RUTCOR
RESEARCH
REPORT

Separable Concave Optimization Approximately Equals Piecewise-Linear Optimization

Thomas L. Magnanti[a] Dan Stratila[b]

RRR 6-2012, January 2012

RUTCOR
Rutgers Center for
Operations Research
Rutgers University
640 Bartholomew Road
Piscataway, New Jersey
08854-8003
Telephone: 732-445-3804
Telefax: 732-445-5472
Email: rrr@rutcor.rutgers.edu
http://rutcor.rutgers.edu/∼rrr

[a]School of Engineering and Sloan School of Management, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room 32-D784, Cambridge, MA 02139. E-mail: magnanti@mit.edu.

[b]Rutgers Center for Operations Research and Rutgers Business School, Rutgers University, 640 Bartholomew Road, Room 107, Piscataway, NJ 08854. E-mail: dstrat@rci.rutgers.edu.

Separable Concave Optimization Approximately Equals Piecewise-Linear Optimization*

Thomas L. Magnanti† Dan Stratila‡

January 13, 2012

Abstract

We study the problem of minimizing a nonnegative separable concave function over a compact feasible set. We approximate this problem to within a factor of $1+\epsilon$ by a piecewise-linear minimization problem over the same feasible set. Our main result is that when the feasible set is a polyhedron, the number of resulting pieces is polynomial in the input size of the polyhedron and linear in $1/\epsilon$. For many practical concave cost problems, the resulting piecewise-linear cost problem can be formulated as a well-studied discrete optimization problem. As a result, a variety of polynomial-time exact algorithms, approximation algorithms, and polynomial-time heuristics for discrete optimization problems immediately yield fully polynomial-time approximation schemes, approximation algorithms, and polynomial-time heuristics for the corresponding concave cost problems.

We illustrate our approach on two problems. For the concave cost multicommodity flow problem, we devise a new heuristic and study its performance using computational experiments. We are able to approximately solve significantly larger test instances than previously possible, and obtain solutions on average within 4.27% of optimality. For the concave cost facility location problem, we obtain a new $1.4991+\epsilon$ approximation algorithm.

1 Introduction

Minimizing a nonnegative separable concave function over a polyhedron arises frequently in fields such as transportation, logistics, telecommunications, and supply chain management. In a typical application, the polyhedral feasible set arises due to network structure, capacity requirements, and other constraints, while the concave costs arise due to economies of scale, volume discounts, and other practical factors [see e.g. GP90]. The concave functions can be nonlinear, piecewise-linear with many pieces, or more generally given by an oracle.

*This research is based on the second author's Ph.D. thesis at the Massachusetts Institute of Technology [Str08]. An extended abstract of this research has appeared in [MS04].

†School of Engineering and Sloan School of Management, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room 32-D784, Cambridge, MA 02139. E-mail: magnanti@mit.edu.

‡Rutgers Center for Operations Research and Rutgers Business School, Rutgers University, 640 Bartholomew Road, Room 107, Piscataway, NJ 08854. E-mail: dstrat@rci.rutgers.edu.

A natural approach for solving such a problem is to approximate each concave function by a piecewise-linear function, and then reformulate the resulting problem as a discrete optimization problem. Often this transformation can be carried out in a way that preserves problem structure, making it possible to apply existing discrete optimization techniques to the resulting problem. A wide variety of techniques is available for these problems, including heuristics [e.g. BMW89, HH98], integer programming methods [e.g. Ata01, OW03], and approximation algorithms [e.g. JMM$^+$03].

For this approach to be efficient, we need to be able to approximate the concave cost problem by a single piecewise-linear cost problem that meets two competing requirements. On one hand, the approximation should employ few pieces so that the resulting problem will have small input size. On the other hand, the approximation should be precise enough that by solving the resulting problem we would obtain an acceptable approximate solution to the original problem.

With this purpose in mind, we introduce a method for approximating a concave cost problem by a piecewise-linear cost problem that provides a $1 + \epsilon$ approximation in terms of optimal cost, and yields a bound on the number of resulting pieces that is polynomial in the input size of the feasible polyhedron and linear in $1/\epsilon$. Previously, no such polynomial bounds were known, even if we allow any dependence on $1/\epsilon$.

Our bound implies that polynomial-time exact algorithms, approximation algorithms, and polynomial-time heuristics for many discrete optimization problems immediately yield fully polynomial-time approximation schemes, approximation algorithms, and polynomial-time heuristics for the corresponding concave cost problems. We illustrate this result by obtaining a new heuristic for the concave cost multicommodity flow problem, and a new approximation algorithm for the concave cost facility location problem.

Under suitable technical assumptions, our method can be generalized to efficiently approximate the objective function of a maximization or minimization problem over a general feasible set, as long as the objective is nonnegative, separable, and concave. In fact, our approach is not limited to optimization problems. It is potentially applicable for approximating problems in dynamic programming, algorithmic game theory, and other settings where new solutions methods become available when switching from concave to piecewise-linear functions.

1.1 Previous Work

Piecewise-linear approximations are used in a variety of contexts in science and engineering, and the literature on them is expansive. Here we limit ourselves to previous results on approximating a separable concave function in the context of an optimization problem.

Geoffrion [Geo77] obtains several general results on approximating objective functions. One of the settings he considers is the minimization of a separable concave function over a general feasible set. He derives conditions under which a piecewise-linear approximation of the objective achieves the smallest possible absolute error for a given number of pieces.

Thakur [Tha78] considers the maximization of a separable concave function over a convex set defined by separable constraints. He approximates both the objective and constraint functions, and bounds the absolute error when using a given number of pieces in terms of feasible set parameters, the maximum values of the first and second derivatives of the functions, and certain dual optimal solutions.

Rosen and Pardalos [RP86] consider the minimization of a quadratic concave function over a polyhedron. They reduce the problem to a separable one, and then approximate the resulting univariate concave functions. The authors derive a bound on the number of pieces needed to guarantee a given approximation error in terms of objective function and feasible polyhedron parameters. They use a nonstandard definition of approximation error, dividing by a scale factor that is at least the maximum minus the minimum of the concave function over the feasible polyhedron. Pardalos and Kovoor [PK90] specialize this result to the minimization of a quadratic concave function over one linear constraint subject to upper and lower bounds on the variables.

Güder and Morris [GM94] study the maximization of a separable quadratic concave function over a polyhedron. They approximate the objective functions, and bound the number of pieces needed to guarantee a given absolute error in terms of function parameters and the lengths of the intervals on which the functions are approximated.

Kontogiorgis [Kon00] also studies the maximization of a separable concave function over a polyhedron. He approximates the objective functions, and uses techniques from numerical analysis to bound the absolute error when using a given number of pieces in terms of the maximum values of the second derivatives of the functions and the lengths of the intervals on which the functions are approximated.

Each of these prior results differs from ours in that, when the goal is to obtain a $1+\epsilon$ approximation, they do not provide a bound on the number of pieces that is polynomial in the input size of the original problem, even if we allow any dependence on $1/\epsilon$.

Meyerson et al. [MMP00] remark, in the context of the single-sink concave cost multicommodity flow problem, that a "tight" approximation could be computed. Munagala [Mun03] states, in the same context, that an approximation of arbitrary precision could be obtained with a polynomial number of pieces. They do not mention specific bounds, or any details on how to do so.

Hajiaghayi et al. [HMM03] and Mahdian et al. [MYZ06] consider the unit demand concave cost facility location problem, and employ an exact reduction by interpolating the concave functions at points $1, 2, \ldots, m$, where m is the number of customers. The size of the resulting problem is polynomial in the size of the original problem, but the approach is limited to problems with unit demand.

1.2 Our Results

In Section 2, we introduce our piecewise-linear approximation approach, on the basis of a minimization problem with a compact feasible set in $\mathbb{R}^n_+$ and a nonnegative separable concave function that is nondecreasing. In this section, we assume that the problem has an optimal solution $x^* = (x^*_1, \ldots, x^*_n)$ with $x^*_i \in \{0\} \cup [l_i, u_i]$ and $0 < l_i \le u_i$. To obtain a $1+\epsilon$ approximation, we need only $1 + \lceil \log_{1+4\epsilon+4\epsilon^2} \frac{u_i}{l_i} \rceil$ pieces for each concave function. As $\epsilon \to 0$, the number of pieces behaves as $\frac{1}{4\epsilon} \log \frac{u_i}{l_i}$.

In Section 2.1, we show that any piecewise-linear approach requires at least $\Omega\left(\frac{1}{\sqrt{\epsilon}} \log \frac{u_i}{l_i}\right)$ pieces to approximate a certain function to within $1+\epsilon$ on $[l_i, u_i]$. Note that for any fixed ϵ, the number of pieces required by our approach is within a constant factor of this lower bound. It is an interesting open question to find tighter upper and lower bounds on the number of pieces as $\epsilon \to 0$. In Section 2.2, we extend our approximation approach to objective functions that are not monotone and feasible sets that are not contained in $\mathbb{R}^n_+$.

In Sections 3 and 3.1, we obtain the main result of this paper. When the feasible set is a polyhedron and the cost function is nonnegative separable concave, we can obtain a $1 + \epsilon$ approximation with a number of pieces that is polynomial in the input size of the feasible polyhedron and linear in $1/\epsilon$. We first obtain a result for polyhedra in $\mathbb{R}^n_+$ and nondecreasing cost functions in Section 3, and then derive the general result in Section 3.1.

In Section 4, we show how our piecewise-linear approximation approach can be combined with algorithms for discrete optimization problems to obtain new algorithms for problems with concave costs. We use a well-known integer programming formulation that often enables us to write piecewise-linear problems as discrete optimization problems in a way that preserves problem structure.

In Section 5, we illustrate our method on the concave cost multicommodity flow problem. We derive considerably smaller bounds on the number of required pieces than in the general case. Using the formulation from Section 4, the resulting discrete optimization problem can be written as a fixed charge multicommodity flow problem. This enables us to devise a new heuristic for concave cost multicommodity flow by combining our piecewise-linear approximation approach with a dual ascent method for fixed charge multicommodity flow due to Balakrishnan et al. [BMW89].

