© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

CONTENTS

Figures and Algorithms xiii

Preface xv

1 Introduction 1

1.1 Minimizers at Infinity 2

1.2 Extending First-Order Optimality Conditions 5

1.3 A Quick Introduction to Astral Space 9

1.4 Minimization Revisited 13

1.5 Overview of the Book 16

PART I. PRELIMINARIES AND BACKGROUND

2 General Preliminaries and Review of Linear Algebra 21

2.1 Notational Conventions 21

2.2 Working with $\pm\infty$ 23

2.3 Linear Algebra 26

2.4 Orthogonal Complement, Projection Matrices, Pseudoinverse 28

2.5 Zero Vectors, Zero Matrices, Zero-Dimensional Euclidean Space 30

3 Review of Topology 32

3.1 Topological Space, Open Sets, Base, Subbase 32

3.2 Closed Sets, Neighborhoods, Dense Sets, Separation Properties 33

3.3 Subspace Topology 34

3.4 Continuity, Homeomorphism, Compactness 35

3.5 Sequences, First-Countability, Second-Countability 37

3.6 Product Topology, Tychonoff's Theorem, Weak Topology 39

3.7 Lower Semicontinuity on First-Countable Spaces 41

4 Review of Convex Analysis 47

4.1 Convex Sets 47

4.2 Hull Operators 48

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

4.3 Affine Sets 49
4.4 Closure, Interior, Relative Interior 49
4.5 Cones 51
4.6 Separation Theorems 53
4.7 Faces and Polyhedral Sets 56
4.8 Convex Functions 57
4.9 Lower Semicontinuity and Continuity 62
4.10 Conjugacy 64
4.11 Subgradients, Subdifferentials, and Directional Derivatives 70
4.12 Recession Cone and Constancy Space 73

PART II. ASTRAL SPACE AND ASTRAL POINTS

5 Constructing Astral Space 79
5.1 The Construction 79
5.2 The Coupling Function 84
5.3 Astral Points 86

6 Astral Topology 88
6.1 Astral Points as Functions 88
6.2 Characterizing Functional Form 93
6.3 Astral Space, Astrons, Leftward Addition 99
6.4 A Sufficient Condition for Convergence of Sequences 105
6.5 First-Countability 107
6.6 Not Second-Countable and Not Metrizable 109

7 Linear Maps and the Matrix Representation of Astral Points 113
7.1 Linear and Affine Maps 113
7.2 Astral Points in Matrix Form 120
7.3 Canonical Representation 126

8 Representation and Sequences 129
8.1 Characterizing All Sequences Converging to an Astral Point 129
8.2 Strong Equivalence and Span-Bound Sequences 133
8.3 Convergence to Points Represented by Specific Expressions 135

9 Astral-Point Decompositions and the Structure of Astral Space 140
9.1 Icons 140

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

9.2 Galaxies and Prefix Structure 143
9.3 Dominant Directions 148
9.4 Convergence in Direction 151
9.5 Comparison to Cosmic Space 153

PART III. EXTENDING FUNCTIONS TO ASTRAL SPACE

10 Lower Semicontinuous Extension 159
10.1 Definition and Basic Properties 159
10.2 Continuity 164
10.3 Working with Epigraphs 167
10.4 Astronic Reductions 171

11 Astral Conjugates and Biconjugates 176
11.1 Definitions and Basic Properties 176
11.2 Conjugates and Biconjugates of Extensions 182
11.3 Functional Operations Involving Linear Maps 183

12 Astral Closedness of Extensions 187
12.1 Barrier Cone of a Function 188
12.2 Exponential Composition 191
12.3 Reductions and Conjugacy 194
12.4 Reduction-Closedness 197
12.5 A Dual Characterization of Reduction-Closedness 202

13 Calculus Rules for Extensions 207
13.1 Scalar Multiple 207
13.2 Sums of Functions 207
13.3 Composition with an Affine Map 213
13.4 Linear Image of a Function 217
13.5 Pointwise Supremum of a Collection of Functions 220
13.6 Composition with a Nondecreasing Function 223

PART IV. CONVEX SETS AND CONES

14 Convex Sets 231
14.1 Halfspaces, Segments, Outer Hull, Convexity 231
14.2 Astral Polytopes and Polyhedral Sets 240

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

14.3 Convex Hull 245

15 Constructing and Operating on Convex Sets 247

15.1 Convexity Under Affine Transformations 247

15.2 Segments with a Finite Endpoint 251

15.3 Interior of a Convex Set and Convex Hull of an Open Set 256

15.4 Closure of a Convex Set and Convex Hull of a Closed Set 258

16 Sequential Operations 261

16.1 Sequential Sum of Sets 261

16.2 Sequential Multiples 271

16.3 Sequential Convex and Conic Combinations 274

17 Astral Cones 280

17.1 Definition and Basic Properties 280

17.2 Convex Astral Cones 284

17.3 Conic Hull Operations 287

17.4 Astral Halflines 292

18 Separation Theorems 296

18.1 Separating Convex Sets 296

18.2 Astral Dual Separation of Sets in $\mathbb{R}^n$ 306

19 Representational Closure and Astral Polarity 317

19.1 Extending a Standard Convex Cone to Astral Space 317

19.2 The Astral Conic Hull of a Polyhedral Cone 322

19.3 Astral Polar Cones 325

20 Astral Linear Subspaces 331

20.1 Definitions and Characterizations 331

20.2 Astral Span and Bases 336

20.3 Orthogonal Complements 342

PART V. FUNCTIONS ON ASTRAL SPACE

21 Convex Functions 349

21.1 Definition and Properties 349

21.2 Constructing and Operating on Convex Functions 353

21.3 Lower Envelope and Lower Support Functions 358

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

22 Astral Closedness of Functions in General 365
22.1 Astral Primal Closedness 365
22.2 Lower Semicontinuity and Primal Biconjugacy 370
22.3 Astral Dual Closedness 374

23 Minimizers and Their Structure 382
23.1 Astral Recession Cone 382
23.2 Recursive Characterization and Generation of Minimizers 390
23.3 Dual Characterization 397
23.4 Reductions Revisited 401

24 Universal Reduction and Canonical Minimizers 404
24.1 The Universal Reduction 404
24.2 Coercivity of the Universal Reduction 410
24.3 Universal Reducers and Canonical Minimizers 417

25 The Structure of Minimizers in Some Particular Cases 422
25.1 Astral Rank of Minimizers 422
25.2 Empirical Risk Minimization 427

26 Continuity 439
26.1 Characterizing Where $\bar{f}$ Is Continuous 440
26.2 Conditions for Continuity 446

PART VI. DIFFERENTIAL THEORY

27 Defining Astral Subgradients 453
27.1 Astral Primal Subgradients 453
27.2 Astral Dual Subgradients 462

28 Characterizations of Astral Subgradients 471
28.1 Conditions for Astral Subgradients 471
28.2 Characterizing Primal Subgradients with Linear Tilts 476
28.3 Linear-Tilt Characterization for an Extension 479
28.4 Inverse Relation Between Primal and Dual Subgradients 484

29 More Properties of Astral Subdifferentials 490
29.1 Nonemptiness and Disjointness of Dual Subdifferentials 490
29.2 Monotonicity, Convexity, and Closedness of Subdifferentials 495

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

29.3 Subdifferentials of Indicator Functions and Normal Cones 497

30 Calculus Rules for Astral Subgradients 502
30.1 Scalar Multiple 502
30.2 Sums of Functions 503
30.3 Composition with an Affine Map 505
30.4 Reductions 507
30.5 Linear Image of a Function 514
30.6 Pointwise Supremum of a Collection of Functions 515
30.7 Composition with a Nondecreasing Function 520

PART VII. OPTIMALITY AND OPTIMIZATION

31 Fenchel Duality 531
31.1 Generalizing Fenchel Duality to Astral Space 531
31.2 Constrained Optimization 537
31.3 Application to Astral Matrix Games 542

32 KKT Conditions and Lagrange Multipliers 547
32.1 The Standard Setting 547
32.2 Defining Astral Optimality 551
32.3 Astral Optimality Conditions 554

33 Differential Continuity and Descent Methods 567
33.1 Convergence of Subgradient Sequences 567
33.2 Existence of Convergent Subgradient Sequences 573
33.3 Convergence of Iterative Methods 577

34 Exponential-Family Distributions 585
34.1 The Standard Setting 585
34.2 Extending to Astral Space 588
34.3 Conjugate and Astral Subgradients 590
34.4 Maximum Likelihood and Maximum Entropy 595
34.5 Astral Galaxies and Faces of the Marginal Polytope 599

Bibliography 609
Notation Index 613
Subject Index 615
Author Index 631

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

1

Introduction

Convex functions play a fundamental role in economics, operations research, statistics, machine learning, information theory, game theory, and many other disciplines of engineering and science (Boyd and Vandenberghe, 2004). For instance, many objective functions that are minimized in operations research and machine learning, such as those that arise in portfolio optimization and in fitting predictive models to data, are convex. Convexity is an attractive structural property, because minimization of convex functions is algorithmically very well-behaved, with a vast array of efficient algorithms available (Boyd and Vandenberghe, 2004; Nemirovski and Yudin, 1983; Nesterov, 2018).

For instance, for many convex functions, it is possible simply to "go down the slope" of the function (that is, in the direction of the negative gradient), which, for appropriately chosen step sizes, is guaranteed to converge to a minimizer (see Figure 1.1, left, for an illustration in one dimension).

However, there are important examples of well-behaved functions (say, lower-bounded and differentiable on all of $\mathbb{R}^n$), for which no algorithm can find a minimizer, because no minimizer exists. This can happen even in the simplest of cases, like the exponential function $f(x) = e^x$ shown in Figure 1.1 (right). In this case, going down the slope decreases function values, but the algorithm diverges, with its iterates going to (negative) infinity. This function has no finite minimizer, and can only be minimized in a limit as $x \to -\infty$.

Similarly, in multiple dimensions, there are examples of convex functions arising in machine learning, statistics, and economics, which can only be minimized by following a sequence to infinity. In these applications, we are often interested in various properties

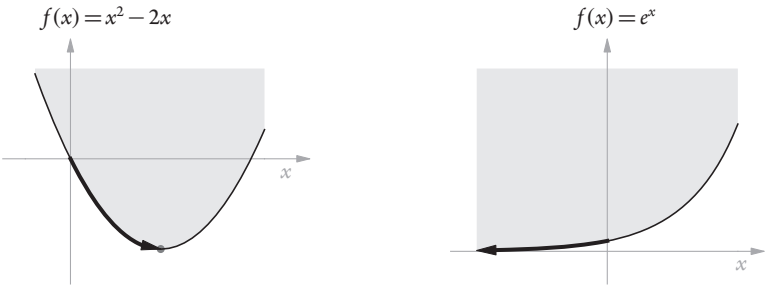


FIGURE 1.1. *Minimization of convex functions in one dimension. Left:* An example where the approach of going down the slope (in the direction of negative gradient) converges to a minimizer. *Right:* An example where the approach diverges to (negative) infinity.

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

of such minimizing sequences, but the analysis of these sequences can be somewhat laborious and specialized to each case (see, for instance, Collins et al., 2002; Zhang and Yu, 2005; and Section 3.6 of Wainwright and Jordan, 2008).

This book seeks to move past such specialized analyses. Our overarching goal is to develop a more complete theory of convex functions by extending $\mathbb{R}^n$ with suitable "points at infinity." Ideally, with such an extension, there would be no need to analyze various types of divergent sequences; instead, we could work directly in the extended space, which would contain new "infinite points" to which these sequences converge.

In later chapters, we construct a specific extension of $\mathbb{R}^n$ called *astral space*, study its properties, and develop an expansive theory of convex functions on this space. Before embarking on that journey, we aim in this introductory chapter to bring out some key issues that arise in constructing such an extension, while also giving an overview of our approach and introducing some of the themes (like differential theory and conjugacy) that will be examined in detail later in the book.

1.1 Minimizers at Infinity

The task of extending $\mathbb{R}^n$ with points at infinity is deceptively simple in one dimension, where we only need to add $+\infty$ and $-\infty$. The resulting extension is called the *extended real line* and denoted $\overline{\mathbb{R}} = [-\infty, +\infty]$. It is far from obvious how to generalize this concept to multiple dimensions. To develop some intuition about what is required of such an extension, we next look at several examples of convex functions and their minimizing sequences.

Before proceeding, we make some brief notational comments: Vectors $\mathbf{x}$ in $\mathbb{R}^n$ are usually written in bold, with components x_i given in italics so that $\mathbf{x} = [x_1, \dots, x_n]^\top$. We use $t = 1, 2, \dots$ as the index of sequences, unless noted otherwise, and write $(\mathbf{x}_t)$ to mean the sequence $\mathbf{x}_1, \mathbf{x}_2, \dots$. Limits and convergence are taken as $t \to +\infty$, unless stated otherwise. For example, $\lim \mathbf{x}_t$ means $\lim_{t\to\infty} \mathbf{x}_t$, and $\mathbf{x}_t \to \mathbf{x}$ means $\mathbf{x}_t$ converges to $\mathbf{x}$ as $t \to +\infty$. (More will be said about notation in Section 2.1.)

