The Convex Geometry of Network Flows

Theo Diamandis Guillermo Angeris

August 2024

Abstract

In this paper, we derive a number of interesting properties and extensions of the convex flow problem from the perspective of convex geometry. We show that the sets of allowable flows always can be imbued with a downward closure property, which leads to a useful 'calculus' of flows, allowing easy combination and splitting of edges. We then derive a conic form for the convex flow problem, which we show is equivalent to the original problem and almost self-dual. Using this conic form, we consider the nonconvex flow problem with fixed costs on the edges, *i.e.*, where there is some fixed cost to send any nonzero flow over an edge. We show that this problem has almost integral solutions by a Shapley–Folkman argument, and we describe a rounding scheme that works well in practice. Additionally, we provide a heuristic for this nonconvex problem which is a simple modification of our original algorithm. We conclude by discussing a number of interesting avenues for future work.

1 Introduction

Network flows model systems in a wide range of applications, from flows of cars to flows of bits. Traditional network flow models assume a linear input-output relationship for each edge: the flow out of an edge is a linear function of the flow into that edge, and an extensive literature covers the theory, algorithms, and applications of these models. (See, *e.g.*, [AMO88], [Wil19], and references therein.) Unfortunately, many real-world systems do not exhibit this linear relationship. For example, in many practical cases, the output of an edge is a concave function of the input. While nonlinear cost functions have also been extensively explored in the literature (*e.g.*, see [Ber98] and references therein), nonlinear edge flows—when the flow out of an edge is a nonlinear function of the flow into it—has received considerably less attention despite its increased modeling capability.

To model these systems, [DAE24] introduced the convex flow problem, which generalizes the traditional network flow problem to allow for concave input-output relationships and hypergraph structures. This problem also generalizes and extends those considered in [Tru78; Shi06; Vég14]. (See [DAE24] for additional discussion of this problem and related literature.) In this paper, we explore the geometry of the convex flow problem to elucidate a number of interesting properties, many with immediate practical implications.

We begin by requiring a downward closure property on the sets of allowable flows. We prove that requiring this property is equivalent to requiring that the utility functions are nondecreasing, as in [DAE24]. Having downward closed allowable flows has a number of important implications. First, downward closure allows for a useful 'calculus' of flows, enabling the straightforward merging and splitting of network edges. These composition rules have immediate practical implications for solvers.

Second, we derive a conic form for the convex flow problem, which we show is equivalent to the original problem. This conic form allows us to derive a number of theoretical properties, including a dual that closely resembles the original problem, coming close to the self-duality of Bertsekas' extended monotropic programming problem [Ber08].

Finally, the conic problem allows us to extend our analysis to nonconvex flow problems where each edge has an associated fixed cost for sending any nonzero flow, a scenario that commonly appears in real-world problems. We show, via an application of the Shapley–Folkman lemma, that this nonconvex flow problem has 'almost' integral solutions and propose a simple modification to the algorithm from [DAE24] to accommodate these nonconvex scenarios.

Outline. In section 2, we define the convex flow problem and discuss interpretations. We also prove the equivalence between requiring nondecreasing utility functions and requiring a downward closure condition on the sets of allowable flows. In section 3, we show a number of composition rules of downward closed sets which preserve the downward closure. Then, in section 4, we derive an equivalent conic form for the convex flow problem and show that this conic form is almost self-dual. Finally, we consider the nonconvex flow problem with fixed costs on the edges in section 5. We show that this problem is intimately related to the conic form of the convex flow problem described in the previous section and that it has 'almost' integral solutions.

2 Problem set up

In this section, we present the convex flow problem, first defined in [DAE24], give a few simple properties, and discuss some important special cases.

2.1 Problem definition

The *convex flow* problem is the following problem:

$$\begin{array}{ll} \text{maximize} & U(y) + \sum_{i=1}^m V_i(x_i) \\ \text{subject to} & y = \sum_{i=1}^m A_i x_i \\ & x_i \in T_i, \quad i = 1, \dots, m. \end{array} \tag{1}$$

Here, the *network utility function* $U : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ and the *edge utility functions* $V_i : \mathbf{R}^{n_i} \to \mathbf{R} \cup \{+\infty\}$ are concave and nondecreasing, the sets $T_i \subseteq \mathbf{R}^{n_i}$ are nonempty, closed, and convex, and the matrices $A_i \in \mathbf{R}^{n \times n_i}$ with $n_i \le n$ are *selector matrices*. Specifically, A_i is a matrix of the form

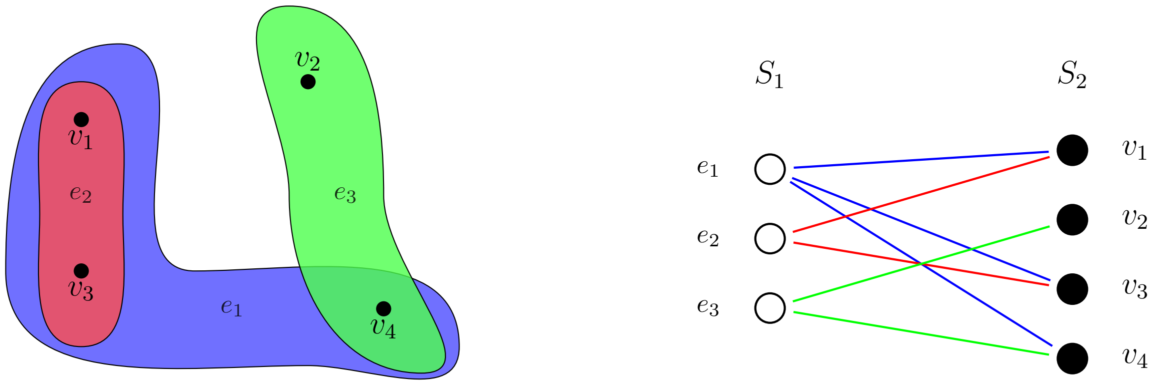
$$A_i = \begin{bmatrix} a_1 & \dots & a_{n_i} \end{bmatrix}, \tag{2}$$

where each $a_k \in \mathbf{R}^n$ is a distinct unit basis vector. The matrix A_i therefore maps the local indices of edge i to the global indices. We also assume that an edge need not be used; *i.e.*, that $0 \in T_i$ for all i. This condition makes the proofs simpler and can always be satisfied by appropriately translating the problem variables and absorbing this translation into the objective terms.

Interpretation: hypergraph flows. Following [DAE24], we can interpret this problem as that of finding the highest-utility allowable flows over a hypergraph with n nodes and m edges, which may be incident to more than two vertices. The variable x_i indicates the flow over the ith hyperedge, which is adjacent to n_i vertices. By convention, we use positive numbers to denote flows out of an edge (equivalently, into a node) and negative numbers to denote flow into an edge (equivalently, out of a node). For example, $(x_i)_k > 0$ is the flow out of the ith edge and into the edge's kth incident vertex. Flows x_i over edge i must lie in T_i, the set of allowable flows. The selector matrices map these edge flows to their corresponding vertices so that y_j is the net flow at node $j = 1, \dots, n$. Figure 1a illustrates this interpretation.

Interpretation: bipartite graph. Alternatively, we may interpret this problem as finding the highest-utility flows over a bipartite graph. Here, a set of m vertices S_1 is connected to a set of n vertices S_2. We denote the flow from node $i \in S_1$ to its kth incident node in S_2 by x_{ik}. The vector x_i is then the set of all flows incident to vertex $i \in S_1$, and we it must lie in the set T_i. Again, positive numbers denote flows from S_1 to S_2, and negative numbers denote flows in the opposite direction. The selector matrices A_i map the flows on edge i to their incident vertices in S_2. Thus, y_j is the net flow at node $j \in S_2$. Figure 1b illustrates this interpretation.

Examples. A number of optimization problems from the literature are special cases of (1). Classic problems, including maximum flow, minimum cost flow, and multi-commodity flows, and their generalization to concave edge input-output relationships, naturally fit into this framework. In engineering, this framework models the problems of optimal power flow and resource allocation in wireless networks, which both have nonlinear edge flow relationships. In economics, this framework includes and generalizes the classic Fisher market model equilibrium problem. And, in finance, this framework models trading through multiple markets, where the output of a market is often a concave function of the input—the more one trades, the worse the price one gets. See [DAE24, §3] for details and discussion of these examples.



(a) Hypergraph interpretation

(b) Bipartite graph interpretation

Figure 1: A hypergraph with 3 edges and 4 nodes (left) and its corresponding bipartite graph representation (right).

2.2 Downward closure and monotonicity

We say that a set $T \subseteq \mathbf{R}^n$ is *downward closed* if, for any $x \in T$ and $x' \leq x$, we have $x' \in T$. In other words, if a flow is feasible, then any smaller flow is also feasible. If $x' \geq x$, we say that the flow x' *dominates* the flow x, since, under any nonnegative utility function, the flow x' is always at least as 'good' as x. In [DAE24], the authors assumed that the functions U and $\{V_i\}$ in the convex flow problem are nondecreasing. This assumption is, in fact, equivalent to the sets $\{T_i\}$ being downward closed in the following sense: if the sets $\{T_i\}$ are downward closed, then the functions U and $\{V_i\}$ can be replaced with their nondecreasing concave envelopes without affecting the optimal objective value. Similarly, if the functions U and $\{V_i\}$ are nondecreasing, then the sets $\{T_i\}$ can be replaced by their downward closures, *i.e.*,

$$\tilde{T}_i = T_i - \mathbf{R}^n_+,$$

without affecting the objective value. This downward closedness property has a number of immediate and useful implications.

