Chapter 1

Introduction to the Central Concepts

"Discrete Convex Analysis" aims at establishing a new theoretical framework of discrete optimization through mathematical studies of "convex functions with combinatorial structures" or "discrete functions with convexity structures." This chapter is a succinct introduction to the central issues discussed in this book, including the role of convexity in optimization, several classes of well-behaved discrete functions, and duality theorems. We start with an account of the aim and the history of discrete convex analysis.

1.1 Aim and History of Discrete Convex Analysis

The motive for "Discrete Convex Analysis" is explained in general terms of optimization. Also included in this section is a brief chronological account of discrete convex functions in relation to the theory of matroids and submodular functions.

1.1.1 Aim

An *optimization problem*, or a *mathematical programming problem*, may be expressed generically as:

$$\text{Minimize } \ f(x) \ \text{ subject to } \ x \in S.$$

This means that we are to find an x that minimizes the value of $f(x)$ subject to the *constraint* that x should belong to the set S. Both f and S are given as the problem data, whereas x is a variable to be determined. The function f is called the *objective function* and the set S the *feasible set*.

In *continuous optimization*, variable x typically denotes a finite-dimensional real vector, say $x \in \mathbf{R}^n$, and accordingly we have $S \subseteq \mathbf{R}^n$ and $f : \mathbf{R}^n \to \mathbf{R}$ (or $f : S \to \mathbf{R}$).[1] An optimization problem with S being a convex set and f a convex function is referred to as a *convex program*, where a set S is *convex* if the line

[1] Notation $\mathbf{R}$ means the set of all real numbers, and $\mathbf{R}^n$ the set of n-dimensional real vectors.

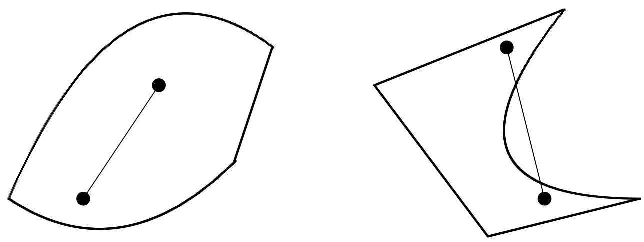

Figure 1.1. *Convex set and nonconvex set*

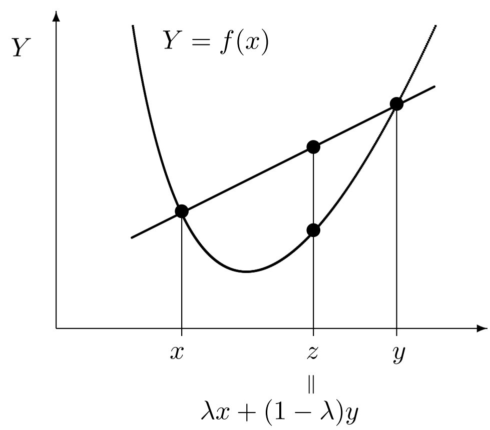

Figure 1.2. *Convex function*

segment joining any two points in S is contained in S (see Fig. 1.1), and a function $f : S \to \mathbf{R}$ defined on a convex set S is *convex* if

$$\lambda f(x) + (1-\lambda) f(y) \geq f(\lambda x + (1-\lambda) y) \tag{1.1}$$

whenever $x, y \in S$ and $0 \leq \lambda \leq 1$ (see Fig. 1.2). Convex programs constitute a class of optimization problems that are tractable both theoretically and practically, with a firm theoretical basis provided by "convex analysis." The tractability of convex programs is largely based on the following properties of convex functions:

1. *Local optimality* (or minimality) guarantees *global optimality*. This implies, in particular, that a global optimum can be found by descent algorithms;

2. *Duality* such as *min-max relation* and *separation theorem* holds good. This leads, for instance, to primal-dual algorithms using dual variables and also to sensitivity analysis in terms of dual variables.

Some more details on these issues will be discussed in §1.2.

