **1** is the constant function in $C(D)$. Thence Slater's condition (1953) holds.

3. The Lagrange multiplier theorem assures a multiplier $\lambda \in S^+$. By the Riesz representation of $C(D)^*$, given above, we may treat λ as a measure and write

$$r + \int_D (g(x,y) - r)\,\lambda(\mathrm{dy}) \geq \mathrm{p}$$

for all $x \in C$ and all $r \in \mathbb{R}$. Since C is nonempty and r is arbitrary we deduce that $\lambda(D) = \int_D \lambda(\mathrm{dy}) = 1$ and so λ is a probability measure on D.

4. Consequently, we derive that for all $x \in C$

$$\int_D g(x,y)\,\lambda(\mathrm{dy}) \geq \mathrm{p}.$$

5. We now consider the barycentre $\widehat{b} := b_D(\lambda)$ guaranteed in the prior section.

Since λ is a probability measure and g is continuous in y we deduce, using the integral form of *Jensen's inequality*[2] for the concave function $g(x,\cdot)$, that for each $x \in C$

$$g(x, \int_D y\,\lambda(\mathrm{dy})) \geq \int_{\mathrm{D}} \mathrm{g(x,y)}\,\lambda(\mathrm{dy}) \geq \mathrm{p}.$$

But this says that

$$d = \sup_{y\in D} \inf_{x\in C} g(x,y) \geq \inf_{x\in C} g(x,\widehat{b}) \geq p.$$

This show the left-hand supremum is attained at the barycentre of the Lagrange multiplier. This completes the proof. ■

[2] Fix $k := y \to g(x,y)$ and observe that for any affine majorant a of of k we have $k(\widehat{b}) = \inf_{a\geq k} a(\widehat{b}) = \inf_{a\geq k} \int_D a(y)\,\lambda(\mathrm{dy}) \geq \int_{\mathrm{D}} \mathrm{k(y)}\,\lambda(\mathrm{dy})$, where the leftmost equality is a consequence of upper semicontinuity of k, and the second since λ is a probability and we have a weak integral.

Extensions

At the expense of some more juggling with the formulation, this proof can be adapted to allow for $g(x,y)$ only to be upper-semicontinuous in y, as is assumed in Fan's theorem.

- One looks at continuous perturbations maximizing G.

Extensions

At the expense of some more juggling with the formulation, this proof can be adapted to allow for $g(x,y)$ only to be upper-semicontinuous in y, as is assumed in Fan's theorem.

- One looks at continuous perturbations maximizing G.

> *I will be glad if I have succeeded in impressing the idea that it is not only pleasant to read at times the works of the old mathematical authors, but this may occasionally be of use for the actual advancement of science.* (Constantin Carathéodory in 1936 speaking to the MAA)

Contents

Conclusions

Too often we teach the principles of functional analysis and of convex analysis with only the most obvious applications in the subject we know the most about—be it operator theory, partial differential equations, or optimization and control.

Conclusions

Too often we teach the principles of functional analysis and of convex analysis with only the most obvious applications in the subject we know the most about—be it operator theory, partial differential equations, or optimization and control.

- But important mathematical results do not arrive in such prepackaged form. In my books, [1, 2, 3], my coauthors and I have tried in part to redress this imbalance. It is in this spirit that I offer this modest article.

Acknowledgements. Thanks are due to many but especially to Heinz Bauschke, Adrian Lewis, Jon Vanderwerff, Jim Zhu, and Brailey Sims who have been close collaborators on matters relating to this work over many years.

Contents

Key References

- J.M. Borwein and A.S. Lewis, *Convex Analysis and Nonlinear Optimization. Theory and Examples*, CMS (Canadian Mathematical Society) Springer-Verlag, New York, Second extended edition, 2005. Paperback, 2009.
- J.M. Borwein and J.D. Vanderwerff, *Convex Functions*, Cambridge University Press, 2010.
- J.M. Borwein and Qiji Zhu, *Techniques of Variational Analysis*, CMS/Springer, 2005.
- J.M. Borwein and D. Zhuang, "On Fan's minimax theorem," *Math. Programming*, **34** (1986), 232–234.
- S. Boyd, and L. Vandenberghe, *Convex Optimization*, **317**, Cambridge Univ. Press, 2004.
- M. Giaquinto, *Visual Thinking in Mathematics. An Epistemological Study*, Oxford University Press, New York, 2007.
- D.G. Luenberger, *Optimization by vector space methods*, John-Wiley & Sons, 1972.
- R.T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.
- H.L. Royden, *Real Analyis*, Prentice Hall, First edition, 1963. Third edition, 1988.
- W. Rudin, *Functional Analyisis*, First edition, McGraw Hill, 1973.
- M. Sion, "On general minimax theorems", *Pacific J. Math.* **8**, Number 1 (1958), 1–186.
- A.W. Tucker, "Solving a matrix game by linear programming." *IBM J. Res. Develop.* **4** (1960). 507–517.

The end with some fractal desert

The end with some fractal desert

Thank you