Example. As a simple example, consider a directed edge i with maximum input capacity 1 that, when w units of flow enter the edge, outputs $h(w)$ units of Flow. The corresponding set of allowable flows is

$$T_i = \{z \in \mathbf{R}^2 \mid -1 \leq z_1 \leq 0 \;\; \text{and} \;\; z_2 \leq h(-z_1)\}.$$

This set is easily verified to be closed and convex, as it is the intersection of two halfspaces and the hypograph of a concave function, but note that it is not downward closed. Figure 2 shows a T_i and its downward closure $\tilde{T}_i$. The downward closure clearly satisfies the same properties: it is also closed and convex. We show next that, in a general sense, the set T and its downward closure $\tilde{T}$ are 'equivalent' for problem (1) if the functions U and $\{V_i\}$ are nondecreasing.

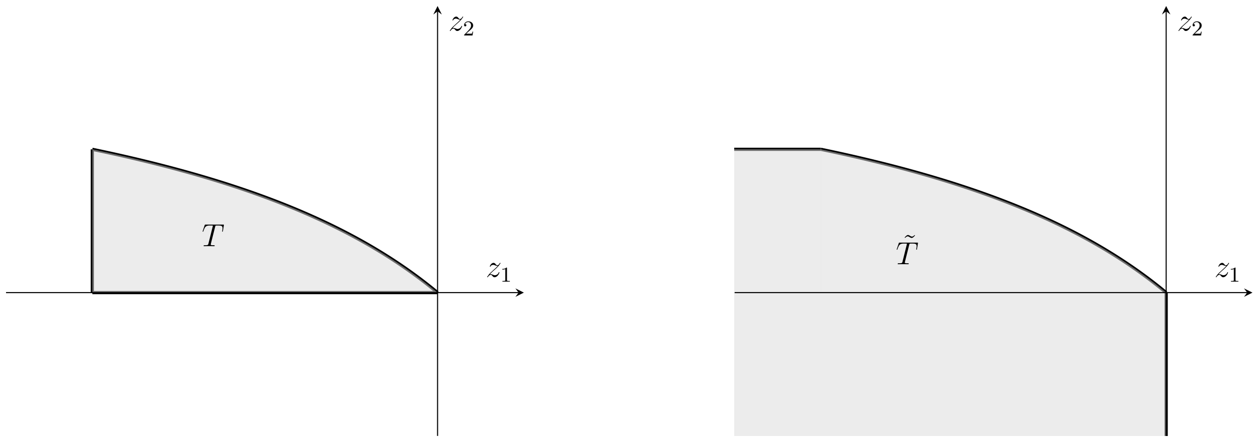


Figure 2: A set of allowable flows T (left) and its downward closure $\tilde{T}$ (right) for a two-node directed edge that, for input w, outputs $h(w) = w/(1+w)$ units of flow.

Equivalence proof. Consider a finite solution $(y^\star, \{x_i^\star\})$ to (1). If the objective functions are nondecreasing, then $U(x') \le U(x)$ for any $x' \le x$, and similarly for the functions $\{V_i\}$. If this solution is not at the boundary, *i.e.*, if there exists some $x_i^\star$ in the relative interior of the corresponding T_i, then we can find a nonnegative direction $d \in \mathbf{R}^{n_i}$ such that $x_i^\star + td \in T_i$ for some $t > 0$. Since the objective functions are nondecreasing, this new point will have a objective value equal to the original solution. As a result, there exists a solution at the boundary, and we can replace the sets $\{T_i\}$ with their 'downward extension',

$$\tilde{T}_i = T_i - \mathbf{R}^n_+,$$

without affecting the solution.

Conversely, if the sets are downward closed and the optimal value is finite, then, by the downward closure of the T_i, there does not exist a nonpositive direction d such that, for some $t > 0$, $x_i^\star + td \in T_i$ and the objective value is larger. (Otherwise, we could find a new point dominated by $x_i^\star$, *i.e.*, in the downward closure of T_i, with a higher objective value.) Equivalently, all subgradients at this solution must be nonnegative: $\partial U(y^\star) \subseteq \mathbf{R}^n_+$ and $\partial V_i(x_i^\star) \subseteq \mathbf{R}^{n_i}_+$ for $i = 1, \dots, m$. This fact immediately suggests that there exists a solution on the boundary and we can replace the objective functions with their monotonic concave envelopes without affecting the solution. We will give an alternate proof of this equivalence in §B, after we have derived a dual problem for (1).

3 A calculus of flows

In light of the previous discussion, we will assume that the sets of allowable flows $\{T_i\}$ are downward closed for the remainder of this paper. We next discuss a number of properties that directly follow from this condition. Much of this section generalizes the authors' previous work in the context of automated market makers [Ang+23, §2]. In the remainder of this section, we will drop subscripts for convenience.

Definition and interpretation. Recall that a set of allowable flows T can be any set satisfying the following properties:

1. The set T is closed and convex.
2. The set T is downward closed: if $x \in T$ and $x' \leq x$, then $x' \in T$.
3. The set T contains the zero vector: $0 \in T$.

The three conditions imposed on the set of allowable flows have a natural interpretation. Convexity means that as more flow enters an edge, the marginal output does not increase. Downward closure means that positive flow (*i.e.*, flow out of an edge) can be dissipated. This property often has a nice interpretation. In power systems, it means that we can dissipate power by, for example, adding a resistive load. In financial markets, it means that we have the option to 'overpay' for an asset. Finally, the last condition means that we need not use an edge. This assumption is not fundamental; we can always translate a set T and absorb the translation into the utility functions. This assumption, however, will simplify some of the proofs later in this paper.

3.1 Composition rules

As a result of the downward closure condition, sets of allowable flows satisfy certain composition rules. Many of these rules follow directly from the calculus of convex sets [BV04, §2.3]. For example, the intersection of two sets of allowable flows yields another set of allowable flows. We discuss a few important composition rules that will be useful in the rest of this paper below.

Nonnegative matrix multiplication. Multiplication of a set of allowable flows by a nonnegative matrix $A \in \mathbf{R}^{p \times k}$ with $\mathcal{N}(A) = \{0\}$, followed by taking the downward closure, results in another set of allowable flows:

$$AT - \mathbf{R}^p_+ = \{x \mid x \leq Ax' \text{ for some } x' \in T\}.$$

This resulting set is downward closed by definition, and also closed and convex. Convexity follows from the fact that convexity is preserved under linear transforms [BV04, §2.3.2] and under downward closure. Closedness of the set follows from [Roc70, Theorem 9.1], as we require A to be injective. This set has a nice interpretation: given some $x \in T$, each element of the vector Ax is a weighted 'meta-flow' with weights given by the rows of A.

Lifting. As a special case of nonnegative matrix multiplication, the lifting of a set of allowable flows into a larger space is also a set of allowable flows. Specifically, let A be a selector matrix (as defined in (2)). Then the set $AT - \mathbf{R}^k_+$ is a set of allowable flows in $\mathbf{R}^{k'}$. This set describes an edge that connects k' vertices but only allows flow between a subset of k of them.

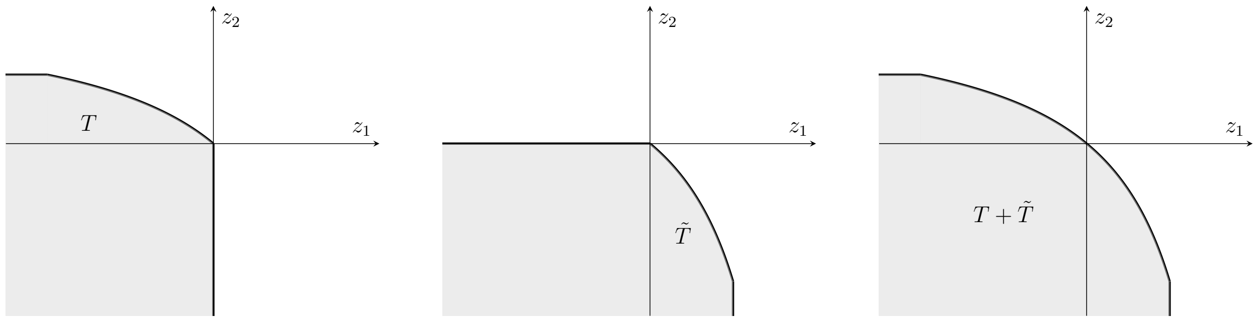


Figure 3: We take the Minkowski sum of the sets of allowable flows T and $\tilde{T}$ for two directed edges with the form of that in figure 2 (left and middle) to obtain a new set of allowable flows $T + \tilde{T}$ that corresponds to an undirected edge (right).

Set addition. Finally, under an additional boundedness assumption, the Minkowski sum of allowable flow sets T and $\tilde{T}$,

$$T + \tilde{T} = \{x + \tilde{x} \mid x \in T, \tilde{x} \in \tilde{T}\},$$

is also a set of allowable flows. For this composition rule, we require that the sets are bounded from above: for a set T, there exists some b such that $x \leq b\mathbf{1}$ for all $x \in T$. This condition means that a bounded input flow cannot produce infinite output flow. We can interpret the combined set $T + \tilde{T}$ as an aggregate edge that can use either of the two original edges; see the example in figure 3.

Aggregate edges. Using the previous two rules, we can combine edges with possibly non-overlapping incident vertices. Importantly, we can view the net flow vector y in (1) as the flow over an 'aggregate edge' that connects all vertices with associated allowable flows

$$T = \sum_{i=1}^{m} \left(A_i T_i - \mathbf{R}_+^{n_i}\right).$$

Thus, when the edge utility functions are equal to zero, the convex network flow problem (1) is equivalent to the following problem over one large aggregate edge:

$$\begin{array}{ll} \text{maximize} & U(y) \\ \text{subject to} & y \in T. \end{array}$$

While this particular rewriting is not immediately useful, combining or splitting certain trading sets, for example those with the same incident vertices, can sometimes help us compute a solution more efficiently. (See [DAE24, §4-5]) for details.)