Example 1.1 (Log-sum-exp). We first consider the log-sum-exp function, which arises, for example, when fitting a multinomial model by maximum likelihood. Specifically, let

$$\Delta_n = \left\{ \mathbf{p} \in [0,1]^n : \sum_{i=1}^n p_i = 1 \right\} \tag{1.1}$$

be the probability simplex in $\mathbb{R}^n$, and let $\hat{\mathbf{p}}$ be any point in Δ_n. Consider minimization of the function $f : \mathbb{R}^n \to \mathbb{R}$ defined, for $\mathbf{x} \in \mathbb{R}^n$, by

$$f(\mathbf{x}) = \ln\left(\sum_{i=1}^n e^{x_i}\right) - \hat{\mathbf{p}} \cdot \mathbf{x}.$$

As is standard, f can be rewritten as

$$f(\mathbf{x}) = -\sum_{i=1}^n \hat{p}_i \ln\left(\frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}\right) = -\sum_{i=1}^n \hat{p}_i \ln p_i(\mathbf{x}), \tag{1.2}$$

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

where $p_i(\mathbf{x}) = e^{x_i} / \sum_{j=1}^{n} e^{x_j}$, so $\mathbf{p}(\mathbf{x}) = [p_1(\mathbf{x}), \ldots, p_n(\mathbf{x})]^\top$ is a probability vector in Δ_n. For this reformulation of f, it can be proved that

$$\inf_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) = -\sum_{i=1}^{n} \hat{p}_i \ln \hat{p}_i \tag{1.3}$$

and that $\mathbf{x}$ minimizes f if and only if $\mathbf{p}(\mathbf{x}) = \hat{\mathbf{p}}$ (see Proposition 34.6a). So, if $\hat{p}_i > 0$ for all i, the minimum is attained by the vector $\mathbf{x}$ with entries $x_i = \ln \hat{p}_i$.

However, if $\hat{p}_i = 0$ for some i, then there is no $\mathbf{x} \in \mathbb{R}^n$ that attains the minimum. In this case, the infimum in Eq. (1.3) is instead reached by any sequence $(\mathbf{x}_t)$ for which $\mathbf{p}(\mathbf{x}_t) \to \hat{\mathbf{p}}$. For example, if $n = 3$ and $\hat{\mathbf{p}} = \left[0, \frac{1}{3}, \frac{2}{3}\right]^\top$, then the sequence $\mathbf{x}_t = \left[-t, \ln \frac{1}{3}, \ln \frac{2}{3}\right]^\top$ converges to the infimum as $t \to +\infty$. ■

This example suggests that, in extending $\mathbb{R}^n$, it perhaps would suffice to allow individual coordinates to take on values of $\pm\infty$. In other words, we could consider a Cartesian product of the extended reals, $(\overline{\mathbb{R}})^n$. The minimizer of Example 1.1 would then be $\left[-\infty, \ln \frac{1}{3}, \ln \frac{2}{3}\right]^\top$. However, the next example shows that this is not enough because minimization may take us in a direction that is not aligned with coordinate axes.

Example 1.2 (Diagonal valley). Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined, for $\mathbf{x} \in \mathbb{R}^2$, as

$$f(\mathbf{x}) = f(x_1, x_2) = e^{-x_1} + (x_2 - x_1)^2 \tag{1.4}$$

(see Figure 1.2). The infimum is obtained in the limit of any sequence $(\mathbf{x}_t)$ that satisfies $x_{t,1} \to +\infty$ while also $x_{t,2} - x_{t,1} \to 0$. One such sequence is $\mathbf{x}_t = [t, t]^\top$, which goes to infinity in the direction of the vector $\mathbf{v} = [1, 1]^\top$ (see Figure 1.2, right). On this sequence, $f(\mathbf{x}_t) = e^{-t} \to 0$.

If we were to just work in $(\overline{\mathbb{R}})^2$, we would find that $\mathbf{x}_t \to [+\infty, +\infty]^\top$. However, to minimize f, the direction in which $(\mathbf{x}_t)$ goes to infinity is critical, and that direction is not

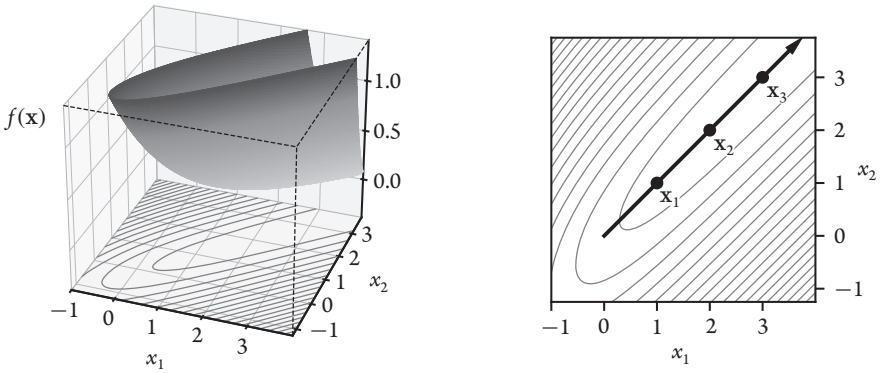


FIGURE 1.2. *Diagonal valley.* The function f from Example 1.2 (surface plot on the left, contour plot on the right). On the right, we also show the first few points of the sequence $\mathbf{x}_t = [t, t]^\top$, which minimizes f in the limit as it follows a halfline to infinity.

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

represented by limit points in $(\overline{\mathbb{R}})^2$. For instance, the sequence $\mathbf{x}'_t = [2t, t]^\top$, which goes to infinity in the direction of $[2, 1]^\top$, also converges to $[+\infty, +\infty]^\top$ in $(\overline{\mathbb{R}})^2$, but it fails to minimize f since $f(\mathbf{x}'_t) \to +\infty$. ■

So perhaps we should extend $\mathbb{R}^n$ with "limit points" of sequences going to infinity in various directions in $\mathbb{R}^n$. For instance, the sequences $(\mathbf{x}_t)$ and $(\mathbf{x}'_t)$, which follow along halflines with different directions, would then also have different limits. Such limit points are considered by Rockafellar and Wets (1998, Section 3), who call them "direction points" and develop the topology and properties of the resulting "cosmic space." This concept was also studied by Hansen and Dupin (2001) under the name "enlarged space." Another related concept is that of "ideal points" in the real projective plane, where a separate ideal point is introduced for each class of parallel lines (see Cox et al., 2025, Chapter 8).

However, these abstractions are too coarse to characterize minimizing sequences in the diagonal valley example. For example, in the cosmic space formalism, the sequence $\mathbf{x}''_t = [t + 1, t]^\top$ converges to the same limit point as $\mathbf{x}_t = [t, t]^\top$, since they both go to infinity in the same direction, but the sequence $(\mathbf{x}''_t)$ fails to minimize f since $f(\mathbf{x}''_t) \to 1$.

Maybe we just need to consider an "offset" in addition to the direction of a sequence, and consider limit points corresponding to following a halfline from a specified starting point in a specific direction. All the minimizing sequences in the preceding examples can be written as $\mathbf{x}_t = t\mathbf{v} + \mathbf{q}$ for a suitable choice of $\mathbf{v}$ and $\mathbf{q}$, so each indeed proceeds along a halfline with a starting point $\mathbf{q}$ in the direction of $\mathbf{v}$. It turns out that this is still not enough, as we show in the next example:

Example 1.3 (Two-speed exponential). Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined, for $\mathbf{x} \in \mathbb{R}^2$, as

$$f(\mathbf{x}) = f(x_1, x_2) = e^{-x_1} + e^{-4x_2 + x_1^2}$$

(see Figure 1.3). The infimum is obtained in the limit of any sequence $(\mathbf{x}_t)$ that satisfies $x_{t,1} \to +\infty$ while also $-4x_{t,2} + x_{t,1}^2 \to -\infty$. This means f cannot be minimized along any halfline; rather, $x_{t,2}$ must go to $+\infty$ faster than $x_{t,1}^2/4$. One such sequence is $\mathbf{x}_t = [t, t^2/2]^\top$, on which $f(\mathbf{x}_t) \to 0$ (see Figure 1.3, right). ■

The above examples show that the task of adding "infinite points" is subtle already in $\mathbb{R}^2$. So perhaps we should just stick to sequences to maintain the broadest flexibility. The downside is that sequences seem to only be indirect proxies for the underlying "infinite points," so working with sequences makes it harder to discover their structural properties. Moreover, we hope that by adding the extra "infinite points," the theory of convex functions will become more complete, for example, by ensuring that all lower-bounded and continuous convex functions (like all those from our examples) have minimizers.

We aim to extend $\mathbb{R}^n$ analogously to how the set of rational numbers is extended to obtain reals, or the set of reals is extended to obtain complex numbers. When moving from rational numbers to reals to complex numbers, basic arithmetic operations naturally extend to the enlarged space, and the enlarged space is more complete; for example, bounded sets of reals have real-valued infima (which is not true for rational numbers), and

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

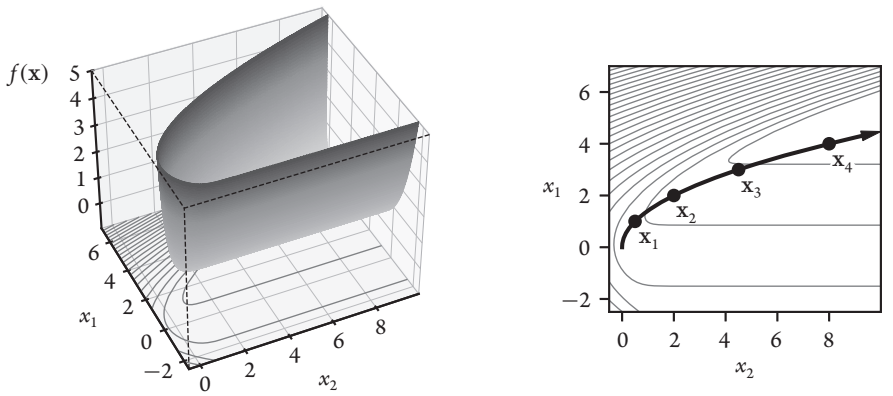


FIGURE 1.3. *Two-speed exponential.* The function f from Example 1.3. On the right, we show the first few points of the sequence $\mathbf{x}_t = [t, t^2/2]^\top$, which minimizes f in the limit as it follows a parabolic curve to infinity.

all polynomials with complex coefficients have complex roots (whereas polynomials with real coefficients might have no real roots).

In the examples above, we considered several ways of extending $\mathbb{R}^n$. With each successive approach, we saw that we could better capture the breadth of possible minimizers of convex functions, but at the cost of the extended space becoming larger and more intricate. This tradeoff is inevitable, and so will require that we seek a middle ground that balances the extent to which we can characterize various minimizers at infinity against our ability to generalize standard analytical tools to the resulting space.

Specifically, we seek an extension of $\mathbb{R}^n$ that will lend itself to natural extensions of the key concepts in convex analysis, like convex sets and functions, conjugacy, subdifferentials, and optimality conditions, but that will also be more complete, for example, when it comes to the existence of minimizers. Our aim is to reveal the structure of how convex functions behave at infinity. We seek to build up the foundations of an expanded theory of convex analysis that will, for instance, make proving the convergence of optimization algorithms rather routine, even if the function being optimized has no finite minimizer.

1.2 Extending First-Order Optimality Conditions

To get a taste of such an extended theory, we next examine how the first-order optimality conditions, which characterize minimizers of convex functions, might generalize to include points at infinity in just one dimension.

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex. One of the fundamental properties of convex functions is that if $\nabla f(\mathbf{x}) = \mathbf{0}$ at a point $\mathbf{x} \in \mathbb{R}^n$ where f is differentiable, then f must be (globally) minimized by $\mathbf{x}$. More generally, $\mathbf{x}$ minimizes f if and only if $\mathbf{0}$ is a subgradient of f at $\mathbf{x}$, meaning geometrically that there exists a horizontal hyperplane that is tangent to f's epigraph at $\mathbf{x}$, as illustrated in Figure 1.4 (left), where $x = 1$ is the minimizer of the depicted function. (Subgradients are generalizations of gradients and will be discussed further in Example 1.8; see also Section 4.11. And the epigraph of a function, defined formally in Eq. 2.2, is the set of points above its graph, for instance, the shaded regions in Figure 1.4.)

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

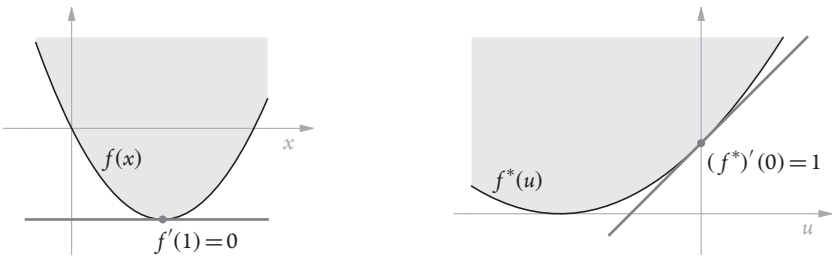


FIGURE 1.4. *A quadratic function and its conjugate.* The function $f(x) = x^2 - 2x$ and its conjugate $f^*(u) = u^2/4 + u + 1$ (calculated using Eq. 1.5). The function f is minimized at $x = 1$, with $f'(1) = 0$. Dually, the derivative of f^* at $u = 0$ is equal to 1. Thus, 1 is the sole subgradient of f^* at 0, demonstrating that subgradients of f^* at the origin correspond to minimizers of f.