In Section 5.1, we report on computational experiments. The new heuristic is able to solve large-scale test problems to within 4.27% of optimality, on average. The concave cost problems have up to 80 nodes, 1,580 edges, 6,320 commodities, and 9,985,600 variables. These problems are, to the best of our knowledge, significantly larger than previously solved concave cost multicommodity flow problems, whether approximately or exactly. A brief review of the literature on concave cost flows can be found in Sections 5 and 5.1.

In Section 6, we illustrate our method on the concave cost facility location problem. Combining a 1.4991-approximation algorithm for the classical facility location problem due to Byrka [Byr07] with our approach, we obtain a $1.4991 + \epsilon$ approximation algorithm for concave cost facility location. Previously, the lowest approximation ratio for this problem was that of a $3 + \epsilon$ approximation algorithm due to Mahdian and Pal [MP03]. In the second author's Ph.D. thesis [Str08], we obtain a number of other algorithms for concave cost problems, including a 1.61-approximation algorithm for concave cost facility location. Independently, Romeijn et al. [RSSZ10] developed 1.61 and 1.52-approximation algorithms for this problem. A brief review of the literature on concave cost facility location can be found in Section 6.

2 General Feasible Sets

We examine the general concave cost minimization problem

$$Z_1^* = \min \{\phi(x) : x \in X\}, \tag{1}$$

defined by a compact feasible set $X \subseteq \mathbb{R}^n_+$ and a nondecreasing separable concave function $\phi : \mathbb{R}^n_+ \to \mathbb{R}_+$. Let $x = (x_1, \ldots, x_n)$ and $\phi(x) = \sum_{i=1}^n \phi_i(x_i)$, and assume that the functions ϕ_i are nonnegative. The feasible set need not be convex or connected—for example, it could be the feasible set of an integer program.

In this section, we impose the following technical assumption. Let $[n] = \{1, \dots, n\}$.

Assumption 1. The problem has an optimal solution $x^* = (x_1^*, \dots, x_n^*)$ and bounds l_i and u_i with $0 < l_i \le u_i$ such that $x_i^* \in \{0\} \cup [l_i, u_i]$ for $i \in [n]$.

Let $\epsilon > 0$. To approximate problem (1) to within a factor of $1+\epsilon$, we approximate each function ϕ_i by a piecewise-linear function $\psi_i : \mathbb{R}_+ \to \mathbb{R}_+$. Each function ψ_i consists of $P_i + 1$ pieces, with $P_i = \lceil \log_{1+\epsilon} \frac{u_i}{l_i} \rceil$, and is defined by the coefficients

$$s_i^p = \phi_i' \left(l_i(1+\epsilon)^p \right), \qquad p \in \{0, \dots, P_i\}, \tag{2a}$$

$$f_i^p = \phi_i \left(l_i(1+\epsilon)^p \right) - l_i(1+\epsilon)^p s_i^p, \qquad p \in \{0, \dots, P_i\}. \tag{2b}$$

If the derivative $\phi_i' \left(l_i(1+\epsilon)^p \right)$ does not exist, we take the right derivative, that is $s_i^p = \lim_{x_i \to (l_i(1+\epsilon)^p)^+} \frac{\phi_i(l_i(1+\epsilon)^p) - \phi_i(x_i)}{l_i(1+\epsilon)^p - x_i}$. The right derivative always exists at points in $(0, +\infty)$ since ϕ_i is concave on $[0, +\infty)$. We proceed in this way throughout the paper when the derivative does not exist.

Each coefficient pair (s_i^p, f_i^p) defines a line with nonnegative slope s_i^p and y-intercept f_i^p, which is tangent to the graph of ϕ_i at the point $l_i(1+\epsilon)^p$. For $x_i > 0$, the function ψ_i is defined by the lower envelope of these lines:

$$\psi_i(x_i) = \min\{ f_i^p + s_i^p x_i : p = 0, \dots, P_i \}. \tag{3}$$

We let $\psi_i(0) = \phi_i(0)$ and $\psi(x) = \sum_{i=1}^n \psi_i(x_i)$. Substituting ψ for ϕ in problem (1), we obtain the piecewise-linear cost problem

$$Z_4^* = \min\{ \psi(x) : x \in X \}. \tag{4}$$

Next, we prove that this problem provides a $1+\epsilon$ approximation for problem (1). The following proof has an intuitive geometric interpretation, but does not yield a tight analysis of the approximation guarantee. A tight analysis will follow.

Lemma 1. $Z_1^* \le Z_4^* \le (1+\epsilon) Z_1^*$.

Proof. Let $x^* = (x_1^*, \dots, x_n^*)$ be an optimal solution to problem (4); an optimal solution exists since $\psi(x)$ is concave and X is compact. Fix an $i \in [n]$, and note that for each $p \in \{0, \dots, P_i\}$, the line $f_i^p + s_i^p x_i$ is tangent from above to the graph of $\phi_i(x_i)$. Hence $\phi_i(x_i^*) \le \min\{ f_i^p + s_i^p x_i^* : p = 0, \dots, P_i\} = \psi_i(x_i^*)$. Therefore, $Z_1^* \le \phi(x^*) \le \psi(x^*) = Z_4^*$.

Conversely, let x^* be an optimal solution of problem (1) that satisfies Assumption 1. It suffices to show that $\psi_i(x_i^*) \le (1+\epsilon)\phi_i(x_i^*)$ for $i \in [n]$. If $x_i^* = 0$, then the inequality holds. Otherwise, let $p = \left\lfloor \log_{1+\epsilon} \frac{x_i^*}{l_i} \right\rfloor$, and note that $p \in \{0, \dots, P_i\}$ and $\frac{x_i^*}{l_i} \in [(1+\epsilon)^p, (1+\epsilon)^{p+1})$. Because ϕ_i is concave, nonnegative, and nondecreasing,

$$\begin{aligned} \psi_i(x_i^*) &\le f_i^p + s_i^p x_i^* \le f_i^p + s_i^p l_i (1+\epsilon)^{p+1} \\ &= f_i^p + s_i^p l_i (1+\epsilon)(1+\epsilon)^p \le (1+\epsilon) \left(f_i^p + s_i^p l_i (1+\epsilon)^p \right) \\ &= (1+\epsilon) \phi_i \left(l_i (1+\epsilon)^p \right) \le (1+\epsilon) \phi_i(x_i^*). \end{aligned} \tag{5}$$

(See Figure 1 for an illustration.) Therefore, $Z_4^* \le \psi(x^*) \le (1+\epsilon)\phi(x^*) = (1+\epsilon) Z_1^*$. □

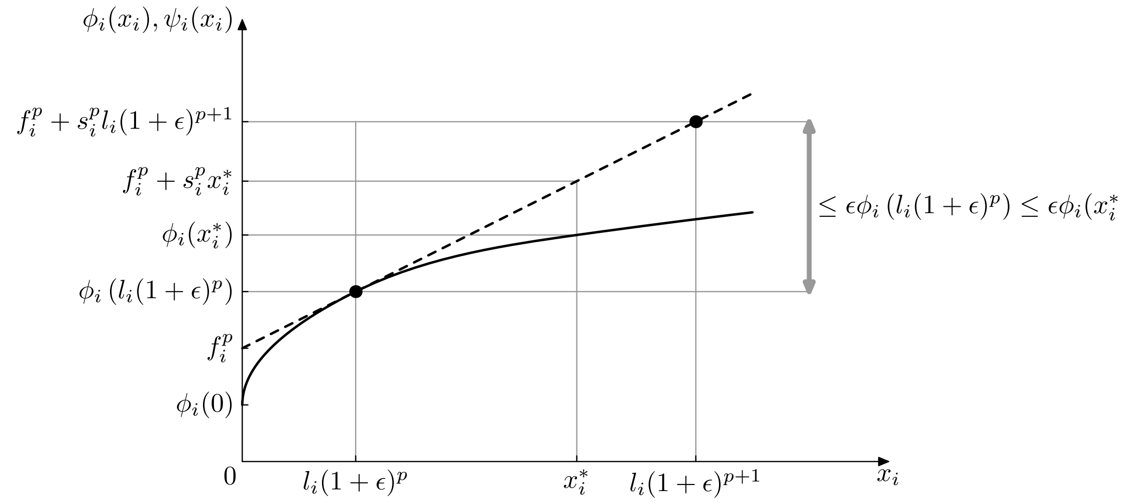

Figure 1: Illustration of the proof of Lemma 1. Observe that the height of any point inside the box with the bold lower left and upper right corners exceeds the height of the box's lower left corner by at most a factor of ϵ.

We now present a tight analysis.

Theorem 1. $Z_1^* \leq Z_4^* \leq \frac{1+\sqrt{\epsilon+1}}{2} Z_1^*$. *The approximation guarantee of* $\frac{1+\sqrt{\epsilon+1}}{2}$ *is tight.*

Proof. We have shown that $Z_1^* \leq Z_4^*$ in Lemma 1. Fix an $i \in [n]$, and consider the approximation ratio achieved on $[l_i, u_i]$ when approximating ϕ_i by ψ_i. If $\phi_i(l_i) = 0$, then $\phi_i(x_i) = 0$ for all $x_i \geq 0$, and we have a trivial case. If $\phi_i(l_i) > 0$, then $\phi_i(x_i) > 0$ for all $x_i > 0$, and the ratio is

$$\min\{1+\gamma : \psi_i(x_i) \leq (1+\gamma)\phi_i(x_i) \text{ for } x_i \in [l_i, u_i]\} \\ = \max\{\psi_i(x_i)/\phi_i(x_i) : x_i \in [l_i, u_i]\}. \tag{6}$$

We derive an upper bound of $\frac{1+\sqrt{\epsilon+1}}{2}$ on this ratio, and then construct a family of functions that, when taken as ϕ_i, yield ratios converging to this upper bound.

