A VERY COMPLICATED PROOF OF THE MINIMAX THEOREM

JONATHAN M. BORWEIN

Abstract. The justly celebrated von Neumann minimax theorem has many proofs. Here I reproduce the most complex one I am aware of. This provides a fine didactic example for many courses in convex analysis or functional analysis.

1. Introduction

We work primarily in a real Banach space with norm dual X^* or indeed in Euclidean space, and adhere to the notation in [4, 5]. From time to time we mention more general Hausdorff topological vector spaces [20].

The justly celebrated von Neumann minimax theorem has many proofs.

Theorem 1 (Concrete von Neumann minimax theorem)**.** *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 sets. 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 simplices of the form $\Sigma := \{z\colon \sum_{i \in I} z_i = 1, z_i \geq 0, \ \forall\ i \in I\}$ for a finite set of indices I.

More generally we have:

Theorem 2 (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*

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

Date: `April 20, 2014`.

2010 *Mathematics Subject Classification.* Primary 45Q05; Secondary 49N45 90C25 49N15 90C90.

Key words and phrases. Minimax theorem, Hahn-Banach theorem, Fenchel duality theorem, weak integrals, barycentre.

Research partially supported by ARC Grant #DP1093769.

To deduce Theorem 1 from Theorem 2 we simply consider $g(x, y) := \langle Ax, y \rangle$.

1.1. **Various proof techniques.** In my books and papers I have reproduced a variety of proofs of Theorem 2. All have their benefits and additional features:

(1) The original proof via *Brouwer's fixed point theorem* [4, §8.3] and more refined subsequent algebraic-topological treatments such as the *KKM principle* [4, §8.1, Exer. 15].
(2) Tucker's proof of Theorem 1 via schema and linear programming [23].
(3) From a compactness and Hahn Banach separation or subgradient argument [7], [5, §4.2, Exercise 14], [6, Theorem 3.6.4]. This approach also yields *Sion's convex-concave-like minimax theorem*, see [5, Thm 2.3.7] and [22] which contains a nice early history of the minimax theorem.
(4) From Fenchel's duality theorem applied to indicator functions and their conjugate support functions see [4, §4.3, Exercise 16], [5, Exercise 2.4.25] or [10, 18] in Euclidean space, [1] in the Hilbert space setting, and in generality [4, 5, 6]. In [3] we show that much of this theory can be implemented in a computer algebra system.

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 that 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 [15, 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. When first hunting for certainty it is reasonable to use whatever tools one possess. For example, I have often used the *Pontryagin maximum principle* [17] of optimal control theory to discover an inequality for which I subsequently find a direct proof, say from Jensen-like inequalities [5].

So it seems fitting to write the proof down in the first issue of the new journal M*inimax Theory and its Applications* dedicated to minimax matters.

2. Prerequisite Tools

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

2.1. **The Hahn-Banach separation theorem.** If $C \subset X$ is a closed convex subset of a Banach space and $x \in X \setminus C$ then there exists $\phi \neq 0$ in X^* such that $\phi(x) > \sup_{x \in C} \phi(x)$, as I learned from multiple sources including [16, 17, 18, 21]. We shall need explicitly only the Euclidean case which can be deduced directly from the existence and characterisation of the best approximation of a point to a closed convex set [4, §2.1, Exer. 8].

2.2. **Lagrange duality.** *Lagrangian duality* for the *abstract convex programme*, see[4, 5, 6], and [8, 18] for the standard formulation, as I learned first from Luenberger [17].

Theorem 3 (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

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

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

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

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

In Luenberger this result is derived directly from the Separation theorem. In [4, 5] it is derived from the nonemptyness of of the subdifferential of a continuous convex function, from Fenchel duality, and otherwise (all being equivalent). In order to handle equality constraints, one needs to use cones with empty interior and to relax Slater's condition, see [2] and via Fenchel duality in [4, §4.3], [5, §4.4] or [6, Thm. 4.4.3].

2.3. **Riesz representation theorem.** For a 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 [16], Luenberger [17] for $\Omega := [a, b]$, Rudin [20] and Royden [19]. See also [11]. Moreover, the positive dual functionals correspond to positive measures, as follows from the lattice structure.

2.4. **Vector integration.** The concept of a *weak vector integral*, as I learned from Rudin [20, 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 $\phi \in Y^*$ we have

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

and the necessarily unique value of $y = \int_Q F(x)\, \mu(\mathrm{dx})$ defines the *weak integral* of F. In [20, Thm. 3.27], Rudin establishes the existence of the weak integral for any 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 then $\int_Q F(x)\, \mu(\mathrm{dx}) \in \overline{\mathrm{conv}}\, \mathrm{F(Q)}$.