Example. Often, a directed edge between two nodes has a gain function defined in a piecewise manner. For example, consider a financial market between two assets given by an order book: sellers list the amount of one asset they are willing to sell for the other at a given price. We can view each 'tick' as an individual linear edge, which, when combined, define an aggregate edge corresponding to the entire orderbook. We provide a simple example in figure 4.

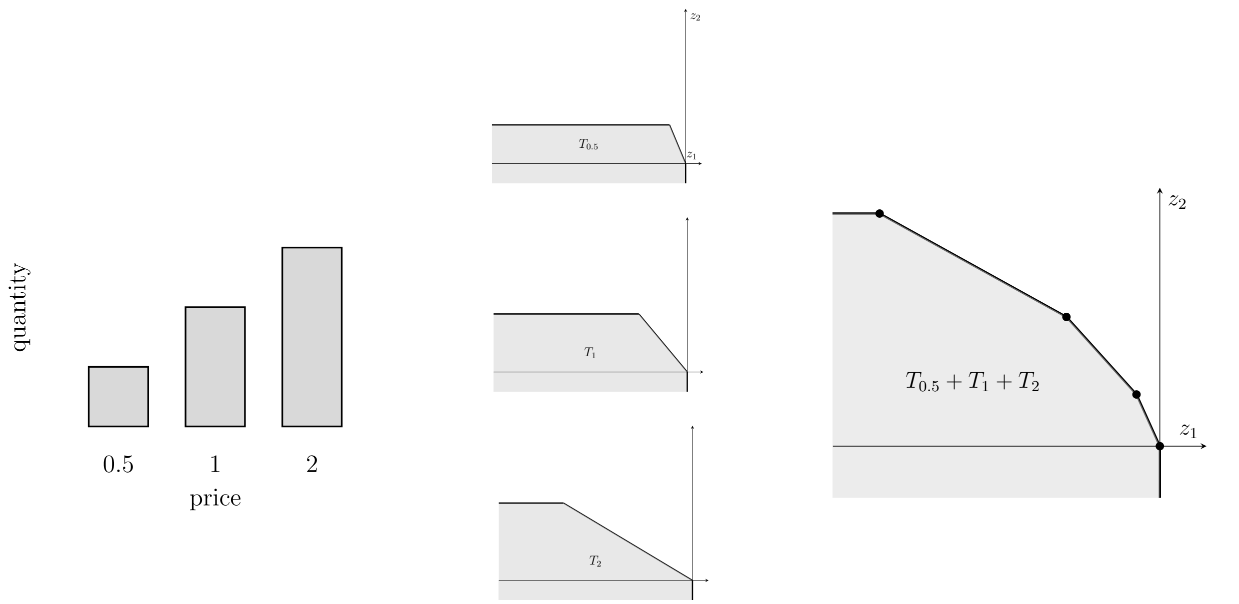


Figure 4: Each tick of the orderbook (left) corresponds to a linear edge with a coefficient corresponding to the price (middle). These linear edges can be combined into an aggregate edge defining the entire orderbook (right).

4 The conic problem

In this section, we will introduce what looks like a restriction of problem (1), which we will call the *conic network flow* problem, defined as

$$\begin{array}{ll} \text{maximize} & U(y) + \sum_{i=1}^m V_i(x_i) \\ \text{subject to} & y = \sum_{i=1}^m A_i x_i \\ & x_i \in K_i, \quad i = 1, \dots, m. \end{array} \tag{3}$$

This problem set up is identical to that of (1), except that the sets K_i, instead of being downward closed convex sets, are downward closed convex cones. A set K_i is called a *cone* if it satisfies the following property: if $x \in K_i$, then, for any $\alpha \geq 0$, we must have that $\alpha x \in K_i$. We call any downward closed convex cone K_i an *allowable flow cone*.

Certainly, every conic flow problem (3) is an instance of a convex network flow problem (1) as every downward closed convex cone is also, by definition, a downward closed convex set. In this section, we will show that the converse is also true: every instance of a convex network flow problem can be turned into an instance of a conic network flow problem. In this sense, problem (1) and problem (3) are equivalent. We will use the conic problem (3) for the remainder of this paper to give a number of important theoretical properties, extensions, and a duality result, all of which easily translate to the original (1), but are much simpler in the conic formulation.

4.1 Basic properties

All of the composition rules presented in §3.1 for the allowable flow sets also hold for the allowable flow cones. More specifically, given two allowable flow cones (*i.e.*, cones that are downward closed) summation, intersection, nonnegative scaling, and nonnegative injective matrix multiplication all yield another allowable flow cone.

Cone is nonpositive. One immediate consequence of the fact that a $K \subseteq \mathbf{R}^d$ is both a cone and downwards closed is that either $K = \mathbf{R}^d$ or K contains no strictly positive vectors; that is,

$$K \cap \mathbf{R}^d_{++} = \varnothing.$$

To see this, let $x \in K$ be any element of K that has only strictly positive entries $x > 0$. Then for every d vector $y \in \mathbf{R}^d$, there exists some $\alpha \geq 0$ such that $y \leq \alpha x$. Since αx is in K, as it is a cone, and K is downward closed, then $y \in K$, as required.

Polar cone. As is standard in convex optimization, given a cone $K \subseteq \mathbf{R}^d$ there exists a *polar cone*, defined

$$K^\circ = \{y \in \mathbf{R}^d \mid y^T x \leq 0 \;\text{ for all }\; x \in K\}. \tag{4}$$

This cone K° is always a closed convex cone (even when K is not). If K is also a closed convex cone, then we have the following duality result $(-K^\circ)^\circ = -K$; in other words, the polar cone of the negative polar cone (called the dual cone) is the negation of the original cone. If, in addition, the cone K is a downward closed cone with any strictly negative element (*i.e.*, there is some $x \in K$ with $x < 0$), then we must have that

$$K^\circ \cap -\mathbf{R}^d_+ = \{0\}.$$

(A sufficient condition for this to hold is, *e.g.*, if the cone K has nonempty interior, which is almost always the case in practice.)

4.2 Reduction

It is clear that the conic problem (3) is a special case of the original problem (1). In this subsection, we will show that the converse is also true: any instance of the original problem can be reduced to an instance of the conic problem.

High level outline. We begin with an instance of (1), which we write again for convenience:

$$\begin{array}{ll}\text{maximize} & U(y) + \sum_{i=1}^m V_i(x_i) \\ \text{subject to} & y = \sum_{i=1}^m A_i x_i \\ & x_i \in T_i, \quad i = 1, \dots, m.\end{array} \tag{1}$$

As in (1) we have some nondecreasing convex network utility function $U : \mathbf{R}^n \to \mathbf{R} \cup \{-\infty\}$, edge utility functions $V_i : \mathbf{R}^{n_i} \to \mathbf{R} \cup \{-\infty\}$, selector matrices $A_i \in \mathbf{R}^{n \times n_i}$, and downward closed sets $T_i \subseteq \mathbf{R}^{n_i}$. (From the previous discussion in section 2.2, we need to require only one of downward closure or montonicity.) The variables are the edge flows $x_i \in \mathbf{R}^{n_i}$ and the network flows $y \in \mathbf{R}^n$. Our goal will be to construct some (simple) nondecreasing net utility function $\tilde{U} : \mathbf{R}^{n+1} \to \mathbf{R} \cup \{-\infty\}$, edge utility functions $\tilde{V}_i : \mathbf{R}^{n_i+1} \to \mathbf{R} \cup \{-\infty\}$, selector matrices $\tilde{A}_i \in \mathbf{R}^{(n+1)\times(n_i+1)}$, and downward closed convex cones $\tilde{K}_i \subseteq \mathbf{R}^{n_i+1}$, such that any solution to the corresponding conic problem (3) over these new functions, matrices, and sets,

$$\begin{array}{ll}\text{maximize} & \tilde{U}(\tilde{y}) + \sum_{i=1}^m \tilde{V}_i(\tilde{x}_i) \\ \text{subject to} & \tilde{y} = \sum_{i=1}^m \tilde{A}_i \tilde{x}_i \\ & \tilde{x}_i \in \tilde{K}_i, \quad i = 1, \dots, m,\end{array}$$

can be (easily) converted to a solution for the original problem (1). We do this process in two steps. First, we define a basic cone $\tilde{K}_i$ associated with each T_i, which is essentially the perspective transformation of T_i, done in such a way as to ensure that $\tilde{K}_i$ is downward closed when T_i is. We then show that any solution over this cone, with an additional constraint, always corresponds to a solution to the original set. Finally, we add this constraint to the objective as an extra term in the edge cost $\tilde{V}_i$.

4.2.1 Flow cone

We define the flow cone corresponding to a set $T_i \subseteq \mathbf{R}^{n_i}$ as

$$\tilde{K}_i = \mathbf{cl}\{(x, -\lambda) \in \mathbf{R}^{n_i} \times \mathbf{R} \mid x/\lambda \in T_i,\ \lambda > 0\}, \tag{5}$$

where **cl** denotes the closure of a set. This definition is just the perspective transformation of the set T_i, with a sign change in the last argument. This set is closed (by definition) and convex (see [BV04, §2.3.3]). It is also downward closed, which we show in appendix A. We can write the downward closedness of the flow cone in the following (slightly more useful) way: given any λ' such that $-1 \le \lambda' \le 0$ then

$$(x, \lambda') \in \tilde{K}_i \quad \text{implies} \quad (x, -1) \in \tilde{K}_i. \tag{6}$$

Recovering the set. We make use of the following (perhaps obvious) observation repeatedly. The set T_i can be easily recovered from the cone $\tilde{K}_i$ by restricting the last coordinate to be equal to -1; that is,

$$T_i = \{x \in \mathbf{R}^{n_i} \mid (x, -1) \in \tilde{K}_i\}, \tag{7}$$

which follows by definition.