There is also a related "dual" characterization of the minimizers of f. Recall that every convex function f has a conjugate $f^* : \mathbb{R}^n \to \overline{\mathbb{R}}$ defined as

$$f^*(\mathbf{u}) = \sup_{\mathbf{x} \in \mathbb{R}^n} \left[\mathbf{x} \cdot \mathbf{u} - f(\mathbf{x}) \right] \tag{1.5}$$

for $\mathbf{u} \in \mathbb{R}^n$, which is itself a convex function. The conjugate is closely related to the problem of minimization of f: If $f^*(\mathbf{0}) < +\infty$, then f is bounded below. Further, under mild conditions on f, it is known that if the conjugate is differentiable at $\mathbf{0}$, then $\nabla f^*(\mathbf{0})$ must be the unique minimizer of f. More generally, the subgradients of f^* at $\mathbf{0}$ are precisely the minimizers of f (as follows from Proposition 4.61ac). This is illustrated in Figure 1.4, which shows a quadratic function f and its conjugate f^*. The function f is minimized at $x = 1$, with its derivative equal to 0; dually, the derivative of the conjugate at $u = 0$ is equal to 1. Since the subgradients of a function at any given point are exactly the slopes of nonvertical tangents of the function's epigraph at that point, the analysis of minimizers of any convex function f can be conducted by analyzing the tangents of the epigraph of f^* at the origin.

Can this style of reasoning—a conjugate, subgradients at the origin, tangents of the epigraph—be lifted to an extended space that includes points at infinity? And can we extend the space in such a way that the abundant structure that these characterizations reflect is still retained? The next example considers how all this might work out for the simple case of the exponential function.

Example 1.4 (Extended exponential). Let $f(x) = e^x$ for $x \in \mathbb{R}$. As already discussed, this function has no finite minimizer. Correspondingly, and consistent with the characterizations just discussed, there is no point $x \in \mathbb{R}$ where the function's derivative is equal to 0. Further, the function's conjugate f^* can be computed to be, for $u \in \mathbb{R}$,

$$f^*(u) = \begin{cases} u \ln u - u & \text{if } u \geq 0, \\ +\infty & \text{if } u < 0, \end{cases} \tag{1.6}$$

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

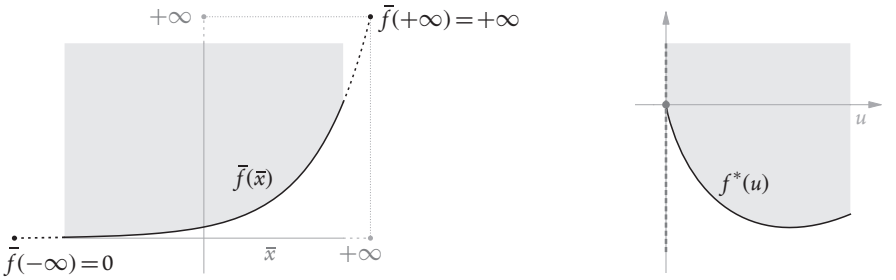


FIGURE 1.5. *Extended exponential and its conjugate.* Both the extension $\bar{f}$ and the conjugate f^* are derived in Example 1.4. *Left:* Although $f(x) = e^x$ has no minimizer, its extension $\bar{f}$ is minimized at $\bar{x} = -\infty$. *Right:* The conjugate f^* is not differentiable at 0, but its epigraph has a vertical tangent there, corresponding to the "infinite rate of decrease" and suggesting that it might be reasonable to define the derivative at 0 to be $-\infty$, which would dually correspond to $-\infty$ being a minimizer of $\bar{f}$.

with the convention $0 \ln 0 = 0$ (see Figure 1.5, right). This function is differentiable at all $u > 0$, with derivative $(f^*)'(u) = \ln u$. However, at $u = 0$, even though the function is finite, it is not differentiable since its epigraph has no nonvertical tangents at that point. This represents a certain incompleteness in the standard theory, though consistent with f having no finite minimizers.

Let us see how all this might change if we extend f to $\overline{\mathbb{R}}$. A natural, continuous extension of f to $\overline{\mathbb{R}}$ is the function $\bar{f} : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ defined as

$$\bar{f}(x) = e^x \text{ when } x \in \mathbb{R},$$

$$\bar{f}(-\infty) = \lim_{x \to -\infty} e^x = 0,$$

$$\bar{f}(+\infty) = \lim_{x \to +\infty} e^x = +\infty$$

(see Figure 1.5, left). Unlike f, which has no minimum, the extension $\bar{f}$ appears better-behaved because it does attain its minimum (at $-\infty$). Further, if we could somehow define the "derivative" of $\bar{f}$ at $-\infty$, we might reasonably expect that it should be equal to 0, since it would seem that an extended version of the horizontal x-axis is tangent to the epigraph of $\bar{f}$ at $-\infty$ (as suggested by Figure 1.5, left).

Next, let us attempt to generalize the notion of conjugacy to the extension $\bar{f}$. Analogous to Eq. (1.5), this conjugate $\bar{f}^* : \mathbb{R} \to \overline{\mathbb{R}}$ could perhaps be defined, for $u \in \mathbb{R}$, as

$$\bar{f}^*(u) = \sup_{\bar{x}} \big[\bar{x} \cdot u - \bar{f}(\bar{x}) \big],$$

where it is understood that the supremum here is only over those $\bar{x} \in \overline{\mathbb{R}}$ for which the difference $\bar{x} \cdot u - \bar{f}(\bar{x})$ is defined (that is, not the difference of $+\infty$ with itself, or $-\infty$ with itself). This definition of $\bar{f}^*$ differs superficially from the one that we introduce later in Section 11.1, but the two are nonetheless equivalent. It can be checked that the function

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

$\overline{f}^*$ turns out to be identical to the standard conjugate f^* given in Eq. (1.6). Therefore, as noted above, this function's epigraph has no nonvertical tangents at 0.

Nonetheless, this function's epigraph does have a *vertical* tangent at 0 (see Figure 1.5, right). It would be natural to represent this vertical tangent by positing the "derivative" at this point to be equal to $-\infty$, corresponding to how $\overline{f}^*$ is "decreasing infinitely fast" at 0 (faster than any finite slope). If we could next apply a theorem saying that the (suitably generalized) subgradients of $\overline{f}^*$ at the origin are minimizers of $\overline{f}$, we would thereby obtain that $-\infty$ minimizes $\overline{f}$, as is actually the case. ■

The above example suggests just one area, besides the existence of minimizers, in which the theory of an extended space and extended functions might give rise to a more complete convex analysis. In this case, we considered generalizing subgradients to take on infinite values so that their extended form might represent not only nonvertical, but also vertical tangents of the epigraphs of convex functions. If this can be done successfully, we might expect all suitable convex functions to have generalized subgradients everywhere that they are finite, which is not the case in standard convex analysis.

In the example above, we considered a function in one dimension and suggested defining a generalized derivative that takes values in the extended reals, $\overline{\mathbb{R}}$, and in particular, is equal to $\pm\infty$ at points where the function is increasing or decreasing infinitely fast. It is considerably less clear how gradients or subgradients of a multivariate function might be similarly generalized since, at a given point, the function may be increasing (or decreasing) infinitely fast in some directions, but only increasing at a finite rate in other directions. We will return to this question in Example 1.8, where we will see how such generalized subgradients can themselves be represented using points in an extension of $\mathbb{R}^n$ that includes points at infinity.

To be clear, our goal is not to merely extend subgradients to include vertical tangents; rather, we seek to develop an expansive theory of convex analysis into which extended subgradients would fit like a puzzle piece, together with extended conjugates, optimality conditions, and an array of other fundamental concepts.

In this book, we propose such a theory, extending $\mathbb{R}^n$ to an enlarged topological space called *astral space*. We study the new space's properties, and develop a theory of convex functions on this space. Although astral space includes all of $\mathbb{R}^n$, it is not a vector space, nor even a metric space. However, it is sufficiently well-structured to allow useful and meaningful extensions of such concepts as convexity, conjugacy, and subgradients. We develop these concepts and analyze various properties of convex functions on astral space, including the structure of their minimizers, characterization of continuity, and convergence of descent algorithms.

Although the conceptual underpinnings of astral space are simple, the full formal development is somewhat involved since key topological properties need to be carefully established. As a teaser of what the book is about, we next present a condensed development of astral space. This will allow us to revisit our earlier multidimensional examples and situate them within the theory of functions on astral space. We then finish this introductory chapter with a high-level overview of the entire book.

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

1.3 A Quick Introduction to Astral Space

In constructing astral space, our aim is to derive a topological extension of $\mathbb{R}^n$ in which various "points at infinity" have been added, corresponding to limits of sequences such as those from Examples 1.1, 1.2, and 1.3. In fact, we seek a *compactification* of $\mathbb{R}^n$, in which every sequence has a convergent subsequence. There are many possible compactifications, which differ in how many new points they add and how finely they differentiate among different kinds of convergence to infinity. We would like to add as few points as possible, but the fewer points we add, the fewer functions will be continuous in the new space, because there will be more sequences converging to each new point.

In this work, we choose a tradeoff in which we add as few points as possible while still ensuring that all *linear* functions, which are the bedrock of convex analysis, remain continuous. As a consequence of this property, we will see that "anything linear" is likely to behave nicely when extended to astral space, including such notions as linear maps, hyperplanes, and halfspaces, as well as convex sets, convex functions, conjugates, and subgradients, all of which are based, in their definitions, on linear functions.

Convergence in all directions. To implement this idea, we focus on the limits of all linear functions when evaluated on a given sequence. Every such linear function takes the form $\mathbf{x} \mapsto \mathbf{x} \cdot \mathbf{u}$ for some $\mathbf{u} \in \mathbb{R}^n$, which, when $\mathbf{u}$ is a unit vector, can be viewed geometrically as scalar projection in the direction of $\mathbf{u}$. We define a notion of a well-behaved sequence with respect to linear maps by saying that a sequence $(\mathbf{x}_t)$ in $\mathbb{R}^n$ *converges in all directions* if, for all $\mathbf{u} \in \mathbb{R}^n$, its image $(\mathbf{x}_t \cdot \mathbf{u})$ under such a linear function converges in $\overline{\mathbb{R}}$, meaning $\lim(\mathbf{x}_t \cdot \mathbf{u})$ exists in $\overline{\mathbb{R}}$ for all $\mathbf{u} \in \mathbb{R}^n$.

For example, every sequence that converges in $\mathbb{R}^n$ also converges in all directions, since $\mathbf{x}_t \to \mathbf{x}$ implies, for all $\mathbf{u} \in \mathbb{R}^n$, that $\mathbf{x}_t \cdot \mathbf{u} \to \mathbf{x} \cdot \mathbf{u}$. There are additional sequences that converge in all directions like all those appearing in Examples 1.1, 1.2, and 1.3.

If two sequences $(\mathbf{x}_t)$ and $(\mathbf{y}_t)$ both converge in all directions and also $\lim(\mathbf{x}_t \cdot \mathbf{u}) = \lim(\mathbf{y}_t \cdot \mathbf{u})$ for all $\mathbf{u} \in \mathbb{R}^n$, we say that $(\mathbf{x}_t)$ and $(\mathbf{y}_t)$ are *all-directions equivalent*.

Astral space. All-directions equivalence creates a partition into equivalence classes of the set of sequences that converge in all directions. As formally defined in Chapter 5, to construct *astral space*, we associate an *astral point* with each such equivalence class; this point will then be exactly the common limit of every sequence in the associated equivalence class. We write $\overline{\mathbb{R}^n}$ for n-dimensional astral space, and use bar or overline notation to denote its elements, such as $\overline{\mathbf{x}}$ or $\overline{\mathbf{y}}$.

This definition allows us to naturally extend the inner product. Specifically, for all $\overline{\mathbf{x}} \in \overline{\mathbb{R}^n}$ and all $\mathbf{u} \in \mathbb{R}^n$, we define the *coupling function*

$$\overline{\mathbf{x}} \cdot \mathbf{u} = \lim(\mathbf{x}_t \cdot \mathbf{u}) \in \overline{\mathbb{R}},$$

where $(\mathbf{x}_t)$ is any sequence in the equivalence class associated with $\overline{\mathbf{x}}$. Note that the value of $\overline{\mathbf{x}} \cdot \mathbf{u}$ does not depend on the choice of $(\mathbf{x}_t)$ because all the sequences associated with $\overline{\mathbf{x}}$ are all-directions equivalent and therefore have identical limits $\lim(\mathbf{x}_t \cdot \mathbf{u})$. In fact, the

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

values of these limits uniquely identify each astral point, so every astral point $\overline{\mathbf{x}}$ is uniquely identified by the values of $\overline{\mathbf{x}} \cdot \mathbf{u}$ over all $\mathbf{u} \in \mathbb{R}^n$.

The space $\mathbb{R}^n$ is naturally included in $\overline{\mathbb{R}^n}$, coinciding with equivalence classes of sequences that converge to points in $\mathbb{R}^n$. But there are additional elements, which are said to be *infinite* since they must satisfy $\overline{\mathbf{x}} \cdot \mathbf{u} \in \{-\infty, +\infty\}$ for at least one vector $\mathbf{u}$.