In *discrete optimization* (or *combinatorial optimization*), on the other hand, variable x takes discrete values; most typically, x is an integer vector or a $\{0,1\}$-vector. Whereas almost all discrete optimization problems arising from practical applications are difficult to solve efficiently, network flow problems are recognized as tractable discrete optimization problems. In the minimum cost flow problem with linear arc costs, for instance, we have the following fundamental facts that render the problem tractable:

1. A flow is optimal if and only if it cannot be improved by augmentation along a cycle. This statement means that the global optimality of a solution can be characterized by the "local optimality" with respect to augmentation along a cycle;

2. A flow is optimal if and only if there exists a potential on the vertex set such that the reduced arc cost with respect to the potential is nonnegative on every arc. This is a duality statement characterizing the optimality of a flow in terms of the dual variable (potential). This provides the basis for primal-dual algorithms.

In more abstract terms, it is an accepted understanding that the tractability of the network flow problems stems from the "matroidal structure" (or "submodularity") inherent therein. Whereas the meaning of this statement will be substantiated later, it is mentioned at this point that a matroid is an abstract combinatorial object defined as a pair of a finite set, say V, and a family $\mathcal{B}$ of subsets of V that satisfies certain abstract axioms. We refer to V as the ground set, a member of $\mathcal{B}$ as a base, and a subset of a base as an independent set. Matroid is considered to be fundamental in combinatorial optimization, which is evidenced by the following facts:[2)]

1. A base is optimal with respect to a given weight vector if and only if it cannot be improved by an elementary exchange, which means a modification of a base B to another base $(B \setminus \{u\}) \cup \{v\}$ with u in B and v not in B. Thus the "local optimality" with respect to elementary exchanges guarantees the global optimality. Moreover, an optimal base can be found by the so-called *greedy algorithm*, which may be compared to the steepest descent algorithm in nonlinear optimization;

2. Given a pair of matroids on a common ground set, the *intersection problem* is to find a common independent set of maximum cardinality. *Edmonds' intersection theorem* is a min-max duality theorem that characterizes the maximum cardinality as the minimum of a submodular function defined by the rank functions of the matroids.

With the above facts it is natural to think of matroidal structure as a discrete or combinatorial analogue of convexity. The connection of matroidal structure to

[2)] More specific account of these facts will be given in §1.3.

convexity was formulated in the early 1980s as a relationship between submodular functions and convex functions. It was shown by Frank that Edmonds' intersection theorem can be rewritten as a separation theorem for a pair of submodular/supermodular functions, with an integrality (discreteness) assertion for the separating hyperplane in the case of integer-valued functions. Another reformulation of Edmonds' intersection theorem is Fujishige's Fenchel-type min-max duality theorem for a pair of submodular/supermodular functions, again with an integrality assertion in the case of integer-valued functions. A precise statement, beyond analogy, about the relationship between submodular functions and convex functions was made by Lovász: A set function is submodular if and only if the so-called Lovász extension of that set function is convex. These results led to the recognition that the the essence of the duality for submodular/supermodular functions consists in the discreteness (integrality) assertion in addition to the duality for convex/concave functions. Namely,

$$\text{Duality for submodular functions} \;=\; \text{Convexity} \;+\; \text{Discreteness.}$$

Such developments notwithstanding, our understanding of convexity in discrete optimization seems to be only partial. In convex programming, a convex objective function is minimized over a convex feasible region, which may be described by a system of inequalities in (other) convex functions. In matroid optimization explained above, the objective function is restricted to be linear and the feasible region is described by a system of inequalities using submodular functions. This means that the convexity argument for submodular functions apply to the convexity of feasible regions and not to the convexity of objective functions. In the literature, however, we can find a number of nice structural results on discrete optimization of nonlinear objective functions. For example, the minimum-cost flow problem with a separable convex cost function admits optimality criteria similar to those for linear arc costs (Minoux [131] and others), and this can be carried over to the submodular flow problem with a separable convex cost function (Fujishige [65]). Minimization of a separable convex function over a base polyhedron also admits a local optimality criterion with respect to elementary exchanges (Fujishige [60], Girlich–Kowaljow [78], Groenevelt [81]). This fact is used in the literature of resource allocation problems (Ibaraki–Katoh [93], Hochbaum [90], Hochbaum–Hong [91], Girlich–Kovalev–Zaporozhets [77]). The convexity argument concerning submodular functions, however, does not help us understand these results in relation to convex analysis. We are thus waiting for a more general theoretical framework for discrete optimization that can be compared to convex analysis for continuous optimization.