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 $\{\phi_1, \phi_2, \ldots, \phi_n\}$. We do this by considering $T := (\phi_1, \phi_2, \ldots, \phi_n)$ as a linear mapping from Y into $\mathbb{R}^n$. Consider

$$\mathbf{m} := \left(\int_Q \phi_1(F(x))\, \mu(\mathrm{dx}), \ldots, \int_{\mathrm{Q}} \phi_{\mathrm{n}}(\mathrm{F(x)})\, \mu(\mathrm{dx}) \right)$$

and use the Euclidean space version of the Separation theorem of Section 2.1 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. □

2.5. **The existence and properties of a barycentre.** 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 [9] and Rudin [20, Ch. 3]), the barycentre

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

exists and lies in D. This is a special case of the discussion in Section 2.4. In the case of a polyhedron P with equal masses of $1/n$ at each of the n extreme points $\{e_i\}_{i=1}^n$ this is just $b_P = \frac{1}{n} \sum_{i=1}^n e_i$.

3. The proof of Theorem 2

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

(1) We observe that, on adding a dummy variable,

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

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

We now have an abstract convex programme

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

where the objective is the linear function $f(x,r)=r$. Fix $0<\varepsilon<1$. Then there is some $\widehat{x}\in C$ with $g(\widehat{x},y)\le p+\varepsilon$ for all $y\in D$. We deduce that

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

where $\mathbf{1}$ is the constant function in $C(D)$. Thence Slater's condition holds.

(3) We may now appeal to Theorem 3 and obtain a Lagrange multiplier $\lambda\in S^+$. By the Riesz representation of $C(D)^*$, given in Section 2.3, we may treat λ as a measure and write

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

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 $x\in C$

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

(5) We now consider the barycentre $\widehat{b} := b_D(\lambda)$ guaranteed in section 2.5. 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 $x\in C$

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

But this says that $d=\sup_{y\in D}\inf_{x\in C} g(x,y)\ge \inf_{x\in C} g(x,\widehat{b})\ge p$.

In particular, 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 $k(\widehat{b}) = \inf_{a\ge k} a(\widehat{b}) = \inf_{a\ge k}\int_D a(y)\,\lambda(\mathrm{dy})\ge\int_{\mathrm{D}}\mathrm{k(y)}\,\lambda(\mathrm{dy})$, 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.

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.

4. Concluding Comments

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, [4, 5, 6], 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.

References

[1] H.H. Bauschke and P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS-Springer Books, 2011.

[2] J.M. Borwein, "Multivalued convexity and optimization: a unified approach to inequality and equality constraints," *Math. Programming*, **13** (1977), 183–199.

[3] J.M. Borwein and C. Hamilton, "Symbolic Convex Analysis: Algorithms and Examples," *Math Programming*, **116** (2009), 17–35.

[4] 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.

[5] J.M. Borwein and J.D. Vanderwerff, *Convex Functions*, Cambridge University Press, 2010.

[6] J.M. Borwein and Qiji Zhu, *Techniques of Variational Analysis*, CMS/Springer, 2005.

[7] J.M. Borwein and D. Zhuang, "On Fan's minimax theorem," *Math. Programming*, **34** (1986), 232–234.

[8] S. Boyd, and L. Vandenberghe, *Convex Optimization*, 317, Cambridge University Pres, 2004.

[9] G. Choquet, *Lectures on Analysis* (Three volumes). English publication, Benjamin/Cummings, 1976.

[10] Kenneth Davidson and Allan P. Donsig, *Real Analysis and Applications* Springer Undergraduate Texts in Mathematics, 2008.

[11] M. Fabian, P. Habala, P. Hájek, V. Montesinos and V. Zizler, *Banach Space Theory*, CMS-Springer-Verlag, 2010.

[12] W. Fenchel, "On conjugate convex functions," *Canad. J. Math.* 1 (1949), 73–77.

[13] W. Fenchel, *Convex Cones, Sets and Funmctions* Lecture Notes, Princeton 1953.

[14] D. Gale, "A geometric duality theorem with economic applications," *The Review of Economic Studies*, **34** (1967), 19—24.

[15] M. Giaquinto, *Visual Thinking in Mathematics. An Epistemological Study*, Oxford University Press, New York, 2007.

[16] G.J.O. Jameson, *Topology and Normed Spaces*, Chapman and Hall, 1974.

[17] D.G. Luenberger, *Optimization by vector space methods*, John-Wile & Sons, 1972.
[18] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.
[19] H.L. Royden, *Real Analyis*, Prentice Hall, First edition, 1963. Third edition, 1988.
[20] W. Rudin, *Functional Analyisis*, First edition, McGraw Hill, 1973.
[21] S. Simons, *From Hahn-Banach to Monotonicity*, Lecture Notes in Mathematics, **1693**, Springer-Verlag, 2008.
[22] M. Sion, "On general minimax theorems", *Pacific J. Math.* **8**, Number 1 (1958), 1–186.
[23] A.W. Tucker, "Solving a matrix game by linear programming." *IBM J. Res. Develop.* **4** (1960). 507–517.

Laureate Professor and Director CARMA, University of Newcastle, Callaghan NSW 2308, Australia.

E-mail address: `jonathan.borwein@newcastle.edu.au`