Without loss of generality, we assume $l_i = 1$ and $u_i = (1+\epsilon)^{P_i}$. The approximation ratio achieved on $[1, u_i]$ is the highest of the approximation ratios on each of the intervals $[1, 1+\epsilon], \ldots, [(1+\epsilon)^{P_i-1}, (1+\epsilon)^{P_i}]$. By scaling along the x-axis, it is enough to consider only the interval $[1, 1+\epsilon]$, and therefore we can assume that ψ_i is given by the two tangents to the graph of ϕ_i at 1 and $1+\epsilon$. Suppose these tangents have slopes a and c respectively. We can assume that $\phi_i(0) = 0$, and that ϕ_i is linear with slope a on $[0, 1]$ and linear with slope c on $[1+\epsilon, +\infty)$. By scaling along the y-axis, we can assume that $a = 1$.

We upper bound the approximation ratio between ψ_i and ϕ_i by the ratio between ψ_i and a new function φ_i that has $\varphi_i(0) = 0$ and consists of 3 linear pieces with slopes $1 \geq b \geq c$ on $[0, 1]$, $[1, 1+\epsilon]$, and $[1+\epsilon, +\infty)$ respectively. The approximation ratio between ψ_i and φ_i can be viewed as a function of b and c. Let $1+\xi$ be a point on the interval $[1, 1+\epsilon]$.

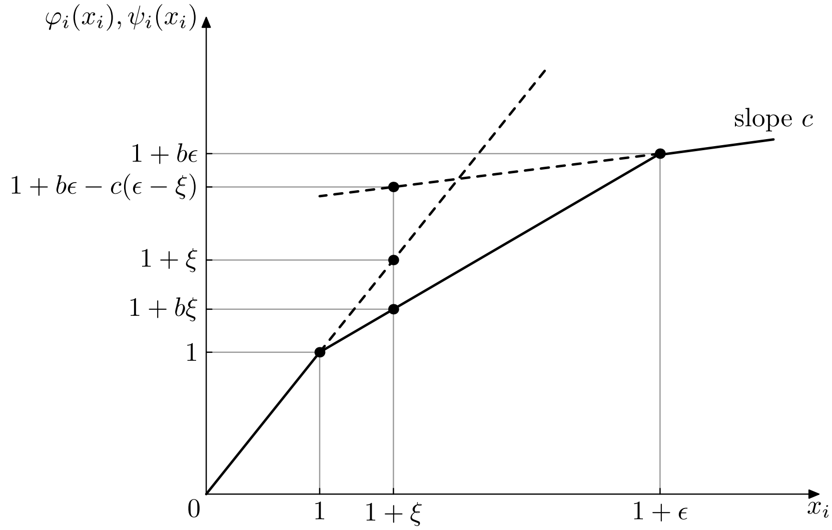

Figure 2: Illustration of the proof of Theorem 1.

We are interested in the following maximization problem with respect to b, c, and ξ:

$$\max\{\psi_i(1+\xi)/\varphi_i(1+\xi) : 1 \geq b \geq c \geq 0, 0 \leq \xi \leq \epsilon\}. \tag{7}$$

Since φ_i consists of 3 linear pieces, while ψ_i is given by the lower envelope of two tangents, we have

$$\varphi_i(1+\xi) = 1+b\xi, \qquad \psi_i(x_i) = \min\left\{1+\xi, 1+b\epsilon - c(\epsilon-\xi)\right\}. \tag{8}$$

(See Figure 2 for an illustration.)

Since $\xi \leq \epsilon$, we can assume that $c = 0$. Next, since we seek to find b and ξ that maximize

$$\frac{\psi_i(1+\xi)}{\varphi_i(1+\xi)} = \frac{\min\{1+\xi, 1+b\epsilon\}}{1+b\xi} = \min\left\{\frac{1+\xi}{1+b\xi}, \frac{1+b\epsilon}{1+b\xi}\right\}, \tag{9}$$

we can assume ξ is such that $\frac{1+\xi}{1+b\xi} = \frac{1+b\epsilon}{1+b\xi}$, which yields $\xi = \epsilon b$. Substituting, we now seek to maximize $\frac{1+\epsilon b}{1+\epsilon b^2}$, and we find that the maximum is achieved at $b = \frac{1}{1+\sqrt{\epsilon+1}}$ and equals $\frac{1+\sqrt{\epsilon+1}}{2}$. Therefore, $\frac{1+\sqrt{\epsilon+1}}{2}$ is an approximation guarantee for our approach.

Finally, we show that this guarantee is tight. First, let ϕ_i be the new function φ_i with b and c taken to have the values that yield the guarantee above. If we were to use our approach to approximate ϕ_i, the tangents at 1 and $1+\epsilon$ would have slopes b and c, instead of the desired 1 and c, since ϕ_i lacks derivatives at 1 and $1+\epsilon$, and our approach uses the derivative from the right when the derivative does not exist. We can cause our approach to produce tangents at 1 and $1+\epsilon$ with slopes 1 and c by taking a sufficiently small ζ and letting ϕ_i have its first breakpoint at $1+\zeta$ instead of 1. When $\zeta \to 0$, the approximation ratio achieved by our approach for ϕ_i converges to the guarantee above. □

To compare the tight approximation guarantee with that provided by Lemma 1, note that $\frac{1+\sqrt{\epsilon+1}}{2} \leq 1 + \frac{\epsilon}{4}$ for $\epsilon > 0$. Moreover, since $\frac{1+\sqrt{\epsilon+1}}{2} \to 1$ and $\frac{d}{d\epsilon}\frac{1+\sqrt{\epsilon+1}}{2} \to \frac{1}{4}$ as $\epsilon \to 0$, it follows that $1+\frac{\epsilon}{4}$ is the lowest ratio of the form $1+\frac{\epsilon}{k}$ that is guaranteed by our approach.

Equivalently, instead of an approximation guarantee of $\frac{1+\sqrt{\epsilon+1}}{2}$ using $1 + \left\lceil \log_{1+\epsilon} \frac{u_i}{l_i} \right\rceil$ pieces, we can obtain a guarantee of $1+\epsilon$ using only $1 + \left\lceil \log_{1+4\epsilon+4\epsilon^2} \frac{u_i}{l_i} \right\rceil$ pieces. Note that $\log_{1+4\epsilon+4\epsilon^2} \frac{u_i}{l_i} = \frac{1}{\log(1+4\epsilon+4\epsilon^2)} \log \frac{u_i}{l_i}$, and as $\epsilon \to 0$, we have $\frac{1}{\log(1+4\epsilon+4\epsilon^2)} \to +\infty$ and $\frac{4\epsilon}{\log(1+4\epsilon+4\epsilon^2)} \to 1$. Therefore, as $\epsilon \to 0$, the number of pieces behaves as $\frac{1}{4\epsilon} \log \frac{u_i}{l_i}$.

This bound on the number of pieces enables us to apply our approximation approach to practical concave cost problems. In Section 3, we will exploit the logarithmic dependence on $\frac{u_i}{l_i}$ of this bound to derive polynomial bounds on the number of pieces for problems with polyhedral feasible sets.

2.1 A Lower Bound on the Number of Pieces

Since the approximation guarantee of Theorem 1 is tight, $1 + \left\lceil \log_{1+4\epsilon+4\epsilon^2} \frac{u_i}{l_i} \right\rceil$ is a lower bound on the number of pieces needed to guarantee a $1+\epsilon$ approximation when using the approach of equations (2)–(4). In this section, we establish a lower bound on the number of pieces needed to guarantee a $1+\epsilon$ approximation when using *any* piecewise-linear approximation approach.

First, we show that by limiting ourselves to approaches that use piecewise-linear functions whose pieces are tangent to the graph of the original function, we increase the number of needed pieces by at most a factor of 3.

Let $\phi_i : \mathbb{R}_+ \to \mathbb{R}_+$ be a nondecreasing concave function, which we are interested in approximating on an interval $[l_i, u_i]$ with $0 < l_i \leq u_i$. Assume that $\phi_i(x_i) > 0$ for all $x_i > 0$; if $\phi_i(x_i) = 0$ for some $x_i > 0$, then ϕ_i must be zero everywhere on $[0, +\infty)$, and we have a trivial case. Also let $\psi_i : \mathbb{R}_+ \to \mathbb{R}_+$ be a piecewise-linear function with Q pieces that approximates ϕ_i on $[l_i, u_i]$ to within a factor of $1+\epsilon$, that is $\frac{1}{1+\epsilon} \leq \frac{\psi_i(x_i)}{\phi_i(x_i)} \leq 1+\epsilon$ for $x_i \in [l_i, u_i]$. We are not imposing any other assumptions on ψ_i; in particular it need not be continuous, and its pieces need not be tangent to the graph of ϕ_i.

Lemma 2. *The function ϕ_i can be approximated on $[l_i, u_i]$ to within a factor of $1+\epsilon$ by a piecewise-linear function $\varphi_i : \mathbb{R}_+ \to \mathbb{R}_+$ that has at most $3Q$ pieces and whose every piece is tangent to the graph of ϕ_i.*

Proof. First, we translate the pieces of ψ_i that are strictly above ϕ_i down, and the pieces strictly below up, until they intersect ϕ_i. Let the modified function be ψ_i'; clearly ψ_i' still provides a $1+\epsilon$ approximation for ϕ_i on $[l_i, u_i]$.

For each piece of ψ_i', we proceed as follows. Let $f_i^p + s_i^p x_i$ be the line defining the piece, and $[a_p, b_p]$ be the interval covered by the piece on the x-axis. Without loss of generality, assume that $[a_p, b_p] \subseteq [l_i, u_i]$. If the piece is tangent to ϕ_i, we take it as one of the pieces composing φ_i, ensuring that φ_i provides a $1+\epsilon$ approximation for ϕ_i on $[a_p, b_p]$.

If the piece is not tangent, it must intersect ϕ_i at either one or two points. If the piece intersects at two points ξ_1 and ξ_2, then the points partition the interval $[a_p, b_p]$ into three subintervals: $[a_p, \xi_1]$, on which the piece is above ϕ_i; $[\xi_1, \xi_2]$, on which the piece is below ϕ_i; and $[\xi_2, b_p]$, on which the piece is again above ϕ_i. If there is one intersection point, we can partition $[a_p, b_p]$ similarly, except that one or two of the subintervals would be empty.