"Discrete Convex Analysis" is aimed at establishing a general theoretical framework for solvable discrete optimization problems by means of a combination of the ideas in continuous optimization and combinatorial optimization. The theoretical framework of convex analysis is adapted to discrete settings and the mathematical results in matroid/submodular function theory are generalized. Viewed from the continuous side, the theory can be classified as a theory of convex functions $f : \mathbf{R}^n \to \mathbf{R}$ that have additional combinatorial properties. Viewed from the discrete side, it is a theory of discrete functions $f : \mathbf{Z}^n \to \mathbf{Z}$ that enjoy certain nice properties comparable to convexity.[3)] Symbolically,

$$\text{Discrete Convex Analysis} = \text{Convex Analysis} + \text{Matroid Theory}.$$

The theory puts emphasis on duality and conjugacy with a view to providing a novel duality framework for nonlinear integer programming. It may be in order to mention that the present theory extends the direction set forth by J. Edmonds, A. Frank, S. Fujishige, and L. Lovász (see §1.1.2), but it is rather independent of the convexity arguments in the theories of greedoids, anti-matroids, convex geometries, and oriented matroids (Björner–Las Vergnas–Sturmfels–White–Ziegler [16], Korte–Lovász–Schrader [114]).

Two convexity concepts, called L-convexity and M-convexity, play primary roles in the present theory. L-convex functions and M-convex functions are both (extensible to) convex functions, and they are conjugate to each other through a discrete version of the Legendre–Fenchel transformation. L-convex functions and M-convex functions generalize, respectively, the concepts of submodular set functions and base polyhedra. It is noted that "L" in "L-convexity" stands for "Lattice" and "M" in "M-convexity" for "Matroid."

1.1.2 History

This section is devoted to an account of the history of discrete convex functions in matroid theory that lead to L-convex and M-convex functions (see Table 1.1). There are, however, many other previous and recent studies on discrete convexity outside the literature of matroid (Hochbaum–Shamir–Shanthikumar [92], Ibaraki–Katoh [93], Kindler [112], Miller [130], and so on).

The concept of matroids was introduced by H. Whitney [218] in 1935, together with the equivalence between submodularity of rank functions and exchange property of independent sets. This equivalence is the germ of the conjugacy between L-convex and M-convex functions in the present theory of discrete convex analysis.

In the late 1960s, J. Edmonds found a fundamental duality theorem on the intersection problem for a pair of (poly)matroids. This theorem, Edmonds' intersection theorem, shows a min-max relation between the maximum of a common independent set and the minimum of a submodular function derived from the rank functions. The famous article of Edmonds [44] convinced us of the fundamental role of submodularity in discrete optimization. Analogies of submodular functions to convex functions and to concave functions were discussed at the same time. The min-max relation supported the analogy to convex functions, whereas some other facts pointed to concave functions. No unanimous conclusion was reached at this point.

The relationship between submodular functions and convex functions was made clear in the early 1980s through the works of A. Frank, S. Fujishige, and L. Lovász, which have been described already in §1.1.1 but is repeated here in view of its importance. The fundamental relationship between submodular functions and convex functions, due to Lovász [123], says that a set function is submodular

[3)] Notation $\mathbf{Z}$ means the set of all integers, and $\mathbf{Z}^n$ the set of n-dimensional integer vectors.