Boundary. From the flow cone and downward closure, we may define the set T_i via the homogenous, nondecreasing, convex function

$$\varphi_i(x) = \min\{\lambda \geq 0 \mid (x, -\lambda) \in K_i\}.$$

Equivalently, we may write φ_i as the Minkowski functional

$$\varphi_i(x) = \inf\{\lambda > 0 \mid x/\lambda \in T_i\},$$

This function's one-level set, $\varphi_i(x) = 1$ parameterizes the boundary of T_i. When φ_i has a nice closed form, this function can be useful in practical applications (see, for example [Dia+23]).

4.2.2 Rewriting via the flow cone

We start with the original problem (1). Using the cone defined previously in equation (7), we rewrite the original problem using the flow cone as

$$\begin{array}{ll} \text{maximize} & U(y) + \sum_{i=1}^m V_i(x_i) \\ \text{subject to} & y = \sum_{i=1}^m A_i x_i \\ & (x_i, \lambda_i) \in \tilde{K}_i, \quad i = 1, \ldots, m \\ & \lambda_i \geq -1, \quad i = 1, \ldots, m, \end{array} \tag{8}$$

where we have relaxed the constraint that $\lambda_i = -1$ to an inequality. This relaxation is exact by the 'dominating points' result (6): any solution with (x_i, λ_i) may be replaced with a solution $(x_i, -1)$ which is also feasible, so $x_i \in T_i$ by (7). Indeed, in many cases, such as when the set T_i is locally strictly concave around 0, one can show that if $\lambda > -1$, there exists a strictly dominating point x_i' such that $x_i' > x_i$ and $(x_i', -1) \in \tilde{K}_i$, so choosing $\lambda_i > -1$ is never optimal.

4.2.3 Final rewriting

Finally, we will take the conic relaxation given in (8), which, due to the constraint $\lambda_i \geq -1$ is not quite a conic flow problem (3), and replace the matrices A_i, edge cost functions V_i, and variables (x_i, λ_i) to get a problem of the required form.

The first part is easy: let $I(z \geq -1) = 0$ if $z \geq -1$ and $+\infty$ otherwise be the nonnegative indicator function for a scalar. Note that I is nonincreasing so $-I$ is nondecreasing, which means we can rewrite (8) by pulling the constraint into the objective

$$\begin{array}{ll} \text{maximize} & U(y) + \sum_{i=1}^m V_i(x_i) - I(\lambda_i \geq -1) \\ \text{subject to} & y = \sum_{i=1}^m A_i x_i \\ & (x_i, \lambda_i) \in \tilde{K}_i, \quad i = 1, \ldots, m. \end{array} \tag{9}$$

Finally, we define the matrix

$$\tilde{A}_i = \begin{bmatrix} A_i & 0 \\ 0 & 1 \end{bmatrix}.$$

which is just the matrix A_i with an additional row and column. Setting $\tilde{x}_i = (x_i, \lambda_i)$ gives the final result:

$$\begin{array}{ll} \text{maximize} & \tilde{U}(\tilde{y}) + \sum_{i=1}^m \tilde{V}_i(\tilde{x}_i) \\ \text{subject to} & \tilde{y} = \sum_{i=1}^m \tilde{A}_i \tilde{x}_i \\ & \tilde{x}_i \in \tilde{K}_i, \quad i = 1, \dots, m, \end{array} \tag{10}$$

where we have defined

$$\tilde{V}(x_i, \lambda_i) = V_i(x_i) - I(\lambda_i) \qquad \text{and} \qquad \tilde{U}(y, \tilde{\lambda}) = U(y).$$

This problem is exactly of the form of the conic flow problem (3), as required.

4.3 Duality

Now that we know problem (1) and problem (3) are essentially equivalent (even though the conic problem (3) 'seems' more restrictive) we give a dual reformulation of (3) that is 'almost' self-dual in this section.

4.3.1 Dual problem

We will write a simple dual for the conic problem (3) using standard duality results and a basic rewriting of the problem.

Lagrangian. First, we write problem (3) here again for convenience:

$$\begin{array}{ll} \text{maximize} & U(y) + \sum_{i=1}^m V_i(x_i) \\ \text{subject to} & y = \sum_{i=1}^m A_i x_i \\ & x_i \in K_i, \quad i = 1, \dots, m. \end{array} \tag{3}$$

We pull the conic constraint $x_i \in K_i$ into the objective by defining the indicator functions

$$I(x_i \in K_i) = \begin{cases} 0 & x_i \in K_i \\ +\infty & \text{otherwise,} \end{cases}$$

for $i = 1, \dots, m$. We can then rewrite the conic problem as

$$\begin{array}{ll} \text{maximize} & U(y) + \sum_{i=1}^m V_i(x_i) - I(\tilde{x}_i \in K_i) \\ \text{subject to} & y = \sum_{i=1}^m A_i x_i \\ & \tilde{x}_i = x_i, \quad i = 1, \dots, m, \end{array}$$

where we have introduced the new redundant variables $\tilde{x}_i \in \mathbf{R}^{n_i}$ for each $i = 1, \dots, m$. This resulting problem is just a convex problem with linear constraints. Introducing the Lagrange multipliers $\nu \in \mathbf{R}^n$ for the first equality constraint and $\eta_i \in \mathbf{R}^{n_i}$ for the second equality constraint, we get the Lagrangian:

$$L(x, \tilde{x}, y, \nu, \eta) = U(y) + \sum_{i=1}^m (V_i(x_i) - I(\tilde{x}_i \in K_i)) + \nu^T \left(y - \sum_{i=1}^m A_i x_i \right) + \sum_{i=1}^m \eta_i^T (x_i - \tilde{x}_i).$$

Dual function. To find the dual function (and therefore the dual problem) we partially maximize L over the primal variables x, $\tilde{x}$, and y:

$$g(\nu, \eta) = \bar{U}(\nu) + \sum_{i=1}^{m} \bar{V}_i(\eta_i - A_i^T \nu) + \sum_{i=1}^{m} \bar{I}_i(\eta_i). \tag{11}$$

Here we have defined

$$\bar{U}(\nu) = \sup_y \left(U(y) + \nu^T y \right), \quad \bar{V}_i(\xi_i) = \sup_{x_i} \left(V_i(x_i) + \xi_i^T x_i \right),$$

and the functions $\{\bar{I}_i\}$ as

$$\bar{I}_i(\eta_i) = \sup_{\tilde{x}_i} \left(-I(\tilde{x}_i \in K_i) + \tilde{x}_i^T \eta_i \right),$$

for each $i = 1, \dots, m$. Note that the function $\bar{I}_i$ is simply the indicator for the polar cone of K_i, defined in (4). In other words,

$$\bar{I}_i(\eta_i) = \begin{cases} 0 & \eta_i \in K_i^\circ \\ +\infty & \text{otherwise.} \end{cases}$$

Dual problem. The dual problem is then to minimize the dual function g; *i.e.*,

$$\text{minimize} \quad g(\nu, \eta).$$

When there exists a point in the relative interior of the domain, strong duality holds and, therefore, the optimal values of the dual problem and the primal problem are identical. Plugging in the definition of g from (11) into the objective of the dual problem, and pulling out the indicator functions $\{\bar{I}_i\}$ into explicit constraints gives

$$\begin{array}{ll} \text{minimize} & \bar{U}(\nu) + \sum_{i=1}^m \bar{V}_i(\eta_i - A_i^T \nu) \\ \text{subject to} & \eta_i \in K_i^\circ, \quad i = 1, \dots, m. \end{array}$$

We can rewrite the problem to make it more similar to the original (3). If we define $\xi_i = A_i^T \nu$ then

$$D\nu = \sum_{i=1}^{m} A_i \xi_i,$$

where D is a diagonal matrix

$$D = \sum_{i=1}^{m} A_i A_i^T,$$

with nonnegative diagonal entries. The jth diagonal entry, D_{jj}, denotes the degree of node j, for $j = 1, \dots, n$. The diagonal entries of D are strictly positive if the hypergraph corresponding to the A_i has no isolated nodes, or, equivalently, if, for each node $j = 1, \dots, n$

there is some edge $i = 1, \ldots, m$ such that the jth row of A_i is nonzero. In this case, which we may always assume in practice by removing isolated nodes, the inverse of D exists so the relationship between ν and the ξ_i is bijective. This means we can rewrite the dual problem:

$$\begin{array}{ll} \text{minimize} & \bar{U}(\nu) + \sum_{i=1}^m \bar{V}_i(\eta_i - \xi_i) \\ \text{subject to} & D\nu = \sum_{i=1}^m A_i \xi_i \\ & \eta_i \in K_i^\circ, \quad i = 1, \ldots, m. \end{array}$$

We may absorb the matrix D into the definition of $\bar{U}$ by replacing $\bar{U}(\nu)$ with $\bar{U}(D^{-1}\nu)$ to get the slightly more familiar-looking problem

$$\begin{array}{ll} \text{minimize} & \bar{U}(\nu) + \sum_{i=1}^m \bar{V}_i(\eta_i - \xi_i) \\ \text{subject to} & \nu = \sum_{i=1}^m A_i \xi_i \\ & \eta_i \in K_i^\circ, \quad i = 1, \ldots, m. \end{array} \tag{12}$$

The dual variables may be interpreted as node dual prices $\nu \in \mathbf{R}^n$ and edge dual prices $\xi_i \in \mathbf{R}^{n_i}$, for $i = 1, \ldots, n$. We call this problem the *dual conic flow problem.* Compare this problem (12) with the original conic flow problem (3).