On the interval $[a_p, \xi_1]$, the line $f_i^p + s_i^p x_i$ is above ϕ_i and provides a $1+\epsilon$ approximation for ϕ_i. We take the tangent to ϕ_i at ξ_1 as one of the pieces composing φ_i, ensuring that φ_i

provides a $1+\epsilon$ approximation for ϕ_i on $[a_p, \xi_1]$. Similarly, we take the tangent to ϕ_i at ξ_2 as one of the pieces, ensuring a $1+\epsilon$ approximation on $[\xi_2, b_p]$.

Next, note that on the interval $[\xi_1, \xi_2]$, the line $f_i^p + s_i^p x_i$ is below ϕ_i and provides a $1+\epsilon$ approximation for ϕ_i. Therefore, on $[\xi_1, \xi_2]$, the scaled line $(1+\epsilon)f_i^p + (1+\epsilon)s_i^p x_i$ is above ϕ_i and still provides a $1+\epsilon$ approximation for ϕ_i. Since the original line is below and the scaled line above ϕ_i, there is an ϵ^* with $0 < \epsilon^* \leq \epsilon$ such that, on $[\xi_1, \xi_2]$, the line $(1+\epsilon^*)f_i^p + (1+\epsilon^*)s_i^p x_i$ is above ϕ_i and intersects it at one or more points. If this line is tangent, we take it as one of the pieces that define φ_i. If the line is not tangent, it must intersect ϕ_i at exactly one point ξ', and we take the tangent to ϕ_i at ξ' as one of the pieces. In either case, we have ensured that φ_i provides a $1+\epsilon$ approximation for ϕ_i on $[\xi_1, \xi_2]$.

Since $\cup_{p=1}^{Q}[a_p, b_p] = [l_i, u_i]$, the constructed function φ_i provides a $1+\epsilon$ approximation for ϕ_i on $[l_i, u_i]$. Since for each piece of ψ_i, we introduced at most 3 pieces, φ_i has at most $3Q$ pieces. □

Next, we establish a lower bound on the number of pieces needed to approximate the square root function to within a factor of $1+\epsilon$ by a piecewise-linear function that has its every piece tangent to the graph of the original function. Let $\phi_i(x_i) = \sqrt{x_i}$, and let φ_i be a piecewise-linear function that approximates ϕ_i to within a factor of $1+\epsilon$ on $[l_i, u_i]$ and whose every piece is tangent to the graph of ϕ_i.

To write the lower bounds in this section in a more intuitive way, we define the function $\gamma(\epsilon) = \left(1 + 2\epsilon(2+\epsilon) + 2(1+\epsilon)\sqrt{\epsilon(2+\epsilon)}\right)^2$. As $\epsilon \to 0$, $\gamma(\epsilon)$ behaves as $1 + \sqrt{32\epsilon}$, with the other terms vanishing because they contain higher powers of ϵ. In particular, $1 + \sqrt{32\epsilon} \leq \gamma(\epsilon) \leq 1 + 16\sqrt{\epsilon}$ for $0 < \epsilon \leq \frac{1}{10}$.

Lemma 3. *The function φ_i must contain at least $\left\lceil \log_{\gamma(\epsilon)} \frac{u_i}{l_i} \right\rceil$ pieces. As $\epsilon \to 0$, this lower bound behaves as $\frac{1}{\sqrt{32\epsilon}} \log \frac{u_i}{l_i}$.*

Proof. Given a point $\xi_0 \in [l_i, u_i]$, a tangent to ϕ_i at ξ_0 guarantees a $1+\epsilon$ approximation on an interval extending to the left and right of ξ_0. Let us denote this interval by $[\xi_0(1+\delta_1), \xi_0(1+\delta_2)]$. The values of δ_1 and δ_2 can be found by solving with respect to δ the equation

$$\phi_i(\xi_0) + \delta\xi_0\phi_i'(\xi_0) = (1+\epsilon)\phi_i((1+\delta)\xi_0) \tag{10a}$$

$$\Leftrightarrow \sqrt{\xi_0} + \delta\xi_0\frac{1}{2\sqrt{\xi_0}} = (1+\epsilon)\sqrt{(1+\delta)\xi_0} \tag{10b}$$

$$\Leftrightarrow \xi_0 + \delta\xi_0 + \frac{1}{4}\delta^2\xi_0 = (1+\epsilon)^2(1+\delta)\xi_0. \tag{10c}$$

This is simply a quadratic equation with respect to δ, and solving it yields $\delta_1 = 2\epsilon(2+\epsilon) - 2(1+\epsilon)\sqrt{\epsilon(2+\epsilon)}$ and $\delta_2 = 2\epsilon(2+\epsilon) + 2(1+\epsilon)\sqrt{\epsilon(2+\epsilon)}$. Let $\xi_1 = \xi_0(1+\delta_1)$, and note that $[\xi_0(1+\delta_1), \xi_0(1+\delta_2)] = \left[\xi_1, \frac{1+\delta_2}{1+\delta_1}\xi_1\right]$. Therefore, the tangent provides a $1+\epsilon$ approximation on the interval

$$\left[\xi_1, \frac{1+\delta_2}{1+\delta_1}\xi_1\right] = \left[\xi_1, \left(1 + 2\epsilon(2+\epsilon) + 2(1+\epsilon)\sqrt{\epsilon(2+\epsilon)}\right)^2 \xi_1\right] = [\xi_1, \gamma(\epsilon)\xi_1]. \tag{11}$$

Since $\gamma(\epsilon)$ does not depend on ξ_1, the best way to obtain a $1+\epsilon$ approximation on $[l_i, u_i]$ is to iteratively introduce tangents that provide approximations on intervals of the form

$[l_i, \gamma(\epsilon) l_i]$, $[\gamma(\epsilon) l_i, \gamma^2(\epsilon) l_i], [\gamma^2(\epsilon) l_i, \gamma^3(\epsilon) l_i], \ldots$, until the entire interval $[l_i, u_i]$ is covered. It immediately follows that we need at least $\lceil \log_{\gamma(\epsilon)} \frac{u_i}{l_i} \rceil$ pieces to approximate ϕ_i on $[l_i, u_i]$.

This bound can also be written as $\lceil \frac{1}{\log \gamma(\epsilon)} \log \frac{u_i}{l_i} \rceil$. As $\epsilon \to 0$, we have $\frac{1}{\log \gamma(\epsilon)} \to +\infty$ and $\frac{\sqrt{32\epsilon}}{\log \gamma(\epsilon)} \to 1$, and therefore, the lower bound behaves as $\frac{1}{\sqrt{32\epsilon}} \log \frac{u_i}{l_i}$. □

Combining Lemmas 2 and 3, we immediately obtain a lower bound for any piecewise-linear approximation approach. Let $\psi_i : \mathbb{R}_+ \to \mathbb{R}_+$ be a piecewise-linear function that approximates $\phi_i(x_i) = \sqrt{x_i}$ to within a factor of $1 + \epsilon$ on $[l_i, u_i]$. Note that ψ_i need not be continuous or have its pieces tangent to the graph of ϕ_i.

Theorem 2. *The function ψ_i must contain at least $\lceil \frac{1}{3} \log_{\gamma(\epsilon)} \frac{u_i}{l_i} \rceil$ pieces. As $\epsilon \to 0$, this lower bound behaves as $\frac{1}{\sqrt{288\epsilon}} \log \frac{u_i}{l_i}$.*

This lower bound is within a factor of $2 + \frac{3 \log \gamma(\epsilon)}{\log(1+4\epsilon+4\epsilon^2)}$ of the number of pieces required by our approach. This implies that for fixed ϵ, the number of pieces required by our approach is within a constant factor of the best possible. As $\epsilon \to 0$, the number of pieces needed by our approach converges to a factor of $\frac{\sqrt{288\epsilon}}{4\epsilon} = O\big(\frac{1}{\sqrt{\epsilon}}\big)$ of the lower bound. An interesting open question is to find tighter upper and lower bounds on the number of pieces as $\epsilon \to 0$.

2.2 Extensions

Our approximation approach applies to a broader class of problems. In this section, we generalize our results to objective functions that are not monotone and feasible sets that are not contained in $\mathbb{R}^n_+$. Consider the problem

$$Z^*_{12} = \min\{\phi(x) : x \in X\}, \tag{12}$$

defined by a compact feasible set $X \subseteq \mathbb{R}^n$ and a separable concave function $\phi : Y \to \mathbb{R}_+$. The feasible set X need not be convex or connected, and the set Y can be any convex set in $\mathbb{R}^n$ that contains X. Let $\phi(x) = \sum_{i=1}^n \phi_i(x_i)$, and assume that the functions ϕ_i are nonnegative.

Instead of Assumption 1, we impose the following assumption. Let $\operatorname{proj}_{x_i} Y$ denote the projection of Y on x_i, and note that $\operatorname{proj}_{x_i} Y$ is the domain of ϕ_i.

Assumption 2. Problem (12) has an optimal solution $x^* = (x^*_1, \ldots, x^*_n)$, bounds α_i, β_i with $[\alpha_i, \beta_i] \subseteq \operatorname{proj}_{x_i} Y$, and bounds l_i, u_i with $0 < l_i \le u_i$ such that $x^*_i \in \{\alpha_i, \beta_i\} \cup \big([\alpha_i + l_i, \alpha_i + u_i] \cap [\beta_i - u_i, \beta_i - l_i]\big)$ for $i \in [n]$.

Next, we apply the approach of equations (2)–(4) to approximate problem (12) to within a factor of $1 + \epsilon$. We approximate each concave function ϕ_i by a piecewise-linear function ψ_i. Assume that the interval $[\alpha_i + l_i, \alpha_i + u_i] \cap [\beta_i - u_i, \beta_i - l_i]$ is nonempty; if this interval is empty, we have a trivial case. For convenience, we define a new pair of bounds

$$l'_i = \max\{l_i, \beta_i - u_i - \alpha_i\}, \qquad u'_i = \min\{u_i, \beta_i - l_i - \alpha_i\}. \tag{13}$$

Note that $[\alpha_i + l_i, \alpha_i + u_i] \cap [\beta_i - u_i, \beta_i - l_i] = [\alpha_i + l'_i, \alpha_i + u'_i] = [\beta_i - u'_i, \beta_i - l'_i]$. Since ϕ_i is concave, there is a point $\xi^* \in [\alpha_i, \beta_i]$ such that ϕ_i is nondecreasing on $[\alpha_i, \xi^*]$ and nonincreasing on $[\xi^*, \beta_i]$. We do not have to compute ξ^* in order to approximate ϕ_i. Instead, we simply introduce tangents starting from $\alpha_i + l'_i$ and advancing to the right, and starting from $\beta_i - l'_i$ and advancing to the left.