Table 1.1. *History (matroid and convexity)*

Year (ca.)	Author	Result
1935	Whitney [218]	axioms of matroid
		exchange property
		⇔ submodularity
1965	Edmonds [44]	polymatroid
		polyhedral method
		intersection theorem
1975		weighted matroid intersection
	Edmonds [45]	
	Lawler [118]	
	Tomizawa–Iri [201]	potential
	Iri–Tomizawa [96]	potential
	Frank [54]	weight splitting
1982		relationship to convexity
	Frank [55]	discrete separation theorem
	Fujishige [62]	Fenchel-type duality
	Lovász [123]	Lovász (linear) extension
1990	Dress–Wenzel [41] [42]	valuated matroid
		axiom, greedy algorithm
	Favati–Tardella [49]	integrally convex function
1995	Murota [135] [139]	valuated matroid intersection
	Murota [137] [140]	L-/M-convex function
		Fenchel-type duality
		separation theorem
	Murota–Shioura [151]	$M^\natural$-convex function
2000	Fujishige–Murota [68]	$L^\natural$-convex function
	Murota–Shioura [152]	polyhedral L-/M-convex function
	Murota–Shioura [156], [157]	continuous L-/M-convex function

if and only if the Lovász extension of that function is convex. Reformulations of Edmonds' intersection theorem into a separation theorem for a pair of submodular/supermodular functions by Frank [55] and a Fenchel-type min-max duality theorem by Fujishige [62] indicate similarity to convex analysis. The discrete mathematical content of these theorems, which cannot be captured by the relationship of submodularity to convexity, lies in the integrality assertion for integer-valued submodular/supermodular functions. Further analogy to convex analysis such as subgradients was conceived by Fujishige [63]. These developments in the 1980s led us to the understanding that (i) submodularity should be compared to convexity, not to concavity, and (ii) the essence of the duality for a pair of submodular/supermodular functions lies in the discreteness (integrality) assertion in addition to the duality for convex/concave functions:

(i) submodular functions $\simeq$ convex functions,

(ii) duality for submodular functions $\simeq$ convexity + discreteness.

A remark is in order here, although it involves technical terminology from convex analysis. The Lovász extension of a submodular set function is a convex function, but it is bound to be *positively homogeneous* ($f(\lambda x) = \lambda f(x)$ for $\lambda \geq 0$). As a matter of fact, it coincides with the support function of the base polyhedra associated with the submodular function. This suggests that the convexity arguments on submodularity deal with a restricted class of convex functions, namely, the class of support functions of convex sets. The relationship of submodular set functions to convex functions summarized in (i) and (ii) above is generalized to the full extent by the concept of L-convex functions in the present theory.

Addressing the issue of local vs global optimality for functions defined on integer lattice points, P. Favati and F. Tardella [49] came up with the concept of *integrally convex functions* in 1990. This concept successfully captures a fairly general class of functions on integer lattice points, for which a local optimality implies the global optimality. Moreover, the class of *submodular integrally convex functions* (i.e., integrally convex functions that are submodular on integer lattice points) was considered as a subclass of integrally convex functions. It turns out that this concept is equivalent to a variant of L-convex functions, called $\mathrm{L}^\natural$-convex functions, in the present theory.

We have so far seen major milestones on the road towards L-convex functions, and are now turning to M-convex functions.

A weighted version of the matroid intersection problem was introduced by Edmonds [44]. The problem is to find a maximum weight common independent set (or a common base) with respect to a given weight vector. Efficient algorithms for this problem were developed in the 1970s by Edmonds [45], Lawler [118], Tomizawa–Iri [201], and Iri–Tomizawa [96] on the basis of a nice optimality criterion in terms of dual variables. The optimality criterion of Frank [54] in terms of weight splitting can be thought of as a version of such optimality criterion using dual variables. The weighted matroid intersection problem was generalized to the polymatroid intersection problem as well as to the submodular flow problem. It should be noted, however, that in all of these generalizations the weighting remained to be linear or separable convex.