5 Fixed fees

Finally, we consider the convex network flow problem with fixed fees for the use of an edge. In particular, we consider the following extension of the convex network flow problem (1), which we call the *network flow problem with fees*:

$$\begin{array}{ll} \text{maximize} & U(y) + \sum_{i=1}^m V_i(x_i) + q_i \lambda_i \\ \text{subject to} & y = \sum_{i=1}^m A_i x_i \\ & (x_i, \lambda_i) \in \{0\} \cup (T_i \times \{-1\}), \quad i = 1, \ldots, m, \end{array} \tag{13}$$

where the set up and variables are exactly those of the original convex flow problem (1) except with the additional variable $\lambda \in \mathbf{R}^m$ and the problem data has the additional fee vector $q \in \mathbf{R}_+^m$. This objective is also nondecreasing in all of its variables (as V_i and U are, along with the fact that $q \geq 0$). We note that this problem is not convex since the constraint set is not convex (in fact, this constraint set is not even connected!) and the problem is NP-hard to solve, which we prove shortly. However, we will show that this problem can be approximately solved quite efficiently in practice and is intimately related to the conic form problem (3) introduced in the previous section.

Interpretation. The interpretation of the constraint

$$(x_i, \lambda_i) \in \{0\} \cup (T_i \times \{-1\}) \tag{14}$$

is that if x_i is nonzero, then $\lambda_i = -1$. In other words, if we use edge i by putting any nonzero flow through it, then $x_i \neq 0$ and we are charged $q_i \geq 0$ for its use. In general, we note that if $q_i > 0$ and $x_i = 0$, then we will have $\lambda_i = 0$ at optimality, so we may view λ_i as a variable that indicates whether or not edge i is being used.

NP-hardness. We show that the network flow problem with fees is NP-hard by reducing the knapsack problem, which is known to be NP-hard [Kar72], to an instance of (13). The knapsack problem is the following: given a vector of nonnegative integers $c \in \mathbf{Z}_+^m$ and some integer $b \geq 0$, find a binary vector $z \in \{0, 1\}^m$ such that $c^T z = b$. This problem can be reduced to an instance of (13) with $n = 1$ by setting $U(y) = y - I(y \geq b)$, $A_i = 1 \in \mathbf{R}$, $V_i = 0$, $T_i = \{z \mid z \leq c_i\}$, and $q = c$. The problem is

$$
\begin{array}{ll}
\text{maximize} & -I(y \geq b) + c^T \lambda \\
\text{subject to} & y = \sum_{i=1}^m A_i x_i \\
& (x_i, \lambda_i) \in \{(0, 0)\} \cup ((-\infty, c_i] \times \{-1\}), \quad i = 1, \dots, m.
\end{array}
$$

Note that $c^T(-\lambda) \geq y$ for any feasible point. Since y is constrained to be at least b then we have that $c^T(-\lambda) \geq b$ for any feasible point. Finally, maximizing $c^T \lambda$ is the same as minimizing $c^T(-\lambda) \geq b$, and equality is achieved if and only if there exists $\lambda \in \{-1, 0\}^m$ such that $c^T(-\lambda) = b$; or, equivalently, when the optimal objective value of this problem is exactly equal to $-b$. If it were easy (*i.e.*, polynomial time) to solve this problem, it would be easy to find a solution to the knapsack problem by verifying that $c^T(-\lambda^\star) = b$, or to assert that no solution exists if the problem is infeasible or has optimal objective value larger than $-b$, making this problem at least as hard as knapsack, which is known to be NP-hard.

5.1 Integrality constraint

For the sake of convenience, we will define the set

$$
Q_i = \{0\} \cup (T_i \times \{-1\}),
$$

such that the constraint (14) can be written as

$$
(x_i, \lambda_i) \in Q_i,
$$

for each $i = 1, \dots, m$. In a certain sense, this constraint encodes the 'hard' part of the problem: if the set Q_i were convex, then the problem would almost be a special case of the original convex flow problem (1), by pulling the constraint that $\lambda_i \geq -1$ into the objective.

Convex relaxation. Given the above discussion, we next examine the convex hull of Q_i; if we can easily write this convex hull in a compact way, then we immediately have a convex relaxation of the potentially hard problem (13). In general, finding the convex hull of a set may be challenging, *e.g.*, even describing a convex hull can require an exponential number of constraints. In this particular special case, we will show that the convex hull of the set Q_i is intimately related to the flow cone (5) introduced in the rewriting of the original convex flow problem (1) into the conic flow problem (3). In many practical scenarios, finding the flow cone K_i corresponding to the allowable flows T_i is fairly straightforward, which, in turn, means that finding the convex hull of Q_i is also fairly straightforward.

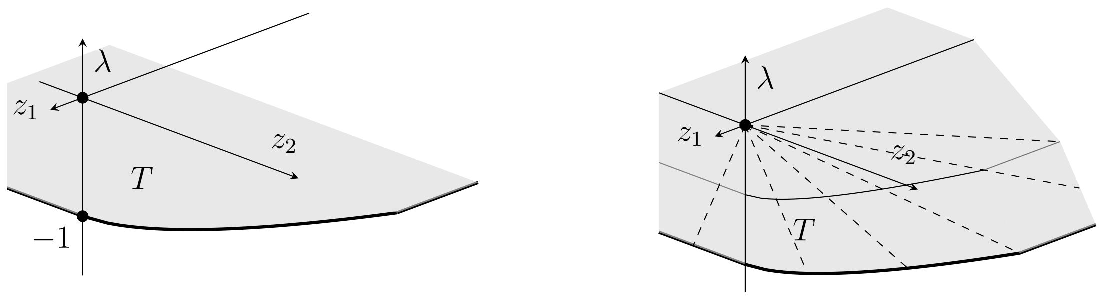


Figure 5: The set Q_i (left) and its convex hull $\mathbf{conv}(Q_i)$ (right).

5.2 Convex hull

Given the above discussion, we will show that the convex hull of Q_i, written $\mathbf{conv}(Q_i)$, is equal to all elements of the corresponding flow cone K_i when the elements' last entry (corresponding to λ_i) lies between -1 and 0. Written out we will show that:

$$\mathbf{conv}(Q_i) = K_i \cap (\mathbf{R}^n \times [-1, 0]).$$

See figure 5 for an example. As a reminder, the flow cone $K_i \subseteq \mathbf{R}^{n+1}$ for a given allowable flow set $T_i \subseteq \mathbf{R}^n$ is defined, using (5), as

$$K_i = \mathbf{cl}\{(x, -\lambda) \in \mathbf{R}^n \times \mathbf{R} \mid x/\lambda \in T_i,\ \lambda > 0\}.$$

To simplify notation, we define

$$\bar{K}_i = K_i \cap (\mathbf{R}^n \times [-1, 0]), \tag{15}$$

which is the cone K_i with the last element restricted to lie between -1 and 0. Of course, this set is also convex as it is the intersection of two convex sets.

Reverse inclusion. First, we show the reverse inclusion: that $\bar{K}_i \subseteq \mathbf{conv}(Q_i)$. Let $(x, -\lambda) \in \bar{K}_i$ (note the negative here) with $\lambda > 0$, then, we will show that $(x, -\lambda)$ can be written as the convex combination of an element in $T_i \times \{-1\}$ and the zero vector, and so also must lie in $\mathbf{conv}(Q_i)$. By definition, if $(x, -\lambda) \in \bar{K}_i$, then $x/\lambda \in T_i$ for $0 < \lambda \leq 1$. But, this is the same as saying

$$(x/\lambda, -1) \in Q_i.$$

Finally, since $0 \in Q_i$, then any convex combination of 0 and $(x/\lambda, -1)$ is in the convex hull of Q_i. Thus,

$$(x, -\lambda) = \lambda(x/\lambda, -1) + (1 - \lambda)0 \in \mathbf{conv}(Q_i),$$

so long as $\lambda > 0$. On the other hand, if $\lambda = 0$, then we know that $x/\lambda' \in T_i$ for all $\lambda' > 0$, so

$$(x, \lambda') = \lambda'(x/\lambda', -1) + (1 - \lambda')0 \in \mathbf{conv}(Q_i).$$

Sending $\lambda' \to 0$ gives the result, since Q_i is closed as it is the union of two closed sets. Putting it all together, this implies that $\mathbf{conv}(Q_i) \supseteq \bar{K}_i$.

Forward inclusion. Now, we show the forward inclusion: that $\mathbf{conv}(Q_i) \subseteq \bar{K}_i$. Note that $Q_i \subseteq \bar{K}_i$ since, by definition

$$T_i \times \{-1\} \subseteq \bar{K}_i,$$

and $0 \in \bar{K}_i$, also essentially by definition. This immediately implies that

$$\mathbf{conv}(Q_i) \subseteq \mathbf{conv}(\bar{K}_i) = \bar{K}_i$$

where we have used the fact that the convex hull of a convex set is itself.

Discussion. Putting the above two points together, we get the claim that

$$\mathbf{conv}(Q_i) = \bar{K}_i. \tag{16}$$

In other words, the convex hull of the 'hard' set is exactly the flow cone K_i with the additional constraint that the last entry must be restricted to lie between 0 and -1. One interesting interpretation of this claim is that we may view the cone K_i as the *conic completion* of the set Q_i. More generally, $\mathbf{cone}(Q_i)$ is defined as the set containing all conic (*i.e.*, nonnegative) combinations of the elements of Q_i. Since is it not hard to show that $\mathbf{cone}(Q_i) = \mathbf{cone}(\mathbf{conv}(Q_i))$, we have

$$\mathbf{cone}(Q_i) = \mathbf{cone}(\mathbf{conv}(Q_i)) = \mathbf{cone}(\bar{K}_i) = K_i,$$

where the second equality follows from (16), while the last simply follows from definitions.