More specifically, we introduce tangents starting from $\alpha_i+l'_i$ only if the slope at this point is nonnegative. We introduce tangents at $\alpha_i+l'_i, \alpha_i+l'_i(1+4\epsilon+4\epsilon^2), \ldots, \alpha_i+l'_i(1+4\epsilon+4\epsilon^2)^{Q_i}$, where Q_i is largest integer such that $\alpha_i + l'_i(1 + 4\epsilon + 4\epsilon^2)^{Q_i} \leq \alpha_i + u'_i$ and the slope at $\alpha_i + l'_i(1 + 4\epsilon + 4\epsilon^2)^{Q_i}$ is nonnegative.

Let $\zeta_i = \min\{\alpha_i + u'_i, \alpha_i + l'_i(1 + 4\epsilon + 4\epsilon^2)^{Q_i+1}\}$. If ϕ_i has a nonnegative slope at ζ_i, we introduce an additional tangent at ζ_i. If the slope at ζ_i is negative, we find the largest integer r_i such that the slope at $\alpha_i + l'_i(1+4\epsilon+4\epsilon^2)^{Q_i}(1+\epsilon)^{r_i}$ is nonnegative, and introduce an additional tangent at that point. Since the slope is nonnegative at $\alpha_i+l'_i(1+4\epsilon+4\epsilon^2)^{Q_i}$, we have $r_i \geq 0$, and since $\zeta_i \leq \alpha_i + l'_i(1 + 4\epsilon + 4\epsilon^2)^{Q_i+1} \leq \alpha_i + l'_i(1 + 4\epsilon + 4\epsilon^2)^{Q_i}(1 + \epsilon)^4$, we have $r_i \leq 3$. Let the tangents introduced starting from $\alpha_i + l'_i$ have slopes $s_i^0, \ldots, s_i^{Q_i+1}$ and y-intercepts $f_i^0, \ldots, f_i^{Q_i+1}$.

We introduce tangents starting from $\beta_i - l'_i$ only if the slope at this point is nonpositive. We proceed in the same way as with the tangents starting from $\alpha_i + l'_i$, and let these tangents have slopes $s_i^{Q_i+2}, \ldots, s_i^{Q_i+R_i+3}$ and y-intercepts $f_i^{Q_i+2}, \ldots, f_i^{Q_i+R_i+3}$. Also let $P_i = Q_i + R_i + 3$.

If α_i and β_i are the endpoints of $\text{proj}_{x_i} Y$, for $x_i \in (\alpha_i, \beta_i)$, the function ψ_i is given by

$$\psi_i(x_i) = \min\{f_i^p + s_i^p x_i : p = 0, ..., P_i\}, \tag{14}$$

while for $x_i \in \{\alpha_i, \beta_i\}$, we let $\psi_i(x_i) = \phi_i(x_i)$.

If α_i and β_i are in the interior of $\text{proj}_{x_i} Y$, we introduce two more tangents at α_i and β_i, with slopes $s_i^{P_i+1}, s_i^{P_i+2}$ and y-intercepts $f_i^{P_i+1}, f_i^{P_i+2}$, and let $\psi_i(x_i) = \min\{f_i^p + s_i^p x_i : p = 0, ..., P_i + 2\}$. If one of α_i, β_i is in the interior and the other is an endpoint, we use the corresponding approach in each case.

We now replace the objective function $\phi(x)$ in problem (12) with the new objective function $\psi(x) = \sum_{i=1}^n \psi_i(x_i)$, obtaining the piecewise-linear cost problem

$$Z^*_{15} = \min\{\psi(x) : x \in X\}. \tag{15}$$

The number of pieces used to approximate each concave function ϕ_i in each direction is at most $2 + \lceil \log_{1+4\epsilon+4\epsilon^2} \frac{u_i}{l_i} \rceil$, and therefore the total number of pieces used for each function is at most $4 + 2\lceil \log_{1+4\epsilon+4\epsilon^2} \frac{u_i}{l_i} \rceil$. As $\epsilon \to 0$, this bound behaves as $\frac{1}{2\epsilon} \log \frac{u_i}{l_i}$.

It remains to show that problem (15) provides a $1 + \epsilon$ approximation for problem (12), which we do by employing Lemma 1 and Theorem 1.

Lemma 4. *$Z^*_{12} \leq Z^*_{15} \leq (1 + \epsilon)Z^*_{12}$.*

Proof. Clearly, $Z^*_{12} \leq Z^*_{15}$. To prove the inequality's other side, let x^* be an optimal solution to problem (12) that satisfies Assumption 2. We will show that $\psi_i(x^*_i) \leq (1 + \epsilon)\phi_i(x^*_i)$ for $i \in [n]$. If $x^*_i \in \{\alpha_i, \beta_i\}$ then $\psi_i(x^*_i) = \phi_i(x^*_i)$. If $x^*_i \notin \{\alpha_i, \beta_i\}$, we must have $x^*_i \in [\alpha_i + l'_i, \alpha_i + u'_i] = [\beta_i - u'_i, \beta_i - l'_i]$. Since ϕ_i is concave, it is nondecreasing on $[\alpha_i, x^*_i]$, nonincreasing on $[x^*_i, \beta_i]$, or both. Without loss of generality, assume that ϕ_i is nondecreasing on $[\alpha_i, x^*_i]$.

Due to the way we introduced tangents starting from $\alpha_i + l'_i$, it follows that $x^*_i \in [\alpha_i + l'_i, \zeta_i]$. We divide this interval into two subintervals, $[\alpha_i + l'_i, \alpha_i + l'_i(1 + 4\epsilon + 4\epsilon^2)^{Q_i}]$

and $\left[\alpha_i + l_i'(1+4\epsilon+4\epsilon^2)^{Q_i}, \zeta_i\right]$. If $x_i^* \in \left[\alpha_i + l_i', \alpha_i + l_i'(1+4\epsilon+4\epsilon^2)^{Q_i}\right]$, then since ϕ_i is nondecreasing on this interval, $\psi_i(x_i^*) \le (1+\epsilon)\phi_i(x_i^*)$ follows directly from Theorem 1.

If $x_i^* \in \left[\alpha_i + l_i'(1+4\epsilon+4\epsilon^2)^{Q_i}, \zeta_i\right]$, additional steps are needed, since ϕ_i is not necessarily nondecreasing on this interval. If ϕ_i has a nonnegative slope at ζ_i, then we introduced a tangent at ζ_i, and $\psi_i(x_i^*) \le (1+\epsilon)\phi_i(x_i^*)$ again follows from Theorem 1. If the slope at ζ_i is negative, we introduced a tangent at $\alpha_i + l_i'(1+4\epsilon+4\epsilon^2)^{Q_i}(1+\epsilon)^{r_i}$. Since r_i is the largest integer such that the slope at $\alpha_i + l_i'(1+4\epsilon+4\epsilon^2)^{Q_i}(1+\epsilon)^{r_i}$ is nonnegative, and the slope at x_i^* is also nonnegative, $x_i^* \in \left[\alpha_i + l_i'(1+4\epsilon+4\epsilon^2)^{Q_i}, \alpha_i + l_i'(1+4\epsilon+4\epsilon^2)^{Q_i}(1+\epsilon)^{r_i+1}\right]$.

We now distinguish two cases. If $x_i^* \in \left[\alpha_i + l_i'(1+4\epsilon+4\epsilon^2)^{Q_i}, \alpha_i + l_i'(1+4\epsilon+4\epsilon^2)^{Q_i}(1+\epsilon)^{r_i}\right]$, then since ϕ_i is nondecreasing on this interval, $\psi_i(x_i^*) \le (1+\epsilon)\phi_i(x_i^*)$ follows by Theorem 1. If $x_i^* \in \left[\alpha_i + l_i'(1+4\epsilon+4\epsilon^2)^{Q_i}(1+\epsilon)^{r_i}, \alpha_i + l_i'(1+4\epsilon+4\epsilon^2)^{Q_i}(1+\epsilon)^{r_i+1}\right]$, note that the right endpoint of this interval is $1+\epsilon$ times farther from α_i than the left endpoint. Since ϕ_i is nondecreasing from the left endpoint to x^*, and we introduced a tangent at the left endpoint, $\psi_i(x_i^*) \le (1+\epsilon)\phi_i(x_i^*)$ follows by Lemma 1.

Taken together, the above cases imply that $Z_{15}^* \le \psi(x^*) \le (1+\epsilon)\phi(x^*) = (1+\epsilon)Z_{12}^*$. □

We conclude this section with two further extensions:

1) We can use secants instead of tangents, in which case we require on the order of one function evaluation per piece, and do not need to evaluate the derivative. The secant approach may be preferable in computational applications where derivatives are difficult to compute.

2) The results in this section can be adapted to apply to concave maximization problems.

3 Polyhedral Feasible Sets

In this section and Section 3.1, we obtain the main result of this paper by applying our approximation approach to concave cost problems with polyhedral feasible sets. We will employ the polyhedral structure of the feasible set to eliminate the quantities l_i and u_i from the bound on the number of pieces, and obtain a bound that is polynomial in the input size of the concave cost problem and linear in $1/\epsilon$.

Let $X = \{x : Ax \le b, x \ge 0\}$ be a nonempty rational polyhedron defined by a matrix $A \in \mathbb{Q}^{m\times n}$ and a vector $b \in \mathbb{Q}^m$. Let $\phi : \mathbb{R}_+^n \to \mathbb{R}_+$ be a nondecreasing separable concave function, with $\phi(x) = \sum_{i=1}^n \phi_i(x_i)$ and each function ϕ_i nonnegative. We consider the problem

$$Z_{16}^* = \min\{\phi(x) : Ax \le b, x \ge 0\}. \tag{16}$$

Following standard practice, we define the size of rational numbers, vectors, and matrices as the number of bits needed to represent them [see e.g. KV02]. More specifically, for an integer r, let $\text{size}(r) = 1 + \lceil \log_2(|r|+1)\rceil$; for a rational number $r = \frac{r_1}{r_2}$ with $r_2 > 0$, and r_1 and r_2 coprime integers, let $\text{size}(r) = \text{size}(r_1) + \text{size}(r_2)$; and for a rational vector or matrix $M \in \mathbb{Q}^{p\times q}$ with elements m_{ij}, let $\text{size}(M) = pq + \sum_{i=1}^p \sum_{j=1}^q \text{size}(m_{ij})$.