The simplest astral points (other than those in $\mathbb{R}^n$) are called *astrons*, each obtained as the limit of a sequence of the form $(t\mathbf{v})$ for some $\mathbf{v} \in \mathbb{R}^n$. The resulting astron is denoted $\omega\mathbf{v}$. By reasoning about the limit of the sequence $(t\mathbf{v} \cdot \mathbf{u})$, for $\mathbf{u} \in \mathbb{R}^n$, it can be checked that

$$(\omega\mathbf{v}) \cdot \mathbf{u} = \begin{cases} +\infty & \text{if } \mathbf{v} \cdot \mathbf{u} > 0, \\ 0 & \text{if } \mathbf{v} \cdot \mathbf{u} = 0, \\ -\infty & \text{if } \mathbf{v} \cdot \mathbf{u} < 0. \end{cases}$$

If $\mathbf{v} \neq \mathbf{0}$, then the associated astron $\omega\mathbf{v}$ is infinite, and is exactly the limit of a sequence that follows a halfline from the origin in the direction of $\mathbf{v}$.

In one-dimensional astral space, $\overline{\mathbb{R}^1}$, the only astrons are 0, and two others corresponding to $+\infty$ and $-\infty$. In multiple dimensions, besides $\mathbf{0}$, there is a distinct astron $\omega\mathbf{v}$ associated with every unit vector $\mathbf{v}$.

The astron construction can be generalized in a way that turns out to yield all additional astral points, including the limits in Examples 1.1, 1.2, and 1.3.

Example 1.5 (Polynomially graded sequence). Let $\mathbf{q} \in \mathbb{R}^n$, let $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$, for some $k \geq 0$, and consider the sequence

$$\mathbf{x}_t = t^k\mathbf{v}_1 + t^{k-1}\mathbf{v}_2 + \cdots + t\mathbf{v}_k + \mathbf{q} = \sum_{i=1}^{k} t^{k-i+1}\mathbf{v}_i + \mathbf{q}.$$

We can verify that this sequence converges in all directions, and therefore corresponds to an astral point, by calculating the limit of $(\mathbf{x}_t \cdot \mathbf{u})$ for any vector $\mathbf{u} \in \mathbb{R}^n$.

If $k = 0$, then $\mathbf{x}_t = \mathbf{q}$ for all t, so trivially $\mathbf{x}_t \cdot \mathbf{u} \to \mathbf{q} \cdot \mathbf{u}$. Otherwise, if $k \geq 1$, then the evolution of the sequence $(\mathbf{x}_t)$ is dominated by its overwhelmingly rapid growth in the direction of $\mathbf{v}_1$. As a result, if $\mathbf{v}_1 \cdot \mathbf{u} > 0$, then $\mathbf{x}_t \cdot \mathbf{u} \to +\infty$, and if $\mathbf{v}_1 \cdot \mathbf{u} < 0$, then $\mathbf{x}_t \cdot \mathbf{u} \to -\infty$. However, if $\mathbf{v}_1 \cdot \mathbf{u} = 0$, then the term involving $\mathbf{v}_1$ vanishes when considering $\mathbf{x}_t \cdot \mathbf{u}$. On the subspace orthogonal to $\mathbf{v}_1$, the direction of $\mathbf{v}_2$ thus becomes dominant. For such vectors $\mathbf{u}$ that are perpendicular to $\mathbf{v}_1$, we find once again that $\mathbf{x}_t \cdot \mathbf{u}$ converges to $+\infty$ or $-\infty$ if $\mathbf{v}_2 \cdot \mathbf{u} > 0$ or $\mathbf{v}_2 \cdot \mathbf{u} < 0$, respectively. This analysis can be continued, so that we next consider vectors $\mathbf{u}$ in the subspace orthogonal to both $\mathbf{v}_1$ and $\mathbf{v}_2$, where $\mathbf{v}_3$ is dominant, and so on. Eventually, for vectors $\mathbf{u}$ that are orthogonal to all $\mathbf{v}_1, \ldots, \mathbf{v}_k$, we find that $\mathbf{x}_t \cdot \mathbf{u}$ converges to the finite value $\mathbf{q} \cdot \mathbf{u}$.

In summary, this argument shows that the sequence $(\mathbf{x}_t)$ converges in all directions, and its corresponding astral point $\overline{\mathbf{x}}$ is described by

$$\overline{\mathbf{x}} \cdot \mathbf{u} = \begin{cases} +\infty & \text{if } \mathbf{v}_1 \cdot \mathbf{u} = \cdots = \mathbf{v}_{i-1} \cdot \mathbf{u} = 0 \text{ and } \mathbf{v}_i \cdot \mathbf{u} > 0 \text{ for some } i, \\ -\infty & \text{if } \mathbf{v}_1 \cdot \mathbf{u} = \cdots = \mathbf{v}_{i-1} \cdot \mathbf{u} = 0 \text{ and } \mathbf{v}_i \cdot \mathbf{u} < 0 \text{ for some } i, \\ \mathbf{q} \cdot \mathbf{u} & \text{if } \mathbf{v}_1 \cdot \mathbf{u} = \cdots = \mathbf{v}_k \cdot \mathbf{u} = 0. \end{cases} \tag{1.7}$$

■

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

Eq. (1.7) implies that if any of the vectors $\mathbf{v}_j$ in Example 1.5 is a linear combination of the preceding vectors, then it can be dropped without affecting the value of the limits $\lim(\mathbf{x}_t \cdot \mathbf{u}) = \overline{\mathbf{x}} \cdot \mathbf{u}$, so that the resulting sequence (after $\mathbf{v}_j$ has been removed) still converges to the same astral point $\overline{\mathbf{x}}$. In particular, this means, for the purpose of constructing astral points as the limits of sequences, that it suffices in Example 1.5 to consider only linearly independent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$, of which there can be at most n.

Although this example analyzes only one type of sequence, we will later see (by a different style of reasoning) that every sequence that converges in all directions must have a similar structure. Namely, every such sequence must have a most dominant direction $\mathbf{v}_1$ in which it is tending to infinity most rapidly, followed by a second most dominant direction $\mathbf{v}_2$, and so on, with residual convergence to some finite point $\mathbf{q}$. As a result, as we prove in Section 6.3, every astral point in $\overline{\mathbb{R}^n}$ is characterized by Eq. (1.7) for some choice of $\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{q} \in \mathbb{R}^n$ (for some finite $k \geq 0$).

Astral topology. In defining astral space, we also must specify a topology, thus determining, for instance, what it means for a sequence to converge in the space, or for a function to be continuous on the space. We specifically endow astral space with the *astral topology*, formally defined in Section 6.1, which turns out to be the "coarsest" (or "smallest") topology under which all linear maps $\mathbf{x} \mapsto \mathbf{x} \cdot \mathbf{u}$ defined over $\mathbb{R}^n$ can be extended continuously to $\overline{\mathbb{R}^n}$ (yielding the maps $\overline{\mathbf{x}} \mapsto \overline{\mathbf{x}} \cdot \mathbf{u}$). As proved in Theorem 6.14 and Corollary 6.26, astral space has several important properties in this topology.

The first and perhaps most important property is *compactness*, which makes it often easier to work with astral space than with $\mathbb{R}^n$. Because of compactness, every continuous function on astral space attains its minimum.

Another crucial property is called *first-countability*. This property allows us to work mainly with sequences and their limits, just as we would in $\mathbb{R}^n$, rather than the more involved machinery afforded by general topology. First-countability implies, for example, that a function F on $\overline{\mathbb{R}^n}$ is continuous if and only if $F(\overline{\mathbf{x}}_t) \to F(\overline{\mathbf{x}})$ for every sequence $(\overline{\mathbf{x}}_t)$ in $\overline{\mathbb{R}^n}$ converging to some point $\overline{\mathbf{x}} \in \overline{\mathbb{R}^n}$. Although all metric spaces (including $\mathbb{R}^n$) are also first-countable, astral space for $n \geq 2$ is not metrizable, meaning its topology cannot be obtained by any metric imposed on the space (see Section 6.6).

Since astral space is both compact and first-countable, it is also *sequentially compact*, meaning every sequence in $\overline{\mathbb{R}^n}$ has a convergent subsequence, greatly simplifying the analysis of optimization algorithms. Many of our proofs depend critically on this property.

Astral space, as already mentioned, is a compactification of $\mathbb{R}^n$. In addition to being compact, this means also that $\mathbb{R}^n$ (in its usual, familiar topology) is a topological subspace of astral space, and that $\mathbb{R}^n$ is *dense* in $\overline{\mathbb{R}^n}$, implying that every astral point is the limit of some sequence of points in $\mathbb{R}^n$.

In one dimension, $\overline{\mathbb{R}^1}$ is homeomorphic with $\overline{\mathbb{R}}$ (that is, the same topologically). In fact, we later define $\overline{\mathbb{R}^1}$ to be *equal* to $\overline{\mathbb{R}}$.

Convergence in astral space in the astral topology is closely related to the notion of convergence in all directions, on which the space was constructed. In particular, a sequence $(\overline{\mathbf{x}}_t)$ in $\overline{\mathbb{R}^n}$ converges to a point $\overline{\mathbf{x}} \in \overline{\mathbb{R}^n}$ if and only if, for all $\mathbf{u} \in \mathbb{R}^n$, the sequence $(\overline{\mathbf{x}}_t \cdot \mathbf{u})$ converges in $\overline{\mathbb{R}}$ to $\overline{\mathbf{x}} \cdot \mathbf{u}$. In this way, convergence of the sequence $(\overline{\mathbf{x}}_t)$ in n-dimensional

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

astral space can be analyzed by considering convergence of $(\overline{\mathbf{x}}_t \cdot \mathbf{u})$ in one-dimensional $\overline{\mathbb{R}}$ (for all $\mathbf{u} \in \mathbb{R}^n$). Many proofs of convergence are based on this property.

Summarizing:

(a) $\overline{\mathbb{R}^n}$ is compact.
(b) $\overline{\mathbb{R}^n}$ is first-countable.
(c) $\overline{\mathbf{x}}_t \to \overline{\mathbf{x}}$ in $\overline{\mathbb{R}^n}$ if and only if $\overline{\mathbf{x}}_t \cdot \mathbf{u} \to \overline{\mathbf{x}} \cdot \mathbf{u}$ in $\overline{\mathbb{R}}$ for all $\mathbf{u} \in \mathbb{R}^n$.
(d) $\mathbb{R}^n$ is a dense subspace of $\overline{\mathbb{R}^n}$.
(e) $\overline{\mathbb{R}^1}$ is homeomorphic with $\overline{\mathbb{R}}$.

Representing astral points. Although astral space is a topological extension of the vector space $\mathbb{R}^n$, it is not a vector space itself, and astral points cannot be added. The root problem is that the sum of $-\infty$ and $+\infty$ is undefined, making it impossible, for example, to establish the identity $(\overline{\mathbf{x}} + \overline{\mathbf{y}}) \cdot \mathbf{u} = \overline{\mathbf{x}} \cdot \mathbf{u} + \overline{\mathbf{y}} \cdot \mathbf{u}$, which is meaningless when $-\infty$ and $+\infty$ appear on the right-hand side.

While standard addition does not generalize to astral space, a noncommutative variant does. For $\overline{x}, \overline{y} \in \overline{\mathbb{R}}$, we write this operation, called *leftward addition*, as $\overline{x} \dot{+} \overline{y}$. It is the same as ordinary addition except that, when adding $-\infty$ and $+\infty$, the argument on the *left* dominates. Thus,

$$\begin{aligned} (+\infty) \dot{+} (-\infty) &= +\infty, \\ (-\infty) \dot{+} (+\infty) &= -\infty, \\ \overline{x} \dot{+} \overline{y} &= \overline{x} + \overline{y} \text{ in all other cases.} \end{aligned} \tag{1.8}$$

This operation can be extended from $\overline{\mathbb{R}}$ to $\overline{\mathbb{R}^n}$: For $\overline{\mathbf{x}}, \overline{\mathbf{y}} \in \overline{\mathbb{R}^n}$, the leftward sum, written $\overline{\mathbf{x}} \dot{+} \overline{\mathbf{y}}$, is defined to be that unique point in $\overline{\mathbb{R}^n}$ for which

$$(\overline{\mathbf{x}} \dot{+} \overline{\mathbf{y}}) \cdot \mathbf{u} = \overline{\mathbf{x}} \cdot \mathbf{u} \dot{+} \overline{\mathbf{y}} \cdot \mathbf{u}$$

for all $\mathbf{u} \in \mathbb{R}^n$. Such a point must always exist (see Proposition 6.17). While leftward addition is not commutative, it is associative. Scalar multiplication of a vector and standard matrix–vector multiplication also extend to astral space, and are distributive with leftward addition.

All astral points can be decomposed as the leftward sum of astrons and a finite part, corresponding exactly to the progression of dominant directions that appeared in Example 1.5. In that example,

$$\overline{\mathbf{x}} = \omega\mathbf{v}_1 \dot{+} \cdots \dot{+} \omega\mathbf{v}_k + \mathbf{q}. \tag{1.9}$$

This can be verified by comparing the definitions and observations above with Eq. (1.7). Such a representation of an astral point is not unique. Nevertheless, every astral point does have a unique *canonical representation* in which the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are orthonormal, and $\mathbf{q}$ is orthogonal to all of them (see Theorem 7.26).