5.3 Convex relaxation

Using the result (16) derived in the previous section, a convex relaxation of the network problem with fees (13) is

$$\begin{array}{ll} \text{maximize} & U(y) + \sum_{i=1}^m (V_i(x_i) + q_i\lambda_i) \\ \text{subject to} & y = \sum_{i=1}^m A_i x_i \\ & (x_i, \lambda_i) \in K_i, \ -1 \le \lambda_i \le 0, \quad i = 1, \dots, m. \end{array} \tag{17}$$

Note that we have replaced the nonconvex constraint $(x_i, \lambda_i) \in Q_i$ with the convex constraint $(x_i, \lambda_i) \in \mathbf{conv}(Q_i)$, or, equivalently, using the facts derived in the previous section, replaced it with the constraint that $(x_i, \lambda_i) \in K_i$ and $-1 \le \lambda_i \le 0$.

Conic formulation. This convex relaxation is also a special case of the conic flow problem (3) in a very natural way. First, note that the constraint that $\lambda \le 0$ is redundant using the definition of K_i. We may then pull the remaining constraint on λ_i, that $\lambda_i \ge -1$ into an indicator function, $I(\lambda_i \ge -1)$ and place it in the objective. This indicator function is nonincreasing, so its negation is nondecreasing, and we get the final problem

$$\begin{array}{ll} \text{maximize} & U(y) + \sum_{i=1}^m \left(V_i(x_i) + q_i\lambda_i - I(\lambda_i \ge -1)\right) \\ \text{subject to} & y = \sum_{i=1}^m A_i x_i \\ & (x_i, \lambda_i) \in K_i, \quad i = 1, \dots, m. \end{array} \tag{18}$$

If we use the same trick used in §4.2 to rewrite the matrices A_i, we receive an instance of the conic flow problem (3) since the objective is nondecreasing and the $\{K_i\}$ are cones. Indeed, the formulation found here (18) is essentially identical to the formulation found in (9), except with the addition of the fixed costs $q \ge 0$.

Integrality gap. In the previous argument in §4.2.2, we used the fact that, if $\lambda_i > -1$, then we could set $\lambda_i = -1$ and always remain feasible with no change in objective value. However in problem (18), the objective value would decrease, so this argument does not apply. We will show next that we expect the solution to be close to integral in the special case that the $V_i = 0$, which is common in practice (see, for example, the applications in [DAE24, §3]).

5.4 Tightness of the relaxation

Just how tight do we expect the relaxation to be? We will show that in the case that $V_i = 0$, if m, the number of edges, is much larger than n, the number of nodes, then most of the λ_i will be integral. More specifically, we will show that, given any solution to the relaxation, we can recover a solution such that at least $m - n - 1$ indices i satisfy $(x_i, \lambda_i) \in Q_i$. If $m \gg n$, *i.e.*, the number of nodes is much smaller than the number of edges, as is usually the case in practice, then most of the solution is integral.

Shapley–Folkman lemma. We state the Shapley–Folkman lemma here in its standard form. Let $S_1, \dots, S_m \subseteq \mathbf{R}^{n+1}$ be any subsets (convex or nonconvex) of $\mathbf{R}^{n+1}$. Then, for any

$$y = x_1 + \cdots + x_m,$$

where $x_i \in \mathbf{conv}(S_i)$ for $i = 1, \dots, m$, there exists $\tilde{x}_i \in \mathbf{conv}(S_i)$ with $i = 1, \dots, m$, such that

$$y = \tilde{x}_1 + \cdots + \tilde{x}_m,$$

which satisfy $\tilde{x}_i \in S_i$ for at least $m - n - 1$ indices i. In other words, given any vector y, which lies in the (Minkowski) sum of the convex hulls of the S_i, we can find $\tilde{x}_i$, which sum to y, such that at least $m-n-1$ lie in the original sets S_i, while the remainder lie in the convex hull, $\mathbf{conv}(S_i)$. Intuitively, this lemma states that the sum of convex sets becomes closer and closer to its convex hull as the number of sets gets large. See figure 6 for an example.

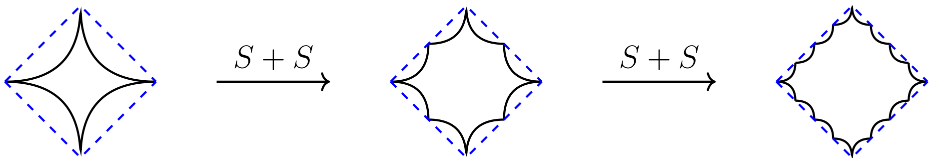


Figure 6: A visual representation of the Shapley–Folkman lemma for the (nonconvex) 1/2-norm ball. As we take the Minkowski sum of the set with itself, it becomes closer and closer to its convex hull, the 1-norm ball.

Given a solution to the convex relaxation (18), we will use this lemma to construct a solution that has the same objective value as the original, yet almost all penalties λ_i will be integral: either -1 or 0. This will then let us construct an approximate solution to (13) and bound the difference between the optimal objective value of (13) and that of the relaxation (17).

Constructing an approximate solution. Assume we are given feasible flows and penalties for the relaxation; *i.e.*, we have a solution $\{(x_i^\star, \lambda_i^\star)\}$ to the relaxation of the fixed-fee problem (17). From this solution to the relaxation (17), we will construct a feasible point for the original fixed-fee problem (13) which we will then show is 'close' to the optimal value, under certain conditions. We write the exact problem we are considering (the special case of (17) when $V_i = 0$) for convenience:

$$\begin{array}{ll} \text{maximize} & U(y) + q^T\lambda \\ \text{subject to} & y = \sum_{i=1}^m A_i x_i \\ & (x_i, \lambda_i) \in \bar{K}_i, \quad i = 1, \dots, m. \end{array}$$

Here, we have used the definition of $\bar{K}_i$ from (15), and, from the previous discussion (16), we know that $\bar{K}_i = \mathbf{conv}(Q_i)$. First, note that, by definition

$$y^\star = \sum_{i=1}^m A_i x_i^\star,$$

and that $(x_i^\star, \lambda_i^\star) \in \mathbf{conv}(Q_i)$ for $i = 1, \dots, m$. Now, from the problem statement, we have

$$c = q^T \lambda^\star,$$

with $q \geq 0$, where $c \leq 0$ stands for the 'fixed cost' part of the objective. Rewriting slightly,

$$\begin{bmatrix} y^\star \\ c \end{bmatrix} = \sum_{i=1}^m \begin{bmatrix} A_i & 0 \\ 0 & q_i \end{bmatrix} \begin{bmatrix} x_i^\star \\ \lambda_i^\star \end{bmatrix},$$

where $(x_i^\star, \lambda_i^\star) \in \bar{K}_i$, or, equivalently, using (16), $(x_i^\star, \lambda_i^\star) \in \mathbf{conv}(Q_i)$, for $i = 1, \dots, m$. From the Shapley–Folkman lemma, there exist $(\tilde{x}_i^\star, \tilde{\lambda}_i^\star) \in \mathbf{conv}(Q_i)$ such that

$$\begin{bmatrix} y^\star \\ c \end{bmatrix} = \sum_{i=1}^m \begin{bmatrix} A_i & 0 \\ 0 & q_i \end{bmatrix} \begin{bmatrix} \tilde{x}_i^\star \\ \tilde{\lambda}_i^\star \end{bmatrix},$$

and at least $m - n - 1$ indices i satisfy $(\tilde{x}_i^\star, \tilde{\lambda}_i^\star) \in Q_i$. In other words, if $m \gg n$, we have an 'almost' feasible solution for the original problem (13), except for $n + 1$ indices i. Consider these nonintegral indices, of which there are at most $n + 1$. In this case, we know, from the dominated point condition (6) for K_i (and therefore for $\bar{K}_i$) that if $(\tilde{x}_i^\star, \tilde{\lambda}_i^\star) \in \bar{K}_i$, then $(\tilde{x}_i^\star, -1) \in \bar{K}_i$. But we know that $(\tilde{x}_i^\star, -1) \in Q_i$, making this point also feasible for the original problem (13). In English: if we were charged less than the full amount due to the relaxation (*i.e.*, we were charged $q_i \lambda_i$), we can always choose to be charged the full amount for the same flow $(-q_i)$ and be feasible for the original problem. This means that a feasible solution for the original problem (13) will be to set

$$(x_i^0, \lambda_i^0) = \begin{cases} (\tilde{x}_i^\star, \tilde{\lambda}_i^\star) & (\tilde{x}_i^\star, \tilde{\lambda}_i^\star) \in Q_i \\ (\tilde{x}_i^\star, -1) & \text{otherwise}, \end{cases} \tag{19}$$

for $i = 1, \dots, m$. Note that $(x_i^0, \lambda_i^0) \in Q_i$ for each i and so is a feasible point for the original fixed-fee problem (13), leading to the same net flows $y^\star$ as the solution to the relaxation, but the cost incurred, $q^T \lambda^0$ differs by at most

$$q^T(\lambda^0 - \tilde{\lambda}^\star),$$

from that of the relaxation, $q^T \tilde{\lambda}^\star = c$. Since most entries of $\lambda^0 - \tilde{\lambda}^\star$ are zero, by the previous argument, then we expect this cost to be small. We give a simple bound on this, and therefore in the objective gap between the relaxation and the original problem, in what follows.