We take the input size of problem (16) to be the input size of the feasible polyhedron, size(A)+size(b). Assume that each function ϕ_i is given by an oracle that returns the function value $\phi_i(x_i)$ and derivative $\phi_i'(x_i)$ in time $O(1)$. When the concave functions are given in other ways than through oracles, the input size of problem (16) is at least $\mathrm{size}(A) + \mathrm{size}(b)$, and therefore our bound applies in those cases as well.

We will use the following classical result that bounds the size of a polyhedron's vertices in terms of the size of the constraint matrix and right-hand side vector that define the polyhedron [see e.g. KV02]. Let $U(A,b) = 4(\mathrm{size}(A) + \mathrm{size}(b) + 2n^2 + 3n)$.

Lemma 5. *If $x' = (x'_1, \ldots, x'_n)$ is a vertex of X, then each of its components has $\mathrm{size}(x'_i) \leq U(A,b)$.*

To approximate problem (16), we replace each concave function ϕ_i with a piecewise-linear function ψ_i as described in equations (2)–(4). To obtain each function ψ_i, we take

$$l_i = \frac{1}{2^{U(A,b)-1} - 1}, \qquad u_i = 2^{U(A,b)-1} - 1, \tag{17}$$

and $P_i = \lceil \log_{1+4\epsilon+4\epsilon^2} \frac{u_i}{l_i} \rceil$, and introduce $P_i + 1$ tangents to ϕ_i at $l_i, l_i(1+4\epsilon+4\epsilon^2), \ldots, l_i(1+4\epsilon+4\epsilon^2)^{P_i}$. The resulting piecewise-linear cost problem is

$$Z^*_{18} = \min\{\psi(x) : Ax \leq b, x \geq 0\}. \tag{18}$$

The number of pieces used to approximate each function ϕ_i is

$$1 + \left\lceil \log_{1+4\epsilon+4\epsilon^2} \frac{u_i}{l_i} \right\rceil \leq 1 + \left\lceil \log_{1+4\epsilon+4\epsilon^2} 2^{2U(A,b)} \right\rceil = 1 + \left\lceil \frac{2U(A,b)}{\log_2(1+4\epsilon+4\epsilon^2)} \right\rceil. \tag{19}$$

As $\epsilon \to 0$, this bound behaves as $\frac{2U(A,b)}{4(\log_2 e)\epsilon} = \frac{U(A,b)}{2(\log_2 e)\epsilon}$. Therefore, the obtained bound is polynomial in the size of the input and linear in $1/\epsilon$. The time needed to compute the piecewise-linear approximation is also polynomial in the size of the input and linear in $1/\epsilon$. Specifically, we can compute all the quantities l_i and u_i in $O(U(A,b))$, and then compute the pieces composing each function ψ_i in $O\big(\frac{U(A,b)}{\log_2(1+4\epsilon+4\epsilon^2)}\big)$ per function, for a total running time of $O\big(U(A,b) + \frac{nU(A,b)}{\log_2(1+4\epsilon+4\epsilon^2)}\big) = O\big(\frac{nU(A,b)}{\epsilon}\big)$.

Next, we apply Theorem 1 to show that problem (18) approximates problem (16) to within a factor of $1+\epsilon$.

Lemma 6. $Z^*_{16} \leq Z^*_{18} \leq (1+\epsilon)Z^*_{16}$.

Proof. It is clear that problem (16) satisfies the assumptions needed by Theorem 1, except for Assumption 1 and the requirement that X be a compact set. Next, we consider these two assumptions.

Because X is a polyhedron in $\mathbb{R}^n_+$ and ϕ is concave and nonnegative, problem (16) has an optimal solution x^* at a vertex of X [HH61]. Lemma 5 ensures that $\mathrm{size}(x^*_i) \leq U(A,b)$ for $i \in [n]$, and hence $x^*_i \in \{0\} \cup \big[\frac{1}{2^{U(A,b)-1}-1}, 2^{U(A,b)-1} - 1\big]$. Therefore, problem (16) together with the bounds l_i and u_i, and the optimal solution x^* satisfies Assumption 1.

If the polyhedron X is bounded, then Theorem 1 applies, and the approximation property follows. If X is unbounded, we add the constraints $x_i \leq 2^{U(A,b)-1} - 1$ for $i \in [n]$ to problems (16) and (18), obtaining the modified problems

$$Z^*_{16\mathrm{B}} = \min\big\{\phi(x) : Ax \leq b, 0 \leq x \leq 2^{U(A,b)-1} - 1\big\}, \tag{16B}$$

$$Z^*_{18\mathrm{B}} = \min\big\{\psi(x) : Ax \leq b, 0 \leq x \leq 2^{U(A,b)-1} - 1\big\}. \tag{18B}$$

Denote the modified feasible polyhedron by X_{B}. Since $X_{\mathrm{B}} \subseteq X$ and $x^* \in X_{\mathrm{B}}$, it follows that $Z^*_{16\mathrm{B}} = Z^*_{16}$ and x^* is an optimal solution to problem (16B). Similarly, let y^* be a vertex optimal solution to problem (18); since $X_{\mathrm{B}} \subseteq X$ and $y^* \in X_{\mathrm{B}}$, we have $Z^*_{18\mathrm{B}} = Z^*_{18}$.

Since X_{B} is a bounded polyhedron, problem (16B), together with the bounds l_i and u_i, and the optimal solution x^* satisfies the assumptions needed by Theorem 1. When we approximate problem (16B) using the approach of equations (2)–(4), we obtain problem (18B), and therefore $Z^*_{16\mathrm{B}} \leq Z^*_{18\mathrm{B}} \leq (1+\epsilon)Z^*_{16\mathrm{B}}$. The approximation property follows. □

Note that it is not necessary to add the constraints $x_i \leq 2^{U(A,b)-1} - 1$ to problem (16) or (18) when computing the piecewise-linear approximation, as the modified problems are only used in the proof of Lemma 6.

If the objective functions ϕ_i of problem (16) are already piecewise-linear, the resulting problem (18) is again a piecewise-linear concave cost problem, but with each objective function ψ_i having at most the number of pieces given by bound (19). Since this bound does not depend on the functions ϕ_i, and is polynomial in the input size of the feasible polyhedron X and linear in $1/\epsilon$, our approach may be used to reduce the number of pieces for piecewise-linear concave cost problems with a large number of pieces.

When considering a specific application, it is often possible to use the application's structure to derive values of l_i and u_i that yield a significantly better bound on the number of pieces than the general values of equation (17). We will illustrate this with two applications in Sections 5 and 6.

3.1 Extensions

Next, we generalize this result to polyhedra that are not contained in $\mathbb{R}^n_+$ and concave functions that are not monotone. Consider the problem

$$Z^*_{20} = \min\{\phi(x) : Ax \leq b\}, \tag{20}$$

defined by a rational polyhedron $X = \{x : Ax \leq b\}$ with at least one vertex, and a separable concave function $\phi : Y \to \mathbb{R}_+$. Here $Y = \{x : Cx \leq d\}$ can be any rational polyhedron that contains X and has at least one vertex. Let $\phi(x) = \sum_{i=1}^n \phi_i(x_i)$, and assume that the functions ϕ_i are nonnegative. We assume that the input size of this problem is $\mathrm{size}(A) + \mathrm{size}(b)$, and that the functions ϕ_i are given by oracles that return the function value and derivative in time $O(1)$.

Since, unlike problem (16), this problem does not include the constraints $x \geq 0$, we need the following variant of Lemma 5 [see e.g. KV02]. Let $V(A,b) = 4(\mathrm{size}(A) + \mathrm{size}(b))$.

Lemma 7. *If $x' = (x'_1, \ldots, x'_n)$ is a vertex of X, then each of its components has* $\mathrm{size}(x'_i) \leq V(A,b)$.

We approximate this problem by applying the approach of Section 2.2 as follows. If $\mathrm{proj}_{x_i} Y$ is a closed interval $[\alpha'_i, \beta'_i]$, we let $[\alpha_i, \beta_i] = [\alpha'_i, \beta'_i]$; if $\mathrm{proj}_{x_i} Y$ is a half-line $[\alpha'_i, +\infty)$ or $(-\infty, \beta'_i]$, we let $[\alpha_i, \beta_i] = \left[\alpha'_i, 2^{V(A,b)-1}\right]$ or $[\alpha_i, \beta_i] = \left[-2^{V(A,b)-1}, \beta'_i\right]$; and if the projection is the entire real line, we let $[\alpha_i, \beta_i] = \left[-2^{V(A,b)}, 2^{V(A,b)}\right]$.

If $\mathrm{proj}_{x_i} Y$ is a closed interval or a half-line, we take

$$l_i = \frac{1}{2^{V(A,b)+V(C,d)-1} - 1} \quad \text{and} \quad u_i = 2^{V(A,b)-1} + 2^{V(C,d)-1} - 1, \tag{21}$$

while if $\text{proj}_{x_i} Y$ is the entire real line, we take $l_i = 2^{V(A,b)-1}$ and $u_i = 3 \cdot 2^{V(A,b)-1}$. We then apply the approach of Section 2.2 as described from Assumption 2 onward, obtaining the piecewise-linear cost problem

$$Z_{22}^* = \min\{\psi(x) : Ax \le b\}. \tag{22}$$

The number of pieces used to approximate each function ϕ_i is at most

$$\begin{aligned} 4 + 2\left\lceil \log_{1+4\epsilon+4\epsilon^2} \frac{u_i}{l_i} \right\rceil &\le 4 + 2\left\lceil \log_{1+4\epsilon+4\epsilon^2} \left(2^{V(A,b)+V(C,d)}\left(2^{V(A,b)} + 2^{V(C,d)}\right)\right) \right\rceil \\ &\le 4 + 2\left\lceil \log_{1+4\epsilon+4\epsilon^2} \left(2^{V(A,b)+V(C,d)} 2^{V(A,b)+V(C,d)}\right) \right\rceil \\ &= 4 + 2\left\lceil \frac{2V(A,b) + 2V(C,d)}{\log_2(1+4\epsilon+4\epsilon^2)} \right\rceil. \end{aligned} \tag{23}$$

As $\epsilon \to 0$, this bound behaves as $\frac{V(A,b)+V(C,d)}{(\log_2 e)\epsilon}$. Note that, in addition to the size of the input and $1/\epsilon$, this bound also depends on the size of C and d. Moreover, to analyze the time needed to compute the piecewise-linear approximation, we have to specify a way to compute the quantities α_i' and β_i'. We will return to these issues shortly.