Every astral point has an intrinsic *astral rank*, which is the smallest number of astrons needed to represent it, and which is also equal to the number of astrons appearing in its canonical representation. In particular, this means that every astral point in $\overline{\mathbb{R}^n}$ has astral

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

rank at most n. Points in $\mathbb{R}^n$ have astral rank 0. Astral points of the form $\omega\mathbf{v} \dotplus \mathbf{q}$ with $\mathbf{v} \neq \mathbf{0}$ have astral rank 1; these are the points obtained as limits of sequences that go to infinity along a halfline with starting point $\mathbf{q}$ in the direction of $\mathbf{v}$, such as $\mathbf{x}_t = t\mathbf{v} + \mathbf{q}$.

1.4 Minimization Revisited

In each of the examples from Section 1.1, the function being considered was minimized only via a sequence of points going to infinity. The preceding development allows us now to take the limit of each of those sequences in astral space. The next step is to extend the functions themselves to astral space so that they can be evaluated at such an infinite limit point. To do so, we focus especially on a natural extension of a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ to astral space called the *lower semicontinuous extension* (or simply, *extension*) of f to $\overline{\mathbb{R}^n}$, which we introduce in Section 10.1. This function, written $\overline{f} : \overline{\mathbb{R}^n} \to \overline{\mathbb{R}}$, is defined at any $\overline{\mathbf{x}} \in \overline{\mathbb{R}^n}$ as

$$\overline{f}(\overline{\mathbf{x}}) = \inf_{\substack{(\mathbf{x}_t) \text{ in } \mathbb{R}^n: \\ \mathbf{x}_t \to \overline{\mathbf{x}}}} \liminf_{t\to\infty} f(\mathbf{x}_t),$$

where the infimum is taken over all sequences $(\mathbf{x}_t)$ in $\mathbb{R}^n$ that converge to $\overline{\mathbf{x}}$. In words, $\overline{f}(\overline{\mathbf{x}})$ is the infimum across all possible limit values achievable by any sequence $\big(f(\mathbf{x}_t)\big)$ for which $\mathbf{x}_t \to \overline{\mathbf{x}}$. For example, this definition yields the extension of the exponential function given in Example 1.4.

In general, the extension $\overline{f}$ is a lower semicontinuous function on a compact space, and therefore always has a minimizer. Moreover, by $\overline{f}$'s construction, an astral point $\overline{\mathbf{x}}$ minimizes the extension $\overline{f}$ if and only if there exists a sequence in $\mathbb{R}^n$ converging to $\overline{\mathbf{x}}$ that minimizes the original function f. Additionally, if $\overline{f}$ is not just lower semicontinuous, but actually continuous at $\overline{\mathbf{x}}$, and if $\overline{\mathbf{x}}$ minimizes $\overline{f}$, then *every* sequence in $\mathbb{R}^n$ converging to $\overline{\mathbf{x}}$ minimizes f. This makes $\overline{f}$'s continuity properties algorithmically appealing; understanding when and where the extension $\overline{f}$ is continuous is therefore one important focus of this book.

Having extended both $\mathbb{R}^n$ and functions on $\mathbb{R}^n$, we are now ready to return to the minimization problems from Section 1.1 as an illustration of the notions discussed above, as well as a few of the general results appearing later in the book (though somewhat specialized for simpler presentation).

In what follows, $\mathbf{e}_1, \ldots, \mathbf{e}_n$ denote the standard basis vectors in $\mathbb{R}^n$ (so $\mathbf{e}_i$ is all 0's except for a 1 in the ith component).

Example 1.6 (Diagonal valley, continued). We derive the extension $\overline{f}$ of the diagonal valley function from Example 1.2, which turns out to be continuous everywhere. First, we rewrite the function using inner products, since these can be continuously extended to astral space:

$$f(\mathbf{x}) = e^{-x_1} + (x_2 - x_1)^2 = e^{\mathbf{x}\cdot(-\mathbf{e}_1)} + \big[\mathbf{x} \cdot (\mathbf{e}_2 - \mathbf{e}_1)\big]^2.$$

Now to obtain the continuous extension, we can just rely on the continuity of the coupling function to obtain

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

$$\overline{f}(\overline{\mathbf{x}}) = e^{\overline{\mathbf{x}}\cdot(-\mathbf{e}_1)} + \left[\overline{\mathbf{x}}\cdot(\mathbf{e}_2 - \mathbf{e}_1)\right]^2.$$

Here, implicitly, the functions e^x and x^2 have been extended in the natural way to $\overline{\mathbb{R}}$ according to their limits as $x \to \pm\infty$. Note that both terms of the summation are nonnegative (though possibly $+\infty$), so the sum is always defined.

The minimizing sequence $\mathbf{x}_t = [t, t]^\top$ from Example 1.2 follows a halfline, and converges in astral space to the point $\overline{\mathbf{x}} = \omega\mathbf{v}$, where $\mathbf{v} = [1, 1]^\top$. This point is an astron and its astral rank is one. That $\overline{f}$ is both continuous everywhere and has a rank-one minimizer is not a coincidence: In Sections 25.1 and 26.2, we prove that if the extension $\overline{f}$ is continuous everywhere, then it must have a minimizer of astral rank at most one. ■

Example 1.7 (Two-speed exponential, continued). Recall the two-speed exponential function from Example 1.3:

$$f(\mathbf{x}) = e^{-x_1} + e^{-4x_2 + x_1^2}.$$

Unlike the previous example, this function's extension $\overline{f}$ is not continuous everywhere. Earlier, we argued that the sequence $\mathbf{x}_t = [t, t^2/2]^\top$ minimizes f, satisfying $f(\mathbf{x}_t) \to 0$. The sequence $(\mathbf{x}_t)$ converges to the astral point $\overline{\mathbf{x}} = \omega\mathbf{e}_2 + \omega\mathbf{e}_1$, with astral rank 2. On the other hand, the sequence $\mathbf{x}'_t = [t, t^2/5]^\top$ also converges to $\omega\mathbf{e}_2 + \omega\mathbf{e}_1$, but $f(\mathbf{x}'_t) \to +\infty$. This shows that $\overline{f}$ is not continuous at $\overline{\mathbf{x}}$, and means, more specifically, that the extension $\overline{f}$ satisfies $\overline{f}(\overline{\mathbf{x}}) = 0$, but not all sequences converging to $\overline{\mathbf{x}}$ minimize f.

It turns out that $\overline{\mathbf{x}}$ is the only minimizer of $\overline{f}$, so $\overline{f}$ is also an example of a function that does not have a rank-one minimizer, and so cannot be minimized by following a halfline to infinity. As discussed in the previous example, the fact that $\overline{f}$ has no rank-one minimizer implies generally that it cannot be continuous everywhere. ■

Example 1.8 (Log-sum-exp, continued). The log-sum-exp function from Example 1.1 has a continuous extension similar to the diagonal valley function, but a direct construction is slightly more involved. Instead, we take a different route that showcases an exact dual characterization of continuity in terms of a geometric property related to f's conjugate, f^*. In Section 26.2, we will see that when f is finite everywhere, its extension $\overline{f}$ is continuous everywhere if and only if the effective domain of f^* has a conic hull that is *polyhedral*, meaning that it is equal to the intersection of finitely many halfspaces in $\mathbb{R}^n$. (The *effective domain* of a function that takes values in $\overline{\mathbb{R}}$ is the set of points where the function is not $+\infty$. The *conic hull* of a set $S \subseteq \mathbb{R}^n$ is obtained by taking all of the nonnegative combinations of points in S.)

For the log-sum-exp function, we have, for $\mathbf{u} \in \mathbb{R}^n$,

$$f^*(\mathbf{u}) = \begin{cases} \sum_{i=1}^n (u_i + \hat{p}_i) \ln(u_i + \hat{p}_i) & \text{if } \mathbf{u} + \hat{\mathbf{p}} \in \Delta_n, \\ +\infty & \text{otherwise,} \end{cases}$$

where, as before, Δ_n is the probability simplex in $\mathbb{R}^n$. Hence, the effective domain of f^* is $\Delta_n - \hat{\mathbf{p}}$, which, being a translation of a simplex, is polyhedral. Its conic hull is therefore also polyhedral, so $\overline{f}$ is continuous everywhere (by Theorem 26.15ea).

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

Earlier, we considered the particular case that $n = 3$ and $\hat{\mathbf{p}} = [0, \frac{1}{3}, \frac{2}{3}]^\top$. We saw, in this case, that f is minimized by the sequence $\mathbf{x}_t = t\mathbf{v} + \mathbf{q}$, where $\mathbf{v} = [-1, 0, 0]^\top$ and $\mathbf{q} = [0, \ln \frac{1}{3}, \ln \frac{2}{3}]^\top$. In astral terms, the sequence $(\mathbf{x}_t)$ converges to $\overline{\mathbf{x}} = \omega\mathbf{v} + \mathbf{q}$, which does indeed minimize $\overline{f}$ (as does every sequence converging to $\overline{\mathbf{x}}$, since $\overline{f}$ is continuous everywhere).

These facts can further be related to differential properties of f^* (which, as in Example 1.4, is the same as $\overline{f}$'s conjugate, as will be defined in Section 11.1). As discussed in Section 1.2, every subgradient of f^* at $\mathbf{0}$ is a minimizer of f. By its standard definition, $\mathbf{x} \in \mathbb{R}^n$ is a *subgradient* of f^* at $\mathbf{u} \in \mathbb{R}^n$ if

$$f^*(\mathbf{u}') \geq f^*(\mathbf{u}) + \mathbf{x} \cdot (\mathbf{u}' - \mathbf{u}) \tag{1.10}$$

for all $\mathbf{u}' \in \mathbb{R}^n$, so that the epigraph of f^* is supported at $\mathbf{u}$ by the affine function (in $\mathbf{u}'$) on the right-hand side of the inequality. In this case, as in Example 1.4, f^* has *no* subgradients at $\mathbf{0}$, even though $\mathbf{0}$ is in the effective domain of f^*, corresponding to f having no finite minimizers.

This incompleteness goes away when working in astral space. As hinted in the discussion of Example 1.4, we can define generalized subgradients that are themselves astral points, which thus include "infinite subgradients" corresponding to vertical tangents to the function's epigraph, in addition to all standard (finite) subgradients. Although our formulation in Section 27.2 is more general, when $f^*(\mathbf{u}) \in \mathbb{R}$, we can define these simply by replacing the inner product in Eq. (1.10) with the astral coupling function; thus, an astral point $\overline{\mathbf{x}} \in \overline{\mathbb{R}^n}$ is an *astral dual subgradient* of f^* at $\mathbf{u}$ if

$$f^*(\mathbf{u}') \geq f^*(\mathbf{u}) + \overline{\mathbf{x}} \cdot (\mathbf{u}' - \mathbf{u}) \tag{1.11}$$

for all $\mathbf{u}' \in \mathbb{R}^n$. Whereas, as we just noted, it is possible for a convex function to have no standard subgradients at a particular point, in astral space, *every* point must have an astral dual subgradient (Theorem 29.1), even points not in the function's effective domain (using a generalized version of the definition in Eq. 1.11). Moreover, the astral dual subgradients of a convex function at a particular point fully determine the function's instantaneous rate of change in every direction, even if infinite in some directions and finite in others (Theorem 27.16).

In this particular case, from the preceding development, it can be checked that $\overline{\mathbf{x}} = \omega\mathbf{v} + \mathbf{q}$, as defined above, does indeed satisfy the condition in Eq. (1.11) at $\mathbf{u} = \mathbf{0}$, and so is an astral dual subgradient. By general results given in Chapter 28, this implies that $\mathbf{0}$ is a so-called astral *primal* subgradient of $\overline{f}$ at $\overline{\mathbf{x}}$ (by Corollary 28.22), which in turn implies that $\overline{\mathbf{x}}$ minimizes $\overline{f}$ (by Proposition 28.5).

Thus, whereas the conjugate of the original function f has no subgradient at $\mathbf{0}$, corresponding to f having no finite minimizer, in astral space, the conjugate does have an astral dual subgradient at $\mathbf{0}$ which, correspondingly, is also a minimizer of the function's extension $\overline{f}$. ■

These examples illustrate some of the key questions and topics that we study in astral space, including continuity, conjugacy, convexity, the structure of minimizers, differential theory, and optimality.

(continued...)