Bounding the optimal objective value. Let p^0 be the optimal objective value for the relaxation (17) and let $p^\star$ be the optimal objective value for the original problem (13). Then, since (17) is a relaxation of (13), we know that

$$p^\star \le p^0.$$

From the previous discussion, we have a feasible point (19) for the original problem. By construction, we know that the net flows $y^\star$ remain unchanged, so the net utility $U(y^\star)$ in the objective similarly remains unchanged. On the other hand, the cost incurred $q^T \lambda^0$ is larger than that of the relaxation, c, by $q^T(\tilde{\lambda}^\star - \lambda^0)$, so we have the following bound

$$p^0 + q^T(\lambda^0 - \tilde{\lambda}^\star) \le p^\star \le p^0.$$

If we solved the relaxation, then we immediately have a two-sided bound on the optimal objective value as given above. On the other hand, we can give a simple bound that does not require solving the relaxation. Since we know that at most $n + 1$ entries of λ^0 will differ from those of $\tilde{\lambda}^\star$ by Shapley–Folkman, then

$$p^0 - (n+1)\left(\max_i q_i\right) \le p^\star \le p^0.$$

Or, equivalently

$$0 \le p^0 - p^\star \le (n+1)(\max_i q_i).$$

5.5 Fixed cost dual problem

Finally, we derive the dual problem of the fixed-fee problem (13) with zero edge costs and show that the algorithm developed in [DAE24] can still be applied to this problem with minimal modifications. In fact, we lose little computational efficiently by solving the fixed-fee problem directly; of course, the solution is not guaranteed to be optimal.

Dual function. Using a similar derivation to that in §4.3, we write the dual function as

$$g(\nu) = \bar{U}(\nu) + \sum_{i=1}^{m} \sup_{(x_i, \lambda_i) \in Q_i} \left((A_i^T \nu)^T x_i + \lambda_i q_i \right).$$

Note that the support function over Q_i, *i.e.*, the expression in the sum, can be easily evaluated: we simply compute the support function of T_i less q_i and compare this value to 0. If we define

$$f_i(\xi) = \sup_{x_i \in T_i} \xi^T x_i, \tag{20}$$

then we can write

$$\sup_{(x_i, \lambda_i) \in Q_i} \left(\xi^T x_i + \lambda_i q_i \right) = \max \left\{ f_i(\xi) - q_i,\, 0 \right\}.$$

The optimal points for the support function are as follows. If $f_i(\xi) \geq q_i$, then $\lambda_i^\star = -1$ and $x_i^\star$ can be any solution to the subproblem (20). On the other hand, if $f_i(\xi) \leq q_i$, then $\lambda_i^\star = 0$ and $x_i^\star = 0$ is a solution. (There may be many solutions if $f_i(\xi) = q_i$.) This observation allows us to apply the algorithm from [DAE24] 'off the shelf' to solve the dual problem.

Primal feasibility. By construction, using the $\lambda_i^\star$ and $x_i^\star$ from above always results in integral solutions to the edge subproblem. Of course, these solutions may not be primal feasible: the net flows $y^\star$ solving the $\bar{U}$ subproblem may not be equal to the sum of the solutions $x_i^\star$ to the edge subproblems; *i.e.*, it is possible that

$$y^\star = \sum_{i=1}^{m} A_i x_i^\star,$$

is not true. However, we can always construct a primal feasible net flow $\hat{y}$ using these subproblem solutions $\{x_i^\star\}$ by setting

$$\hat{y} = \sum_{i=1}^{m} A_i x_i^\star.$$

Verifying optimality. We can verify the optimality of the primal feasible point $(\hat{y}, \{x_i^\star\})$ by checking it's objective value. From duality, we know that

$$g(\nu^\star) = U(y^\star) \geq U(\hat{y}).$$

If these values are equal, we know that the point $(\hat{y}, \{x_i^\star\})$ is also optimal.

5.6 Numerical experiments

Here, we present some simple numerical experiments to illustrate the results of this section. We emphasize the these experiments are by no means exhaustive, but we leave a more thorough investigation to future work.

Setup. We consider the order routing problem from [Dia+23; Ang+22] with a network of n assets and $m = (1/4)n^2$ markets between these assets. Each market is a constant function market maker [AC20; Ang+23] that allows trades between a randomly selected pair of assets and has a strictly concave, increasing edge gain function. To interact with a market (*i.e.*, to use this edge), a trader must pay a fixed fee $q_0 \in \mathbf{R}_+$. The trader's goal is to maximize their utility $U(y)$ of the net trade y, given by

$$U(y) = c^T y - \frac{\mu}{2} y^T y,$$

where $\mu > 0$ is some risk aversion parameter. The objective function is, therefore,

$$c^T y - \frac{\mu}{2} y^T y + q_0(\mathbf{1}^T \lambda)$$

We solve the convex relaxation of the fixed fee problem (17) using the open-source convex optimization solver Clarabel [GC24]. All code is available at

`https://github.com/tjdiamandis/routing-theory-experiments`.

		low fee ($q_0 = 0.01$)			high fee $q_0 = 1.0$		
n	m	$\mu = 0$	$\mu = 10^{-4}$	$\mu = 10^{-2}$	$\mu = 0$	$\mu = 10^{-4}$	$\mu = 10^{-2}$
10	25	0	0	10	0	0	10
17	72	0	0	16	0	3	16
28	196	0	0	28	0	0	26
46	529	0	0	47	0	4	46
77	1,482	0	0	76	0	4	75
129	4,160	0	3	129	0	17	126
215	11,556	2	11	213	0	56	212
359	32,220	7	45	359	1	150	355
599	89,700	9	536	601	3	479	592
1,000	250,000	33	1,143	1,008	5	884	992

Table 1: Number of non-integral λ_i for the fixed fee problem relaxation.

		low fee ($q_0 = 0.01$)			high fee $q_0 = 1.0$		
n	m	$\mu = 0$	$\mu = 10^{-4}$	$\mu = 10^{-2}$	$\mu = 0$	$\mu = 10^{-4}$	$\mu = 10^{-2}$
10	25	8.20e-11	1.10e-10	9.11e-04	1.78e-09	4.26e-09	9.32e-02
17	72	3.44e-10	7.65e-10	8.83e-04	1.92e-08	6.04e-04	9.37e-02
28	196	1.12e-09	6.43e-10	6.99e-04	7.21e-09	4.77e-09	7.01e-02
46	529	2.28e-10	1.15e-10	7.61e-04	1.23e-08	2.60e-04	8.37e-02
77	1,482	2.77e-10	1.46e-10	8.42e-04	5.85e-09	5.77e-05	9.21e-02
129	4,160	4.27e-10	1.57e-07	9.33e-04	2.66e-09	1.05e-04	1.03e-01
215	11,556	3.10e-08	4.06e-07	8.27e-04	6.73e-09	1.87e-04	9.20e-02
359	32,220	6.70e-08	6.72e-07	8.39e-04	1.16e-06	2.60e-04	9.25e-02
599	89,700	2.47e-08	5.36e-06	8.91e-04	4.22e-07	4.85e-04	9.82e-02
1,000	250,000	3.73e-08	7.05e-06	9.20e-04	7.65e-07	5.48e-04	1.02e-01

Table 2: Relative objective difference between the relaxation and the objective value with the rounded solution.

Results. We solve the fixed fee problem for logarithmically spaced range n ranging from 10 to 1,000 ($m = 25$ to 250,000), for $\mu = 0$ (linear objective), 10^{-4}, and 10^{-2}, and for both a 'low fee' setting with $q_0 = 0.01$ while for the high fee setting, we have $q_0 = 1.0$. We record the number of non-integral λ_i in the solution to the relaxation in table 1 and the relative objective difference between the relaxation and the objective value with the rounded solution, using (19), in table 2. We see that when the objective function is linear, the number of non-integral λ_i is negligible, and the objective value is very close to the optimal value as we would expect. As the objective function becomes more concave, the number of non-integral λ_i increases, but remains small relative to m. Additionally, we see that the objective value of the rounded solution is still very close to that of the relaxation, suggesting that we can approximately solve this NP-hard problem in practice. We leave further numerical investigation to future work.

6 Conclusion

In this work, we have shown a number of theoretical properties of the convex flow problem, all of which follow more or less directly from standard results in convex geometry. We first showed that the traditional assumption of nondecreasing objective terms can be replaced by a downward closedness condition on the sets of allowable flows. Using this condition, we showed that there is a natural calculus of flows. We then introduced the flow cone to derive a conic formulation of the convex flow problem, which we showed is equivalent to the original and has a number of interesting properties. We next examined the case of fixed costs for the use of an edge. Via a Shapley–Folkman argument, we showed that the relaxation of the fixed fee problem always has an 'almost' integral solution. Finally, we showed that this fixed fee problem can be solved using the same algorithm for the convex flow problem, and solving the convex relaxation often results in an integral solution that is close to optimal in practice.

Future work. This work prompts a number of questions, any of which suggest an interesting avenue for future research. First, why do we often find integral solutions to the fixed fee problem? Is there a natural condition that guarantees this? Second, what algorithm is best for solving the fixed fee problem? Should we use the same algorithm as for the convex flow problem, with the modification suggested in §5.5? Finally, this problem has a very natural decomposition over the edges (equivalently, over the nodes, since we are working with a bipartite graph). Can we exploit this decomposition to devise efficient distributed, tatonnement-style algorithms? If so, what convergence rates should we expect? These potentially asynchronous algorithms may be of interest in decentralized applications, such as power grids or wireless networks as discussed in [DAE24, §3].

Acknowledgements

Theo Diamandis is supported by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate (NDSEG) Fellowship Program.

References

[AC20] Guillermo Angeris and Tarun Chitra. "Improved Price Oracles: Constant Function Market Makers". In: *Proceedings of the 2nd ACM Conference on Advances in Financial Technologies*. AFT '20: 2nd ACM Conference on Advances in Financial Technologies. New York NY USA: ACM, Oct. 21, 2020, pp. 80–91. ISBN: 978-1-4503-8139-0. DOI: 10.1145/3419614.3423251. (Visited on 02/17/2021).

[AMO88] Ravindra K Ahuja, Thomas L Magnanti, and James B Orlin. *Network flows*. Cambridge, Mass.: Alfred P. Sloan School of Management, Massachusetts ..., 1988.