Next, we prove that problem (22) approximates problem (20) to within a factor of $1+\epsilon$, by applying Lemma 4.

Theorem 3. $Z_{20}^* \le Z_{22}^* \le (1+\epsilon) Z_{20}^*$.

Proof. The assumptions needed by Lemma 4 are satisfied, except for Assumption 2 and the requirement that X be a compact set. We address these two assumptions as follows.

First, note that problem (20) has an optimal solution at a vertex x^* of X, since X is a polyhedron with at least one vertex, and ϕ is concave and nonnegative [HH61]. By Lemma 7, we have $\text{size}(x_i^*) \le V(A,b)$ for $i \in [n]$, and hence $x_i^* \in \left[-2^{V(A,b)-1}+1, 2^{V(A,b)-1}-1\right]$. We add the constraints $-2^{V(A,b)-1}+1 \le x_i \le 2^{V(A,b)-1}-1$ for $i \in [n]$ to X, obtaining the polyhedron X_{B} and the problems

$$Z_{20\text{B}}^* = \min\left\{\phi(x) : Ax \le b, -2^{V(A,b)-1}+1 \le x \le 2^{V(A,b)-1}-1\right\}, \tag{20B}$$

$$Z_{22\text{B}}^* = \min\left\{\psi(x) : Ax \le b, -2^{V(A,b)-1}+1 \le x \le 2^{V(A,b)-1}-1\right\}. \tag{22B}$$

It is easy to see that $Z_{20\text{B}}^* = Z_{20}^*$ and $Z_{22\text{B}}^* = Z_{22}^*$, and that x^* is an optimal solution to problem (20B).

Clearly, X_{B} is a compact set. To see that Assumption 2 is satisfied for problem (20B), consider the following three cases:

1) If $\text{proj}_{x_i} Y = (-\infty, +\infty)$, then $\alpha_i = -2^{V(A,b)}$, $l_i = 2^{V(A,b)-1}$, and $u_i = 3 \cdot 2^{V(A,b)-1}$. As a result, $x_i^* - \alpha_i \in \left[-2^{V(A,b)-1}+1+2^{V(A,b)}, 2^{V(A,b)-1}-1+2^{V(A,b)}\right] \subseteq \{0\} \cup [l_i, u_i]$.
2) If $\text{proj}_{x_i} Y = (-\infty, \beta_i']$, then $\alpha_i = -2^{V(A,b)-1}$, and thus $x_i^* - \alpha_i \ge 1$. On the other hand, $\beta_i = \beta_i'$, implying that β_i is a component of a vertex of Y, and thus $\text{size}(\beta_i) \le V(C,d)$. Now, $x_i^* \le \beta_i$ implies that $x_i^* - \alpha_i \le 2^{V(C,d)-1} - 1 + 2^{V(A,b)-1}$. Since $l_i = \frac{1}{2^{V(A,b)+V(C,d)-1}-1}$ and $u_i = 2^{V(A,b)-1} + 2^{V(C,d)-1} - 1$, we have $x_i^* - \alpha_i \in \{0\} \cup [l_i, u_i]$.

3) If $\text{proj}_{x_i} Y = [\alpha'_i, +\infty)$ or $\text{proj}_{x_i} Y = [\alpha'_i, \beta'_i]$, then let $x^*_i = \frac{p_1}{q_1}$ with $q_1 > 0$, and p_1 and q_1 coprime integers. Similarly let $\alpha_i = \frac{p_2}{q_2}$, and note that $x^*_i - \alpha_i = \frac{p_1 q_2 - p_2 q_1}{q_1 q_2}$. Since $\alpha_i = \alpha'_i$, we know that α_i is a component of a vertex of Y, and hence $\text{size}(\alpha_i) \leq V(C, d)$ and $\text{size}(q_2) \leq V(C, d)$. On the other hand, $\text{size}(x^*_i) \leq V(A, b)$, and thus $\text{size}(q_1) \leq V(A, b)$. This implies that $\text{size}(q_1 q_2) \leq V(A, b) + V(C, d)$, and therefore either $x^*_i = \alpha_i$ or $x^*_i - \alpha_i \geq \frac{1}{2^{V(A,b)+V(C,d)-1}-1}$. Next, since $\text{size}(\alpha_i) \leq V(C, d)$ and $\text{size}(x^*_i) \leq V(A, b)$, we have $x^*_i - \alpha_i \leq 2^{V(A,b)-1} + 2^{V(C,d)-1} - 2$. Given that $l_i = \frac{1}{2^{V(A,b)+V(C,d)-1}-1}$ and $u_i = 2^{V(A,b)-1} + 2^{V(C,d)-1} - 1$, it follows that $x^*_i - \alpha_i \in \{0\} \cup [l_i, u_i]$.

Combining the three cases, we obtain $x^*_i \in \{\alpha_i\} \cup [\alpha_i + l_i, \alpha_i + u_i]$. Similarly, we can show that $x^*_i \in \{\beta_i\} \cup [\beta_i - u_i, \beta_i - l_i]$, and therefore $x^*_i \in (\{\alpha_i\} \cup [\alpha_i + l_i, \alpha_i + u_i]) \cap (\{\beta_i\} \cup [\beta_i - u_i, \beta_i - l_i])$, which is a subset of $\{\alpha_i, \beta_i\} \cup \big([\alpha_i + l_i, \alpha_i + u_i] \cap [\beta_i - u_i, \beta_i - l_i]\big)$.

Therefore, problem (20B), together with the quantities α_i and β_i, the bounds l_i and u_i, and the optimal solution x^* satisfies Assumption 2, and Lemma 4 applies. Using the approach of Section 2.2 to approximate problem (20B) yields problem (22B), which implies that $Z^*_{20\text{B}} \leq Z^*_{22\text{B}} \leq (1 + \epsilon) Z^*_{20\text{B}}$, and the approximation property follows. □

To obtain a bound on the number of pieces that is polynomial in the size of the input and linear in $1/\epsilon$, we can simply restrict the domain Y of the objective function to the feasible polyhedron X, that is let $Y := X$. In this case, bound (23) becomes $4 + 2\left\lceil \frac{4V(A,b)}{\log_2(1+4\epsilon+4\epsilon^2)} \right\rceil$, and can be further improved to $4 + 2\left\lceil \frac{3V(A,b)}{\log_2(1+4\epsilon+4\epsilon^2)} \right\rceil$, which behaves as $\frac{1.5V(A,b)}{(\log_2 e)\epsilon}$ as $\epsilon \to 0$.

When $Y = X$, the time needed to compute the piecewise-linear approximation is also polynomial in the size of the input and linear in $1/\epsilon$. The quantities α'_i and β'_i can be computed by solving the linear programs $\min\{x_i : Ax \leq b\}$ and $\max\{x_i : Ax \leq b\}$. Recall that this can be done in polynomial time, for example by the ellipsoid method [see e.g. GLS93, KV02], and denote the time needed to solve such a linear program by $T_{\text{LP}}(A, b)$. After computing the quantities α'_i and β'_i, we can compute all the quantities α_i and β_i in $O(V(A, b))$, all the bounds l_i and u_i in $O(V(A, b))$, and the pieces composing each function ϕ_i in $O\big(\frac{V(A,b)}{\log_2(1+4\epsilon+4\epsilon^2)}\big)$ per function. The total running time is therefore $O\big(nT_{\text{LP}}(A, b) + V(A, b) + \frac{nV(A,b)}{\log_2(1+4\epsilon+4\epsilon^2)}\big) = O\big(nT_{\text{LP}}(A, b) + \frac{nV(A,b)}{\epsilon}\big)$.

In many applications, the domain Y of the objective function has a very simple structure and the quantities α'_i and β'_i are included in the input, as part of the description of the objective function. In this case, using bound (23) directly may yield significant advantages over the approach that lets $Y := X$ and solves $2n$ linear programs. Bound (23) can be improved to $4+2\left\lceil \frac{2V(A,b)+2\,\text{size}(\alpha'_i)+2\,\text{size}(\beta'_i)}{\log_2(1+4\epsilon+4\epsilon^2)} \right\rceil$, and as $\epsilon \to 0$ it behaves as $\frac{V(A,b)+\text{size}(\alpha'_i)+\text{size}(\beta'_i)}{(\log_2 e)\epsilon}$. Since α'_i and β'_i are part of the input, the improved bound is again polynomial in the size of the input and linear in $1/\epsilon$.

4 Algorithms for Concave Cost Problems

Although concave cost problem (16) can be approximated efficiently by piecewise-linear cost problem (18), both the original and the resulting problems contain the set cover problem as a special case, and therefore are NP-hard. Moreover, the set cover problem does not have an approximation algorithm with a certain logarithmic factor, unless P = NP [RS97]. Therefore, assuming that P $\neq$ NP, we cannot develop a polynomial-time exact algorithm or constant factor approximation algorithm for problem (18) in the general case, and then use it to approximately solve problem (16). In this section, we show how to use our piecewise-linear approximation approach to obtain new algorithms for concave cost problems.