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

SUBJECT INDEX

References to figures and algorithms are marked, respectively, with f *and* a.

absolute value at infinity, 456, 457f
    astral subgradients of extension, 457, 474, 510
absolute value function, subgradients of, 454f, 453–454, 456, 473
AdaBoost, 583
addition, *see* downward addition; leftward addition; sequential sum; set sum; upward addition
addition matrix, 263
affine combinations, 49
affine functions (astral)
    convexity of, 355
    primal conjugate of, 177
affine functions (standard), 57
    extension of, 160
    subgradients of extension, 483
affine hull, 49
affine map in composition with astral function
    convexity of, 355–356
affine map in composition with standard function, 59
    extension of, 213–215
    relative interior of domain, 61–62
    subgradients of extension, 505–507
affine maps, astral, 119
    applied to polytope, 247–250
    astral cones and, 284
    convexity preserved, 250–251
    sequential convex combinations under, 278–279
affine sets, 49
    astral closure of, 216–217
algorithms, minimization, 577–579, 581–582
all-directions equivalence, 9, 80f, 81–83
    defined, 81
astral affine maps, *see* affine maps, astral
astral basis, 339–341
astral closedness, *see* closedness, astral (primal)
astral cones, *see* cones, astral
astral conic hull, *see* conic hull, astral
astral dual closedness, *see* closedness, astral dual
astral linear maps, *see* linear maps, astral
astral linear subspaces, *see* linear subspaces, astral
astral points
    commutativity of, 145–146
    decomposition of, 103–104
    finite, 85–86
    as functions, *see* functional representation of astral points
    iconic decomposition of, 141–142
    infinite, 10, 86
    representations of, *see* representations of astral points
    sequences involving, *see under* sequence convergence
astral rank, *see* rank, astral
astral rays, *see* rays, astral
astral recession cone, *see* recession cone, astral
astral space
    0- and 1-dimensional, 83
    algebraic structure, 140, 143
    as compactification, 9, 100, 105
    compactness of, 99–100
    construction, 9, 79–83
    cosmic space comparison, 153–155
    defined, 83
    first-countability of, 107–109
    galactic structure, 144–145
    not metrizable, 111
    not second-countable, 111
    versus $(\overline{\mathbb{R}})^n$, 85–86
    sequential compactness of, 109
    topology, *see* astral topology

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

astral span, *see* span, astral
astral topology
  closure in, *see* closure, astral
  construction, 89–91
  defined, 90
  Euclidean space as subspace of, 91–92
  properties, 11–12, 99–100
  as weak topology, 90–91
astrons, 10, 100–101
  sequence of, 151–152
  topological isolation of, 110f, 109–112, 147f
asymptotic dominance, 105–106

ball (open Euclidean), 22
barrier cone (of function), 189–191
  astral closedness and, 206
  astral recession cone and, 398–399
  defined, 189
  exponential composition and, 193
  recession cone and, 203
  recessive completeness and, 399–400
  reduction-closedness and, 202–206
barrier cone (of set), 188
  of epigraph, 188–189
base, 32
basis (for linear subspace), 339
basis (topological), *see* base
basis, astral, 339–341
Bernoulli distributions, 594
β-smoothness, 578–579
biconjugate (astral), 179–180
  of astronic reduction, 196–197
  of extension, 182–183
biconjugate (standard), 64
bijection, 22
boundary, 50
  relative, 50–51

canonical minimizers, 418–419
  defined, 418
  generation of all, 408a, 408, 418–419
  of maximum rank, 423–424
  of rank at most one, 425–427
canonical representation, 12, 126–127
  derived from sequence, 151
Carathéodory's theorem, 47
  outer hull and, 242–243
Cartesian product, 39
Case Decomposition Lemma, 104
Cauchy–Schwarz inequality, 22
Cliff ramp, 93f, 93–94
closed (convex function), 62
closed graph property, 268
closed sets, 33
closedness, astral (primal), 180–181
  barrier cone and, 206
  of extension, 200–202, 206, 368–369
  of general function, 366–368, 373
  of lower-bounded function, 369
  in one dimension, 365–366
  preserved by composition with linear map, 184–185
  preserved by linear image, 369–370
  reduction-closedness and, 200–201
closedness, astral dual, 180–181
  characterization, 376–379
  not preserved by linear image, 370
  in one dimension, 374–376
  of positively homogeneous functions, 380–381
  preserved by composition with linear map, 185
  standard closedness and, 181
closure (of convex function), 62–63
closure (of set in $\mathbb{R}^n$), 50–51
closure (topological), 33
  astral versus Euclidean, 92
closure, astral, 92
  of affine set, 216–217
  of convex set, 258–259
  convexity of, 239
  of halfspaces, 226–227
  of intersection of sets, 213
  of inverse image under linear map, 216
  of linear subspace, 325
  of polar, 329–330
  representational closure and, 322–325
  of standard polyhedral sets, 244–245
  of sublevel sets, 226
closure, representational, *see* representational closure
coercivity, 411–412
  defined, 411
  of universal reduction, 413–414
column space

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

astral, 115
standard, 27
column-orthogonal (matrix), 27
commutativity (of astral points), 145–146
compactification, 9, 37
astral space as, 100
compactness, 36–37
of astral space, 11, 99–100
of extended reals, 36
extremum attained, 36, 42–43
of product space, 40
sequential, *see* sequential compactness
complementary slackness
astral, 555
standard, 550
component projection matrices
for pairs in $\mathbb{R}^n \times \mathbb{R}$, 167–168
for tuples in $\mathbb{R}^{mn}$, 263
composition (with function), *see* affine map in composition with; linear map in composition with; nondecreasing function, composition with
concave functions, 57
cones (standard), 51–53
as intersection of dual halfspaces, 316
under linear map, 52
cones, astral, 280–281, 283–287
convex hull of, 285
as convex hull of icons, 286–287
convexity of, 286
defined, 281
as intersection of halfspaces, 306
linear or affine maps and, 283–284
sequential sum of, 285–286
cones, naive, 280
conic combinations, 52
sequential, *see* sequential conic combinations
conic hull (standard), 52
as intersection of standard halfspaces, 54
as intersection of dual halfspaces, 316
conic hull, astral, 288–290
as convex hull, 289
defined, 288–289
as representational closure, 317–320
of union, 289–290
as union of outer conic hulls, 291–292
conic hull, outer, *see* outer conic hull
conjugate (standard), 6, 64–65
of astronic reduction, 195–196
of composition with linear map, 66–67
of composition with nondecreasing function, 67–70
defined, 64
of indicator function, 66
of linear image, 67
of pointwise maximum, 67
re-expressed, 176
of shadow, 195
of sum of functions, 66
conjugate, dual (astral), 178–179, 181
convexity of, 357
defined, 178
expressed with leftward addition, 182
of image under linear map, 185
conjugate, primal (astral), 176–178, 181–182
of affine function, 177
defined, 177
of extensions, 182
of image under linear map, 184–185
constancy space, 75–76
constrained optimality conditions, astral, 537–542
with affine constraints, 538–539, 541–542, 565
for convex program, 558–563
equivalent definitions, 537–538
general, 538, 540–541
constrained optimality conditions, standard, 537
for convex program, 550
constraint region, extension when restricted to, 222–223
continuity, 35
of astral affine maps, 119
of astral linear maps, 114–115
of astral subgradients, 568–573
extensible, *see* extensible continuity
of extensions, *see* continuity of extensions
of standard convex functions, 64
of standard subgradients, 70–71
continuity of extensions, 164
at all minimizers, 448–449
characterizations, 440–447
extensible continuity and, 164–166
minimizing sequences and, 567–568, 571–572, 579–581
of nondecreasing functions, 166–167
polyhedral recession cones and, 448–449

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

continuity of extensions (*continued*)
recessive completeness and, 447–449
sources of discontinuity, 439–440
subgradient sequences and, 567–568, 570–572
when finite everywhere, 448–449
convergence, *see* sequential convergence
convergence in all directions, 9, 80f, 80–83
defined, 80–81
convergence of entries at decreasing rates, 106–107
convex combinations, 47
sequential, *see* sequential convex combinations
convex cone generated by set, *see* conic hull
convex functions (standard)
characterized, 58
closed, 62
closure of, 62–63
constructing, 58–61
continuity of, 64
defined, 57
domain, 57–58
lower semicontinuous hull of, 62–63
nonemptiness of dual subdifferentials and, 491–492
strictly, 58
sublevel sets of, 57–58, 63–64
convex functions, astral, 349–353
bounded along halfline, 353
characterization, 350–352
constructing, 354–358
defined, 349
domain of, 352
in one dimension, 353–354
sublevel sets of, 352
sum of, 355
convex hull (standard), 47
as intersection of standard halfspaces, 54
as intersection of dual halfspaces, 313–314
convex hull, astral, 245–246
under affine map, 251
of astrons, 320–322
closedness preserved, 259–260
defined, 245
openness preserved, 258
sequential sum and, 269–270
as union of polytopes, 246
as union of sequential convex combinations, 277–278
convex hull, outer, *see* outer convex hull
convex programs, *see* ordinary convex programs
convex sets (standard), 47
astral convexity and, 237–238
as intersection of dual halfspaces, 313–314
operations on, 48
convex sets, astral, 237–239
under affine map, 250–251
closure of, 258–259
defined, 237
interior of, 256–258
as intersection of halfspaces, 303
coordinate descent, 581–582
cosmic space, 153
compared to astral space, 153–155
countability, 37
countability axioms, *see* first-countability; second-countability
countable neighborhood base, 37–39
for astral point, 108f, 107–109
convexity of, 238–239
nested, 37
coupling function, 9–10, 84–85
continuity of, 100
defined, 84
distributivity, 84
icons and, 141
as matrix product, 115
matrix representation and, 121–122
on subgradient sequences, 568–570

decompositions, astral, 103–104, *see also* representations of astral points
with dominant direction, 150
as icon and finite part, 141–142
orthogonal, 104
dense (set), 34
derivative, directional, *see* one-sided directional derivative
diagonal valley, 3f, 3–4, 13–14
differentiability, 71
differential continuity, *see under* continuity
dimension
of affine set, 49
of astral linear subspace, 341
of linear subspace, 26

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

direction, 149
directional derivative, *see* one-sided directional derivative
discrimination, sharp, 54–56
domain, effective, 23
    of astral convex function, 352
dominant direction, 148–150
    decomposition using, 150
    defined, 149
    sequence of, 152–153
downward addition, 25
dual biconjugate, *see* biconjugate, dual (astral)
dual conjugate, *see* conjugate, dual (astral)
dual halfspaces, astral, 312–314
    cones as intersection of, 316
    convex sets as intersection of, 313–314
    defined, 312
    homogeneous, 315
dual orthogonal complement, *see* orthogonal complement, dual
dual polar, *see* polar, dual
dual separation, *see* separation, strong astral dual

easy set, 429, *see also* reducible indices
effective domain, *see* domain, effective
embedding of $\overline{\mathbb{R}^n} \times \mathbb{R}$ in $\overline{\mathbb{R}^{n+1}}$, 167–169
empirical distribution, 595
empirical risk functions, 427–438, 449
    canonical minimizers of, 433
    continuity of, 428–429
    extension of, 428–429
    faces of associated convex hull, 433–436
    general form, 427
    in machine learning and statistics, 427
    minimizers of, 436–438
    minimizing sequences, 438
    persistent and reducible indices of, 429–430
    rank of minimizers of, 433
    recession cone of, 428–429
    recessive completeness of, 428–429
    universal reducers of, 430–433
    universal reduction of, 430–433
enlarged space, 153
entropy (of distribution), 590–591
    maximum, 597–598
entropy function, 464, 465f
    astral dual subgradients of, 465f, 465–466, 469–470
    constrained optimality and, 541–542
    directional derivatives of, 469–470
epigraph
    of astral function, 167, 170–171
    barrier cone of, 188–189
    defined, 23
    of extension, 162, 171
    lower semicontinuity and, 42
    relative interior of, 61
    separation of disjoint set from, 376–377
    support function of, 69–70
equality constraints (of convex program), 547
    in matrix form, 564–565
Euclidean norm, 21
Euclidean space
    first-countability of, 37
    second-countability of, 39
    as subspace of astral space, 91–92
    topology, 32–33
    zero-dimensional, 30–31
Euclidean topology, 32–33
exponential function
    composition with, 191–193
    extension of, 7f, 6–8, 191
exponential loss, 583
exponential-family distributions, 585–587
    astral extension, *see* extended exponential-family distributions
    defined, 585
    estimation from data, 584, 586
extended exponential, 7f, 6–8
extended exponential-family distributions, 588–594
    Bernoulli, 594
    defined, 589
    entropy of, 590–591
    estimating from data, 595–599
    support of, 599–604
extended reals, 2, 21
    arithmetic, 23–24
    Cartesian product of, 85–86
    continuity of operations, 26
    convergence in, 25–26
extensible continuity, 164–166
    defined, 164

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

extension (of function), *see* lower semicontinuous extension

faces, 56
    smallest face including set, 435
feasibility (of convex program)
    astral (primal), 552–555
    dual, 548, 550
    standard (primal), 548, 550
feature map, 585
    minimal, 606
    translated, 588
Fenchel duality, 531–536
    astral version, 532–533
    minima characterized by subdifferentials, 533–535
    standard theorem, 531–532
Fenchel–Young inequality, 471
    astral subgradients and, 472–473, 475–476
    linear tilts and, 476
finite everywhere (function), 57
finite part, 141
finitely generated (set), 56–57
    representational closure of, 323–325
first-countability, 37–39
    of astral space, 11, 107–109
    defined, 37
    of Euclidean space, 37
first-order optimality (of convex program)
    astral, 555
    standard, 550–551
flattening valley, 389f, 389
    continuity of, 439, 445
    not minimized with vanishing gradients, 567–568
    recession cone of, 389–390
    recessive completeness and, 400
four-outcome exponential family, 586
    conjugate of log-partition function, 592f
    marginal polytope, 603f, 603–604
    maximum-likelihood distribution, 596–597
    minimality, 606
    negative log-likelihood of, 587f
    no finite likelihood maximizer, 586–587
full column or row rank, 27
functional representation of astral points, 89f, 88–89
    characterizations, 93–99
    defined, 88
    dual conjugate of, 180

galaxies, 143–145
    defined, 143
    homeomorphic to Euclidean space, 147f, 146–148
    marginal polytope faces and, 599–604
    tree structure of, 145
games, matrix, *see also* minimax theorems
    astral, 543
    standard, 542
gliding parabola, 535, 536f
    constrained optimality and, 538–539
    Fenchel duality and, 535–536
gradient descent, 577–579
gradients (standard), 71, *see also* subdifferentials
    sequences of, *see* subgradient sequences

halflines (standard), 47, 292
halflines, astral, 292–295
    defined, 292
    function bounded along, 353
    included in convex set, 294–295
    sequential characterization, 294
halfspaces (standard), 53
    astral closure of, 226–227
halfspaces, astral, 231–232
    astral cones as intersection of, 306
    convex sets as intersection of, 303
    convexity of, 238–239
    homogeneous, 280–281
hard core, 429, *see also* persistent indices
Hausdorff (space), 34
hemispace, 314
homeomorphism, 35–36
homogeneous (halfspace or hyperplane), 53
    astral, 280–281
    dual halfspace, 315
horizon subgradients, 453
hull, *see* astral conic hull; convex hull; outer conic hull; outer convex hull
hull operator, 48–49
hyperplanes (standard), 53
hyperplanes, astral, 231–232
    convexity of, 238–239
hypograph, 23