The concept of *valuated matroids*, introduced by Dress and Wenzel [41], [42] in 1990, provides a nice framework of nonlinear optimization on matroids. A valuation of a matroid is a nonlinear and nonseparable function of bases satisfying a certain exchange axiom. It was shown by Dress and Wenzel that a version of greedy algorithm works for maximizing a matroid valuation, and this property in turn characterizes a matroid valuation. Not only the greedy algorithm but the intersection problem extends to valuated matroids. The valuated matroid intersection problem, introduced by Murota [135], is to maximize the sum of two valuations. This generalizes the weighted matroid intersection problem since linear weighting is a special case of matroid valuation. Optimality criteria such as weight splitting as well as algorithms for the weighted matroid intersection are generalized to the valuated matroid intersection (Murota [136]). Analogy of matroid valuations to concave functions resulted in a Fenchel-type min-max duality theorem for matroid valuations (Murota [139]). This Fenchel-type duality is not a generalization nor a special case of Fujishige's Fenchel-type duality for submodular functions, but these two can be generalized into a single min-max equation, which is the Fenchel-type duality theorem in the present theory.

A further analogy of valuated matroids to concave functions led to the concept of *M-convex/concave functions* in Murota [137], 1996. M-convexity is a concept of "convexity" for functions defined on integer lattice points in terms of an exchange axiom, and affords a common generalization of valuated matroids and (integral) polymatroids. A valuated matroid can be identified with an M-concave function defined on $\{0,1\}$-vectors, and the base polyhedron of an integral polymatroid is a synonym for a $\{0,+\infty\}$-valued M-convex function. The valuated matroid intersection problem and the polymatroid intersection problem are unified into the M-convex intersection problem. The Fenchel-type duality theorem for matroid valuations is generalized for M-convex functions, and the submodular flow problem to the M-convex submodular flow problem (Murota [142]), which involves an M-convex function as a nonlinear cost. The nice optimality criterion using dual variables survives in this generalization. Thus, M-convex functions yield fruitful generalizations of many important optimization problems on matroids.

The two independent lines of developments, namely, the convexity argument for submodular functions in the early 1980s and that for valuated matroids and M-convex functions in the early nineties, were merged into a unified framework of "Discrete Convex Analysis," advocated by Murota [140] in 1998. The concept of *L-convex functions* was introduced as a generalization of submodular set functions. L-convex functions form a conjugate class of M-convex functions with respect to the Legendre–Fenchel transformation. This completes the picture of conjugacy advanced by Whitney (1935) as the equivalence between submodularity of the rank function of a matroid and exchange property of independent sets of a matroid. The duality theorems carry over to L-convex and M-convex functions. In particular, the separation theorem for L-convex functions is a generalization of Frank's separation theorem for submodular functions.

Ramifications of the concepts of L- and M-convexity followed. *M$^\natural$-convex functions*[4)], introduced by Murota–Shioura [151], are essentially equivalent to M-convex functions, but are sometimes more convenient. For example, a convex function in one variable, when considered only for integer values of the variable, is an M$^\natural$-convex function that is not M-convex. *L$^\natural$-convex functions*, due to Fujishige–Murota [68], are an equivalent variant of L-convex functions. It turned out that L$^\natural$-convex functions are exactly the same as submodular integrally convex functions that had been introduced by Favati–Tardella (1990) in their study of local vs global optimality.

The success of *polyhedral methods* in combinatorial optimization naturally suggests the possibility of polyhedral versions of L- and M-convex functions. This idea was worked out by Murota–Shioura [152] with the introduction of the concepts of

4) "M$^\natural$-convex" should be read "M-natural-convex," and similarly for "L$^\natural$-convex."

L- and M-convexity for polyhedral functions (piecewise linear functions in real variables). Those convexity concepts were defined also for quadratic functions (Murota–Shioura [155]) and for closed convex functions (Murota–Shioura [156], [157]).

We conclude this section with a remark on a subtle point of the relationship between submodularity and convexity. From the discussion in the early 1980s we have agreed that submodularity should be compared to convexity. This statement is certainly true for set functions. When it comes to functions on integer points, however, we need to be careful. As a matter of fact, an $\mathrm{M}^{\natural}$-concave function is submodular and concave-extensible (Theorems 6.19 and 6.42), whereas an $\mathrm{L}^{\natural}$-convex function is submodular and convex-extensible (Theorem 7.20). This shows that submodularity and convexity are mutually independent properties for functions on integer points. It is undoubtedly true, however, that submodularity is essentially related to discrete convexity.