[Ang+22] Guillermo Angeris et al. "Optimal routing for constant function market makers". In: *Proceedings of the 23rd ACM Conference on Economics and Computation*. 2022, pp. 115–128.

[Ang+23] Guillermo Angeris et al. "The geometry of constant function market makers". In: *arXiv preprint arXiv:2308.08066* (2023).

[Ber08] Dimitri P Bertsekas. "Extended monotropic programming and duality". In: *Journal of optimization theory and applications* 139.2 (2008), pp. 209–225.

[Ber98] Dimitri Bertsekas. *Network optimization: continuous and discrete models*. Vol. 8. Athena Scientific, 1998.

[Boy+07] Stephen Boyd et al. "Notes on decomposition methods". In: *Notes for EE364B, Stanford University* 635 (2007), pp. 1–36.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. 1st ed. Cambridge, United Kingdom: Cambridge University Press, 2004. 716 pp. ISBN: 978-0-521-83378-3.

[DAE24] Theo Diamandis, Guillermo Angeris, and Alan Edelman. "Convex Network Flows". In: *arXiv preprint arXiv:2404.00765* (2024).

[Dia+23] Theo Diamandis et al. "An Efficient Algorithm for Optimal Routing Through Constant Function Market Makers". In: *arXiv preprint arXiv:2302.04938* (2023).

[GC24] Paul J Goulart and Yuwen Chen. "Clarabel: An interior-point solver for conic programs with quadratic objectives". In: *arXiv preprint arXiv:2405.12762* (2024).

[Kar72] Richard M Karp. *Reducibility among combinatorial problems*. Springer, 1972.

[Roc70] R. Tyrrell Rockafellar. *Convex Analysis*. Vol. 28. Princeton university press, 1970.

[Shi06] Maiko Shigeno. "Maximum network flows with concave gains". In: *Mathematical programming* 107.3 (2006), pp. 439–459.

[Tru78] Klaus Truemper. "Optimal flows in nonlinear gain networks". In: *Networks* 8.1 (1978), pp. 17–36.

[Vég14] László A Végh. "Concave generalized flows with applications to market equilibria". In: *Mathematics of Operations Research* 39.2 (2014), pp. 573–596.

[Wil19] David P Williamson. *Network flow algorithms*. Cambridge University Press, 2019.

A Downward closure of the flow cone

In this section, we show that the flow cone K_i from (5) is downward closed.

Scaling property. An important property of a set of allowable flows is that, for any flow set T_i, if $x \in T_i$, then $\alpha x \in T_i$ for any $0 \le \alpha \le 1$. The proof is nearly immediate by the convexity of T_i and the fact that $0 \in T_i$ by noting that

$$\alpha x = \alpha x + (1-\alpha)0 \in T_i. \tag{21}$$

Downward closure. We will now prove that the flow cone K_i is downward closed. In other words, we will show that, for any $(x, \lambda) \in K_i$ and (x', λ') with $x' \le x$ and $\lambda' \le \lambda$, then $(x', \lambda') \in K_i$. We will first show that $(x, \lambda) \in K_i$ implies that $(x, \lambda') \in K_i$. To see this, note that, if $(x, \lambda) \in K_i$, then there exists a sequence $\{(x_k, \lambda_k)\}$ such that $x_k/(-\lambda_k) \in T_i$ and $\lambda_k < 0$ for each k, while the sequences converge in that $x_k \to x$ and $\lambda_k \to \lambda$ by the definition (5). If $\lambda = \lambda'$ then obviously $(x, \lambda') \in K_i$, so assume that $\lambda' < \lambda$. In this case, there exists some k' such that, $\lambda_k \ge \lambda'$ for all $k \ge k'$. But, since

$$\frac{x_k}{-\lambda_k} \in T_i,$$

for all $k \geq k'$ by definition, then, since $0 \leq \lambda_k/\lambda' \leq 1$ we have that, using the scaling property (21):

$$\frac{\lambda_k}{\lambda'}\frac{x_k}{-\lambda_k} \in T_i,$$

so $x_k/(-\lambda') \in T_i$ for each $k \geq k'$. Since $x_k \to x$ and T_i is closed, then $x/(-\lambda') \in T_i$, so, by definition $(x, \lambda') \in K_i$, as required. The final question, if $x' \leq x$ then $(x', \lambda') \in K_i$ is easy: if $\lambda' < 0$ then $x/(-\lambda') \in T_i$ implies that $x'/(-\lambda') \in T_i$ by the downward closure of T_i, so $(x', \lambda') \in K_i$. If $\lambda' = 0$ then we know that $\lambda = 0$. Set $\delta = x' - x$ and note that $\delta \leq 0$ since $x' \leq x$. This means that, for the sequence $\lambda_k < 0$ with $\lambda_k \to 0$, we have

$$\frac{x_k + \delta}{-\lambda_k} \in T_i,$$

since $x_k/(-\lambda_k) \in T_i$ and $(x_k + \delta)/(-\lambda_k) \leq x_k/(-\lambda_k)$ because $\delta \leq 0$ and T_i is downward closed. But, since $x_k \to x$, then $x_k + \delta \to x + \delta = x'$. By the definition of (5), then we have that $(x + \delta, 0) \in K_i$, so $(x', \lambda') = (x + \delta, 0) \in K_i$ as required.

B Downward closure equivalence using the dual

Here, we provide a brief, intuitive proof of the equivalence between downward closure and monotonicity using the dual problem.

Monotonicity implies downward closure. Assume that the objective functions U and $\{V_i\}$ are nondecreasing. Then the dual variables ν and $\{\eta_i\}$ must all be nonnegative for the dual to be finite-valued. Consider the arbitrage problem for some i:

$$f_i(\eta_i) = \sup_{x_i \in T_i} \eta_i^T x_i.$$

The solution to the arbitrage problem can be broken into two cases. First, if the set T_i contains a positive ray, then the optimal value is unbounded, and the primal problem is therefore infeasible. Second, if the optimal value is finite, then the solution will lie on the boundary of T_i, as T_i is closed and convex. As a result, we can replace T_i with its 'downward extension',

$$\tilde{T}_i = T_i - \mathbf{R}^n_+,$$

without affecting the solution.

Downward closure implies monotonicity. Assume that the sets $\{T_i\}$ are downward closed. We use $x_i^\star$ to denote a solution to the arbitrage problem:

$$x_i^\star = \operatorname*{argmax}_{x \in T_i} \eta_i^T x,$$

for a fixed price vector η_i when this solution is finite. Consider two possible cases for f_i. First, if f_i is unbounded, then the primal problem is again infeasible. Second, if f_i is is finite, then the set T_i is contained in the halfspace $\{x \mid \eta_i^T x \leq f_i(\eta_i)\}$, and $x_i^\star$ is on the boundary of T_i. Since T_i is downward closed, we must have $\eta_i \geq 0$ in this case. As a result, we can replace the objective function with its monotonic concave envelope without affecting the solution.

C Other properties

C.1 Circulation problem

If the net flow utility simply constrains the net flow to be zero, *i.e.*,

$$U(y) = -I_{\{0\}}(y).$$

then we recover a generalized circulation problem. The dual problem becomes

$$\text{minimize} \quad \sum_{i=1}^m \left(\bar{V}_i(\eta_i - A_i^T \nu) + f_i(\eta_i)\right).$$

Taking the infimum of the objective over η and introducing a new variables $z_i \in \mathbf{R}^{n_i}$ for $i = 1, \dots, m$, we can rewrite this problem as

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^m \tilde{V}_i(z_i) \\ \text{subject to} & z_i = A_i^T \nu, \quad i = 1, \dots, m \\ & \nu \geq 0, \end{array}$$

where we define

$$\tilde{V}_i(z_i) = \inf_{\eta_i} \left\{\bar{V}_i(\eta_i - z_i) + f_i(\eta_i)\right\}.$$

(This is very close to, but not quite, the infimal convolution of $\bar{V}_i$ and f_i.) Note that $\tilde{V}_i$ is a convex function, as convexity is preserved under partial minimization [BV04, §3.2.5]. This problem has a nice interpretation: we are finding the pin voltages on the m components in a passive electrical circuit [Boy+07, §6]. The optimality conditions simplify to

$$\begin{aligned} \nabla \tilde{V}_i(z_i) = x_i, &\quad i = 1, \dots, m, \\ A_i^T \nu = z_i, &\quad i = 1, \dots, m, \\ \sum_{i=1}^m A_i x_i \geq 0, &\quad \nu \geq 0. \end{aligned}$$

Viewing z_i and x_i as the voltage and current, respectively, at the terminals of component i, and viewing ν as the voltages at every node in the circuit, then the first equation can can be interpreted as the voltage-current characteristic for component i, the second as the Kirchoff voltage law, and the last as Kirchoff's current law, respectively.

C.2 Cycle condition

When all edges are between two nodes, the optimality conditions have a nice interpretation in terms of cycle conditions, similar to the augmenting path condition for max flow, given by Ford and Fulkerson.

We define the vectors $\delta_1, \dots, \delta_m$ to be an arbitrage with respect to the flows $\{x_i\}$ if the following conditions hold:

1. The vector $\delta_i \in T_i^\star(x_i)$ for all $i = 1, \dots, m$, where

$$T_i^\star(x_i) = \{\delta \mid x_i + t\delta_i \in T_i \text{ for some } t > 0\}.$$

2. For some subgradient $\hat{\nu} \in U\left(\sum_{i=1}^m x_i\right)$, we have that

$$\hat{\nu}^T \left(\sum_{i=1}^m A_i \delta_i\right) \le 0.$$

For the standard max flow problem, this should give the classic augmenting path condition. For more general problems, we can first decompose into a cycle basis to easily check this condition.