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

iconic part, 141–142
icons, 140–143
  decomposition using, 141–142
  defined, 140
  $\boldsymbol{\omega}$, 120
  prefix of, 144
idempotent, 140–141
image (of a set), 22
image of function under linear map, *see* linear image of astral function; linear image of standard function
implicit regularization, 581
increasing function, *see* nondecreasing function; strictly increasing function
indicator functions (astral), 163–164
  conjugate and biconjugate of, 234–235
  conjugate for singleton, 177–178
  convexity of, 354–355
  polar operations and, 328–329
  subgradients of, 500
indicator functions (standard), 58
  conjugate of, 66
  dual subgradients of, 500–501
  extension of, 164
  polar operations and, 328–329
  standard subgradients of, 497
  subgradients of extension, 500
inequality constraints (of convex program), 547
infimal convolution, 219–220
infinite points, 10, 86
injective (function), 22
inner approach, 231
inner product, 21
interior, 33
  of convex set, 256–258
  relative, 50–51
interval convexity, 231
inverse image (of a set), 22

Karush–Kuhn–Tucker, *see* KKT
KKT conditions
  astral, 554–555
  standard, 550–551
KKT vectors, 549
  existence of, 556–558

Lagrangian, 548–550
  astral saddle points of, 555
  defined, 548
  extension of, 555–556
  standard saddle points of, 549–550
  subgradients of extension, 555–556
leftward addition
  of astral points, 12, 102–103
  commutativity of, 145–146
  of scalars, 12, 24–25
  of sets, 103
lexicographic order, 315
lexicographic separation, 314–315
likelihood, 586–587, *see also* maximum likelihood
limit inferior and superior, 26
limits of astrons, 151–152
line segments, 47
linear combinations, 26
linear dependence, 26
linear extended functions, 99
linear functions, *see* affine functions
linear functions, continuous extension of, 9
  all-directions equivalence and, 82
  in astral space, 90–91, 100, 105
  in cosmic space, 154
linear image of astral function, 183–184, 217–218
  astral closedness preserved, 369–370
  conjugate identities, 184–185
  convexity of, 356–357
linear image of standard function, 59–60
  astral conjugate identities, 185
  astral dual closedness not preserved, 370
  extension of, 218–219
  relative interior of domain, 62
  standard conjugate of, 67
  standard subgradients of, 71–72, 515
  subgradients of extension, 514–515
linear independence, 26
linear map in composition with astral function, 183–184
  astral closedness preserved, 184–185
  conjugate identities and inequalities, 184–185
  convexity of, 355–356
linear map in composition with standard function, 59
  astral conjugate identities and inequalities, 185–186
  astral dual closedness preserved, 185
  extension of, 213–216

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

linear map in composition with standard function (*continued*)
standard conjugate of, 66–67
standard subgradients of, 71
subgradients of extension, 505–506
linear map, image of function under, *see* linear image of astral function; linear image of standard function
linear maps (standard), 26
astral closure of inverse image under, 216
linear maps, astral, 113–116, *see also* affine maps
astral cone under, 283–284
characterization, 117–119
continuity of, 114–115
coupling function as, 115
defined, 114
sequential sum under, 270–271
linear subspaces (standard), 26
closure same as representational closure, 325
correspondence with astral linear subspaces, 335, 345
dimension of, 26
linear subspaces, astral, 331–336
characterizations, 331–335
correspondence with linear subspaces, 335, 345
defined, 331
dimension of, 341
linear tilt of astral function, 476–478
defined, 476
subgradients and, 478–479
linear tilt of standard function, 479–482
defined, 479
notational convention, 480
subgradients and, 482–484
log-partition function, 585–586
conjugate of, 590, 591, 593
extension of, 588
gradient of, 590
subgradients of extension, 593–594
for translated feature map, 588
log-sum-exp, 2–3, 14–15
logarithm, extension of, 583
logistic regression, 427, 582–583
lower addition, 25
lower envelope, 358–364
defined, 358
of outer hull of epigraph, 370–373
lower semicontinuity, 41–44
of convex functions, 62
defined, 41–42
minimum attained, 42–43
lower semicontinuous extension, 13, 159–163
of affine function, 160
astral closedness of, 200–202, 206, 368–369
astral recession cone of, 384–386
biconjugate of, 182–183
of composition with affine map, 213–216
of composition with nondecreasing function, 223–226
conjugate of, 182
continuity of, *see* continuity of extensions
convexity of, 349–350
defined, 159–160
epigraph of, 162, 171
of indicator function, 164
of linear image of function, 218–219
minimizers of, *see* minimizers of extensions
minimum attained, 163
of nondecreasing functions, 166–167
of pointwise supremum, 220–222
of restricted function, 222–223
of scalar multiple, 207
simultaneous convergence to multiple, 211–212
subdifferentials of, *see* subdifferentials of extensions
of sum of functions, 207–213
trivial extension and, 161
lower semicontinuous hull, 44–46
of convex function, 62–63
defined, 44
epigraph of, 170–171
lower support function, 360–364
defined, 360
of outer hull of epigraph, 370–373

m-ary sequential sum, 262–265
defined, 262
magnitude function, 579
majorization, 23
marginal polytope, 591–593
faces of, 599–606
when feature map minimal, 606–607

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

matrices, *see also* component projection matrices; positive upper triangular matrices
notation, 21
projection (orthogonal), 28–29
upper triangular, 27
with zero rows or columns, 31
matrix product, astral, 113–117
commutativity and, 146
coupling function of, 114
defined, 114
row-by-row calculation of, 116–117
sequences converging to, 135–136
matrix representation, *see* representations of astral points
maxima, pointwise, *see* suprema and maxima, pointwise
maximin strategy, 542
maximum entropy, 597–598
maximum-likelihood distribution, 586, 595–597
equivalent formulations, 595–596
maximum entropy and, 597–598
methods for computing, 584
support of, 604
mean map, 590–594
defined, 590
image of galaxies under, 599–604
yielding homeomorphism, 604–607
metric topology, 33
metrizability, 33
astral space not metrizable, 111
second-countability and, 39
midsegments and midpoints (sequential), 275, 350–351
sequential characterization, 276–277
minimal (feature map), 606
minimax strategy, 542
minimax theorems
astral, 543–546
standard, 542–543
minimization methods, 577–579, 581–582
minimizers of extensions
canonical, *see* canonical minimizers
characterizations of, 387–388, 395
continuity at all, 448–449
finite parts characterized, 406
functions with any given, 424–425
generation of all, 393a, 393–397
of maximum rank, 422–424
of rank at most one, 425–427
sequences converging to, *see* minimizing sequences
testing for, 392a, 392–393
minimizing sequences
examples, 2–4, 13–15
existence of, 576–577
with progress guarantee, 579–581
using subgradient descent, 577–579
with vanishing subgradients, 570–572
monotone extension, 60
monotonicity of subgradients, 495–496
Moore–Penrose generalized inverse, 29–30
multiples, *see* ω-multiples; scalar multiples (astral); sequential multiples
multipliers (of convex program), 548

naive cones, 280
natural embedding of $\overline{\mathbb{R}^n} \times \mathbb{R}$, 169
natural numbers, 21
negative log-likelihood, 584, 586–587
neighborhood, 33
neighborhood base, 37, *see also* countable neighborhood base
nondecreasing function, composition with
astral convexity of, 357–358
astral subgradients of, 520–527
extension of, 223–226
standard conjugate of, 67–70
standard convexity of, 60–61
standard subgradients of, 72, 522–523
nondecreasing functions, extension of, 166–167
nondeterminism, 394
norm (Euclidean), 21
normal (space), 34
normal cone (primal), 497–499
defined, 498
normal cone (standard), 497
normal cone, dual, 498–500
defined, 498

objective function (of convex program), 547
ω-multiples, 100–101
sequence of, 152
sequences converging to, 138–139
one-sided directional derivatives, 72–73
of an extension, 511–514
astral dual subgradients and, 467–470

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

one-sided directional derivatives (*continued*)
defined for astral functions, 511
one-to-one (function), 22
onto (map), 22
open cover, 36
open sets, 32
optimality conditions
astral, 466, 474–475
constrained, *see* constrained optimality conditions
standard, 5–6, 70
ordinary convex programs, 547–554
astral optimality conditions, 558–563
astral solution definitions, 551–552, 554
defined, 547
existence of astral solution, 560–562
existence of KKT vector, 556–558
extension of, 553–554
no finite solution, 551, 565–566
only equality constraints, 565
standard optimality conditions, 550
subgradients of extension, 553–554
orthogonal complement (primal), 343–346
redefined for astral sets, 343
for standard sets, 28
orthogonal complement, dual, 343–345
defined, 343
orthogonal projection, 28–29
orthogonal projection matrices, 28–29
orthogonality, 26–27, 342–343
orthonormality, 27
outcomes, 585
outer conic hull, 281–282, 287–288
astral conic hull as union of, 291–292
decomposed as rays, 290
defined, 281
as outer convex hull, 281–282
sequential characterization, 291
for sets in $\mathbb{R}^n$, 288
as smallest closed convex astral cone, 305–306
outer convex hull, 232–235
under affine map, 247–250, 305
as biconjugate of indicator, 234–235
convexity of, 238–239
defined, 232–233
of finite set, *see* polytopes, astral
for sets in $\mathbb{R}^n$, 239
as smallest closed convex superset, 303–304
as union of sequential convex combinations, 277–278
outer hull, *see* outer convex hull

persistent indices, 429–430
faces of associated convex hull, 433–436
marginal polytope faces and, 599–604
universal reducers and reduction and, 430–433
polar (primal), 325–329
closure of, 329–330
conjugate of indicator and, 328–329
defined for standard cones, 52–53
redefined for naive cones, 325
polar, dual, 325–330
of closure, 329–330
conjugate of indicator and, 328–329
defined, 325
polyhedral sets (standard), 56–57
astral closure of, 244–245
representational closure of, 323–325
polyhedral sets, astral, 243–245
continuity and, 448–449
defined, 243
recessive completeness and, 387, 399–401
polynomially graded sequence, 10, 86–87
polytopes, astral, 240–243
under affine map, 247–250
convex hull as union of, 246
included in convex set, 243
sequential characterization, 240–242
as union of sequential convex combinations, 277–278
positive upper triangular matrices, 27–28
matrix representation and, 123–124
positively homogeneous functions, 57
astral dual closedness of, 380–381
as support function, 380–381
positives and negatives, 207–208, 213
potential output (of algorithm), 395
prefix (of icon), 144
primal biconjugate, *see* biconjugate, primal (astral)
primal conjugate, *see* conjugate, primal (astral)
product of hyperbolas, 160, 160f
canonical minimizers and, 418
continuity of, 439, 445

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

extension of, 160–161
minimized along halfline, 422
reduction of, 172
shadow of, 174
variant continuous at all minimizers, 449
product space, 40
compactness of, 40
product topology, 40–41
projection (orthogonal), 28–29
Projection Lemma, 122–123
projection map, 40
projection matrices
component, *see* component projection matrices
orthogonal, 28–29
projection orthogonal to vector, 150–151
proper functions, 57
proper separation, 53–54
pseudoinverse, 29–30
Push Lemma, 123–124

$\mathbb{Q}$-compactification, 105
QR factorization, 28
quotient topology, 154

rank (of matrix), 27
astral rank and, 125
rank, astral, 12–13, 125
minimizer of maximum possible, 422–424
minimizers with rank at most one, 425–427
rays (standard), 47, 280
rays, astral, 280–283
defined, 281
as segment, 281–282
sequential characterization, 291
sequential sum of, 290
as union of sequential multiples, 282–283
recession cone (standard), 73–75
astral recession cone and, 386–387, 413
barrier cone and, 203
continuity when polyhedral, 448–449
defined, 73
recessive completeness when polyhedral, 399–401
recession cone, astral, 383–390
as astral cone, 383–384
characterizations, 384–385, 391–392
closedness of, 386
as conic hull of universal reducers, 420
continuity when polyhedral, 448–449
convexity of, 383–384
defined, 383
dual characterization, 397–399
examples, 388–390
extension of, 384–386
minimizers characterized using, 387–388
orthogonal complement of, 413–415
recessive completeness when polyhedral, 399–400
recursive formulation of, 390–391
standard recession cone and, 386–387, 413
recessive completeness, 387
barrier cone and, 399–400
characterizations, 446–447
continuity and, 447–449
ensuring rank-one minimizer, 425–427
examples, 388–390
polyhedral recession cone and, 387, 399–401
reciprocal function, extension of, 589
reducers, universal, *see* universal reducers
reducible indices, 429–430
reduction, universal, *see* universal reduction
reduction-closedness, 199–201
astral closedness and, 200–201
barrier cone and, 202–206
defined, 199
reductions, astronic, 171–175, 194–195
astral biconjugate of, 196–197
astral recession cone of, 390–391
conjugate of, 195–196
defined, 171
extension of, 172–173
as lower semicontinuous hull, 175
reductions, iconic, 197–200
astral analog, 401–402
conjugate of, 402–403
defined, 197
as lower semicontinuous hull, 402
relative interior of domain, 403
subgradients of extension, 507–511
regular (space), 34
relative boundary, 50–51
relative interior, 50–51
relatively open (set), 50

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

representational closure, 317
astral conic hull as, 317–320
convex hull of astrons and, 320–322
of linear subspace, 325
topological closure, when equal to, 322–325
representational span, 337–338
astral span and, 337–338
general definition, 337
for sets in $\mathbb{R}^n$, 338
of singleton, 134–135
representations of astral points
with astrons, 12, 103–104
canonical, *see* canonical representation
coupling function and, 121–122
equivalence of, 127–128
with full-rank matrix, 124–125
as icon and finite part, 141–142
of icons, 141
manipulating, 121–125
matrix form, 120
matrix product of, 121
positive upper triangular matrices and, 123–124
scalar multiple of, 121
sequences converging to, 105, 129–133
restricted linear function, 190–191
astral biconjugate of, 201
astral recession cone and, 398
barrier cone of, 190–191
extension of supremum and, 222
recession cone of, 388–389
recessive completeness and, 400

saddle points (of Lagrangian)
astral, 555
standard, 549–550
scalar multiples (astral), 84–85, 100–102
extension of, 207
by ω, *see* ω-multiples
of sets, 103
subgradients of extension, 502–503
second-countability, 39
astral space not second-countable, 111
of Euclidean space, 39
second-order cone, 208
segments, astral, 233–234
convexity of, 238–239
defined, 233
examples, 235–237, 322
with finite endpoint, 251–256
sequential characterization, 240f, 242
split at a point, 254–256
separation by hyperplane (standard), 53–55, 205
separation by open sets, 34
separation of astral sets, strong, 296–305
from astral cone, 298–300, 305–306
characterized, 297–298
defined, 296–297
main theorem, 298–303
sets in $\mathbb{R}^n$ and, 304–305
separation, lexicographic, 314–315
separation, strong astral dual, 307–311
characterized, 308–310
from convex cone, 315–316
defined, 307
main theorem, 310–311
sequence convergence, 37
of astral points, 100
of astrons, 151–152
to decomposition, 105, 107
of dominant directions, 152–153
to matrix product, 135–136
to matrix representation, 129–133
to minimizers, *see* minimizing sequences
to $\omega\overline{\mathbf{x}} \dotplus \overline{\mathbf{y}}$, 136–139
of ω-multiples, 152
to sequential multiple, 273–274
of subgradients, *see* subgradient sequences
sequences, 22
sequential compactness, 39
of astral space, 11, 109
sequential conic combinations, 275–277
convexity of, 277
defined, 275
sequential characterization, 275–277
sequential convex combinations, 275–279
under affine maps, 278–279
convex hull as union of, 277–278
convexity of, 277
defined, 275
sequential characterization, 275–277
sequential multiples, 271–274
astral rays as union of, 282–283
defined, 271
sequences converging to, 273–274

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

sequential sum, 261–271
associativity of, 264–265
of astral cones, 285–286
of astral linear subspaces, 336
characterizations of, 266–269
closed graph property, 268
closedness preserved, 265–266
convex hull and, 269–270
convexity preserved, 265–266
defined, 261
under linear map, 270–271
m-ary, 262–265
set difference, 22
set sum (standard), 22
shadow (of function), 174
astronic reductions and, 175
conjugate of, 195
iconic reductions and, 402
sharp discrimination, 54–56
sideways cone, 208, 208f
extension of composition with linear map and, 214
extension of sum and, 209
representational and topological closures different, 322
Slater's condition and, 563–564
subgradients of composition with linear map and, 506
simultaneous convergence to multiple extensions, 211–212
Slater's condition, 548
necessity for theory, 563–564
solution (of convex program)
astral, 551–552, 554
standard, 548
span (standard), 26
span, astral, 336–339, 341–342
as conic or convex hull, 336–337
defined, 336
representational span and, 337–338
sequential characterization, 341–342
for sets in $\mathbb{R}^n$, 338
of union, 339
span-boundness, 135
standard basis, 26
standard basis vectors, 21
steepest descent, 581–582
Stone–Čech compactification, 105
strategies, 542
strictly convex functions, 58
astral dual subdifferentials and, 494–495
strictly increasing function, composition with, 191–193
strong equivalence, 133–134
strong separation (standard), 53–54
subbase, 32
subdifferentials of extensions
of affine function, 483
characterizations, 482–483, 511
of composition with affine map, 505–507
of composition with nondecreasing function, 520–527
directional derivatives and, 511–514
dual subgradients of conjugate and, 487
examples, 454f, 456f, 456–457, 483
at finite points, 461–462
of indicator function, 500
of linear image of function, 514–515
linear tilts and, 482–483
nonemptiness of, 512–514
in one dimension, 488–489
of pointwise supremum, 516–520
of reduction, 507–511
of scalar multiple, 502–503
sequences in, *see under* subgradient sequences
of sum of functions, 503–505
subdifferentials, astral (primal), 454–461
0 included in, 474–475
astral closedness and, 460–461
characterizations, 458–460, 478–479
closedness of, 497
convexity of, 496–497
defined, 455–456
of dual conjugate, 475–476
dual subgradients of conjugate and, 472–473, 484–487
examples, 454f, 456f, 456–457
Fenchel–Young and, 472–473
of indicator function, 500
linear tilts and, 478–479
lower semicontinuity and, 460–461
monotonicity of, 495–496
subdifferentials, astral dual, 462–470
0 included in, 466
closedness of, 496

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

subdifferentials, astral dual (*continued*)
of conjugate, 472–473, 481–487
convexity of, 496
defined, 463–464
directional derivatives and, 467–470
disjoint, 494–495
examples, 463f, 465f, 463–466
Fenchel–Young and, 472–473, 475–476
generalizing standard subgradients, 466–467
of indicator function, 500–501
monotonicity of, 495
nonemptiness of, 490–492
in one dimension, 487–488
overlapping, 492–494
primal subgradients of conjugate and, 475–476
sequences in, *see under* subgradient sequences
singleton, 468–469
of support function, 499
subdifferentials, standard, 70–71
of composition with linear map, 71
of composition with nondecreasing function, 72, 522–523
continuity, 70–71
defined, 70
generalized by astral subgradients, 461–462, 466–467
gradient and, 71
of indicator function, 497
of linear image of function, 71–72, 515
of pointwise supremum, 72, 515–516, 518
of sum of functions, 71
subgradient descent, 577–579
subgradient sequences (standard)
continuity and, 567–568, 570–572
convergence of function values, 570–572
converging to **0**, 567–568, 571–572, 576–577
existence of convergent, 575–577
gradients and, 568
subgradient sequences, astral (primal), 568–570
convergence of function values, 569
convergence to subgradient, 569–570
coupling function and, 568–570
subgradient sequences, astral dual
convergence to subgradient, 572–573
existence of convergent, 574–575
subgradients, *see* subdifferentials
sublevel sets
all bounded, 411–412
astral closure of, 226
of astral convex function, 352
lower semicontinuity and, 42
of standard convex function, 63–64
subsequences, 37
subsequential limit, 37
subspace (topological), 34–35
subspaces, linear, *see* linear subspaces
sum of functions (standard)
conjugate of, 66
convexity of, 59
extension of, 207–213
subgradients of, 71
subgradients of extension, 503–505
summability, 24
support (of distribution), 599
support functions (standard), 66
of epigraph, 69–70
support functions, astral, 234
dual conjugate of, 234–235
dual subgradients of, 499
positively homogeneous functions as, 380–381
suprema and maxima, pointwise
astral convexity of, 357
conjugate of, 67
extension of, 220–222
standard convexity of, 60
standard subgradients of, 72, 515–516, 518
subgradients of extension, 516–520
surjective (function), 22

topological spaces, 32
topology
astral, *see* astral topology
defined, 32
Euclidean, 32–33
on extended reals, 33
generated by base or subbase, 32–33
metric, 33
metrizable, 33
of pointwise convergence, 41
product, 40–41
quotient, 154

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

on reals, 32–33
subspace, 34
weak, 41
triangle inequality, 22
trivial extension, 161
conjugate of, 182
tuples, 22–23, 40
two exps and a square, 394
algorithm applied to, 394, 408–409
associated convex hull, 433, 434f
nonminimizer counterexample, 418
persistent and reducible indices of, 430
rank-one minimizer, 426–427
universal reduction of, 433
universal reduction's minimizers, 411
two-speed exponential, 4, 5f, 14
Tychonoff's theorem, 40

universal reducers, 417–421
astral recession cone generated by, 420
convex hull of, 420–421
defined, 417
generation of all, 408a, 408, 418–419
of maximum rank, 423–424
of rank at most one, 425–427
recursive characterization, 419
universal reduction, 404–410
coercivity of, 413–414
construction of, 408a, 408–409
defined, 404
equivalent definitions, 405–406
as iconic reduction, 409–410
minimizers bounded, 411, 413–415
minimizers of extension and, 406
on minimizing sequence, 415
preserved under reduction, 406–407
recessive completeness of, 410
sublevel sets bounded, 413–415
when same as original function, 407, 415–416, 419
updirected (family of sets), 237–238
upper addition, *see* upward addition
upper semicontinuity, 41
upper triangular matrices, 27, *see also* positive upper triangular matrices
upward addition, 25
of convex functions, 355

value (of convex program), 548
value (of game), 542
vertical barrier cone, 189–190

weak topology, 41
astral space and, 90–91

$\langle \overline{\mathbf{x}}, y \rangle$ pairs
embedded in $\overline{\mathbb{R}^{n+1}}$, 167–170
operations on, 170

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

AUTHOR INDEX

Albert, A., 587
Aliprantis, C. D., 32–35, 37, 39–41
Anderson, J. A., 587

Barndorff-Nielsen, O. E., 587
Bartlett, P., 581
Bertsekas, D. P., 63
Border, K. C., 32–35, 37, 39–41
Borwein, J. M., 52, 63
Boyd, S., 1, 577, 581
Brown, L. D., 587

Collins, M., 2, 583
Cover, T. M., 591
Cox, D. A., 4
Csiszár, I., 587

Della Pietra, S., 597
Della Pietra, V., 597
Dudík, M., 577
Dupin, J.-C., 4, 153

Fenchel, W., 16
Freund, Y., 581, 583

Geyer, C. J., 587
Gunasekar, S., 581

Haberman, S. J., 587
Hamel, A. H., 105
Hansen, G. L., 4, 153
Hiriart-Urruty, J.-B., 67, 71, 72
Hoffer, E, 581
Horn, R. A., 27, 28

Jaynes, E. T., 597
Ji, Z., 577, 581
Johnson, C. R., 27, 28
Jordan, M. I., 2, 585

Karush, W., 547
Klee, V. L., 55
Kuhn, H. W., 547
Kullback, S., 597

Lafferty, J., 597
Lee, W. S., 581
Lemaréchal, C., 67, 71, 72
Lewis, A. S., 52, 63
Little, J., 4
Loeb, P. A., 105

Manetti, M., 39
Martínez-Legaz, J.-E., 306, 314, 315
Matúš, F., 587
Moreau, J. J., 177
Munkres, J. R., 32–37, 39, 40, 105

Nacson, M. S., 581
Nemirovski, A. S., 1
Nesterov, Y., 1, 583
Nocedal, J., 577

O'Shea, D., 4

Penrose, R., 29, 30

Rockafellar, R. T., 4, 16, 23, 47–67, 70, 71, 73–75, 153, 294, 305, 412, 453, 495, 531, 535, 537, 548–550, 557
Roman, S., 27, 49

Schapire, R. E., 2, 577, 581, 583
Schrage, C., 105
Silvapulle, M. J., 587
Singer, I., 177–179, 231, 306, 314, 315, 453

For general queries, contact info@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

Singer, Y., 2, 583
Soudry, D., 581
Srebro, N., 581

Takane, Y., 28–30
Takeuchi, K., 28–30
Telgarsky, M., 429, 577, 581
Thomas, J. A., 591
Tucker, A. W., 547

Vandenberghe, L., 1, 577, 581
Vel, M. L. J. van de, 231, 237
von Neumann, J., 542

Waggoner, B., 99, 464, 490, 492, 494
Wainwright, M. J., 2, 585
Wedderburn, R. W. M., 587
Wets, R. J-B, 4, 16, 52, 153, 453
Wright, S. J., 577

Yanai, H., 28–30
Yu, B., 2, 581
Yudin, D. B., 1

Zhang, T., 2, 581

For general queries, contact info@press.princeton.edu