We begin by writing problem (18) as an integer program. Several classical methods for representing a piecewise-linear function as part of an integer program introduce a binary variable for each piece and add one or more coupling constraints to ensure that any feasible solution uses at most one piece [see e.g. NW99, CGM03]. However, since the objective function of problem (18) is also concave, the coupling constraints are unnecessary, and we can employ the following fixed charge formulation. This formulation has been known since at least the 1960s [e.g. FLR66].

$$\min \sum_{i=1}^{n} \sum_{p=0}^{P_i} (f_i^p z_i^p + s_i^p y_i^p)\,, \tag{24a}$$

$$\text{s.t. } Ax \leq b, \tag{24b}$$

$$x_i = \sum_{p=0}^{P_i} y_i^p, \qquad i \in [n], \tag{24c}$$

$$0 \leq y_i^p \leq B_i z_i^p, \qquad i \in [n], p \in \{0, \ldots, P_i\}, \tag{24d}$$

$$z_i^p \in \{0, 1\}, \qquad i \in [n], p \in \{0, \ldots, P_i\}. \tag{24e}$$

Here, we assume without loss of generality that $\psi_i(0) = 0$. The coefficients B_i are chosen so that $x_i \leq B_i$ at any vertex of the feasible polyhedron X of problem (18), for instance $B_i = 2^{U(A,b)-1} - 1$.

A key advantage of formulation (24) is that, in many cases, it preserves the special structure of the original concave cost problem. For example, when (16) is the concave cost multicommodity flow problem, (24) becomes the fixed charge multicommodity flow problem, and when (16) is the concave cost facility location problem, (24) becomes the classical facility location problem. In such cases, (24) is a well-studied discrete optimization problem and may have a polynomial-time exact algorithm, fully polynomial-time approximation scheme (FPTAS), polynomial-time approximation scheme (PTAS), approximation algorithm, or polynomial-time heuristic.

Let $\gamma \geq 1$. The next lemma follows directly from Lemma 6.

Lemma 8. *Let x' be a γ-approximate solution to problem* (18)*, that is $x' \in X$ and $Z^*_{18} \leq \psi(x') \leq \gamma Z^*_{18}$. Then x' is also a $(1+\epsilon)\gamma$ approximate solution to problem* (16)*, that is $Z^*_{16} \leq \phi(x') \leq (1+\epsilon)\gamma Z^*_{16}$.*

Therefore, a γ-approximation algorithm for the resulting discrete optimization problem yields a $(1+\epsilon)\gamma$ approximation algorithm for the original concave cost problem. More specifically, we compute a $1+\epsilon$ piecewise-linear approximation of the concave cost problem; the time needed for the computation and the input size of the resulting problem are both bounded by $O\big(\frac{nU(A,b)}{\epsilon}\big)$. Then, we run the γ-approximation algorithm on the resulting problem. The following table summarizes the results for other types of algorithms.

When the resulting discrete optimization problem has a ...	We can obtain for the original concave cost problem a ...
Polynomial-time exact algorithm	FPTAS
FPTAS	FPTAS
PTAS	PTAS
γ-approximation algorithm	$(1+\epsilon)\gamma$ approximation algorithm
Polynomial-time heuristic	Polynomial-time heuristic

In conclusion, we note that the results in this section can be adapted to the more general problems (20) and (22).

5 Concave Cost Multicommodity Flow

To illustrate our approach on a practical problem, we consider the concave cost multicommodity flow problem. Let (V, E) be an undirected network with node set V and edge set E, and let $n = |V|$ and $m = |E|$. This network has K commodities flowing on it, with the supply or demand of commodity k at node i being b_i^k. If $b_i^k > 0$ then node i is a source for commodity k, while $b_i^k < 0$ indicates a sink. We assume that each commodity has one source and one sink, that the supply and demand for each commodity are balanced, and that the network is connected.

Each edge $\{i, j\} \in E$ has an associated nondecreasing concave cost function $\phi_{ij} : \mathbb{R}_+ \to \mathbb{R}_+$. Without loss of generality, we let $\phi_{ij}(0) = 0$ for $\{i, j\} \in E$. For an edge $\{i, j\} \in E$, let x_{ij}^k indicate the flow of commodity k from i to j, and x_{ji}^k the flow in the opposite direction. The cost on edge $\{i, j\}$ is a function of the total flow of all commodities on it, namely $\phi_{ij}\big(\sum_{k=1}^K (x_{ij}^k + x_{ji}^k)\big)$. The goal is to route the flow of each commodity so as to satisfy all supply and demand constraints, while minimizing total cost.

A mathematical programming formulation for this problem is given by:

$$Z_{25}^* = \min \sum_{\{i,j\}\in E} \phi_{ij}\left(\sum_{k=1}^K (x_{ij}^k + x_{ji}^k)\right), \tag{25a}$$

$$\text{s.t.} \sum_{\{i,j\}\in E} x_{ij}^k - \sum_{\{j,i\}\in E} x_{ji}^k = b_i^k, \qquad i \in V, k \in [K], \tag{25b}$$

$$x_{ij}^k, x_{ji}^k \geq 0, \qquad \{i, j\} \in E, k \in [K]. \tag{25c}$$

Let $B^k = \sum_{i:b_i^k>0} b_i^k$ and $B = \sum_{k=1}^K B^k$. For simplicity, we assume that the coefficients b_i^k are integral.

A survey on concave cost network flows and their applications is available in [GP90]. Concave cost multicommodity flow is also known as the buy-at-bulk network design problem [e.g CK05, CHKS06]. Concave cost multicommodity flow has the Steiner tree problem as a special case, and therefore is NP-hard, and does not have a polynomial-time approximation scheme, unless P = NP [BP89, ALM+98]. Moreover, concave cost multicommodity flow does not have an $O\big(\log^{1/2-\epsilon'} n\big)$ approximation algorithm for ϵ' arbitrarily close to 0, unless NP $\subseteq$ ZTIME$\big(\eta^{\text{polylog}\eta}\big)$ [And04].

Problem (25) satisfies the assumptions needed by Lemma 6, since we can handle the cost functions $\phi_{ij}\big(\sum_{k=1}^{K}(x_{ij}^k + x_{ji}^k)\big)$ by introducing new variables $\xi_{ij} = \sum_{k=1}^{K}(x_{ij}^k + x_{ji}^k)$ for $\{i,j\} \in E$. We apply the approach of equations (17)–(19) to approximate this problem to within a factor of $1+\epsilon$, use formulation (24) to write the resulting problem as an integer program, and disaggregate the integer program, obtaining:

$$
\begin{aligned}
Z_{26}^* = \min \quad & \sum_{\{i,j,p\}\in E'} f_{ijp} z_{ijp} + \sum_{\{i,j,p\}\in E'} \sum_{k=1}^{K} s_{ijp}(x_{ijp}^k + x_{jip}^k), && && \text{(26a)}\\
\text{s.t.} \quad & \sum_{\{i,j,p\}\in E'} x_{ijp}^k - \sum_{\{j,i,p\}\in E'} x_{jip}^k = b_i^k, && i \in V, k \in [K], && \text{(26b)}\\
& 0 \le x_{ijp}^k, x_{jip}^k \le B^k y_{ijp}, && \{i,j,p\} \in E', k \in [K], && \text{(26c)}\\
& y_{ijp} \in \{0,1\}, && \{i,j,p\} \in E'. && \text{(26d)}
\end{aligned}
$$

This is the well-known fixed charge multicommodity flow problem, but on a new network (V, E') with $(P+1)m$ edges, for a suitably defined P. Each edge $\{i,j\}$ in the old network corresponds to $P+1$ parallel edges $\{i,j,p\}$ in the new network, with p being an index to distinguish between parallel edges. For each edge $\{i,j,p\} \in E'$, the coefficient f_{ijp} can be interpreted as its installation cost, and s_{ijp} as the unit cost of routing flow on the edge once installed. The binary variable y_{ijp} indicates whether edge $\{i,j,p\}$ is installed.

For a survey on fixed charge multicommodity flow, see [AMOR95, BMM97]. This problem is also known as the uncapacitated network design problem. The above hardness results from [BP89, ALM+98] and [And04] also apply to fixed charge multicommodity flow.

By bound (19), $P \le \left\lceil \frac{2U(A,b')}{\log_2(1+4\epsilon+4\epsilon^2)} \right\rceil$, with A and b' being the constraint matrix and right-hand side vector of problem (25). However, we can obtain a much lower value of P by taking problem structure into account. Specifically, we perform the approximation with $l_i = 1$ and $u_i = B$, which results in $P \le \left\lceil \frac{\log_2 B}{\log_2(1+4\epsilon+4\epsilon^2)} \right\rceil$.

Lemma 9. $Z_{25}^* \le Z_{26}^* \le (1+\epsilon) Z_{25}^*$.

Proof. Since the objective is concave and nonnegative, problem (25) has an optimal solution at a vertex z of its feasible polyhedron [HH61]. In z, the flow of each commodity occurs on a tree [see e.g. Sch03], and therefore, the total flow $\sum_{k=1}^{K}(z_{ij}^k + z_{ji}^k)$ on any edge $\{i,j\} \in E$ is in $\{0\} \cup [1, B]$. The approximation result follows from Theorem 1. □

5.1 Computational Results

We present computational results for problems with complete uniform demand—there is a commodity for every ordered pair of nodes, and every commodity has a demand of 1. We have generated the instances based on [BMW89] as follows. To ensure feasibility, for each problem we first generated a random spanning tree. Then we added the desired number of edges between nodes selected uniformly at random. For each number of nodes, we considered a dense network with $n(n-1)/4$ edges (rounded down to the nearest multiple of 5), and a sparse network with $3n$ edges. For each network thus generated, we have considered two cost structures.

#	n	m	K	Flow Variables	Pieces
1	10	30	90	8,100	41
2	20	60	380	22,800	77
3	20	95	380	36,100	77
4	30	90	870	78,300	98
5	30	215	870	187,050	98
6	40	120	1,560	187,200	113
7	40	390	1,560	608,400	113
8	50	150	2,450	367,500	124
9	50	610	2,450	1,494,500	124
10	60	180	3,540	637,200	133
11	60	885	3,540	3,132,900	133
12	70	210	4,830	1,014,300	141
13	70	1,205	4,830	5,820,150	141
14	80	240	6,320	1,516,800	148
15	80	1,580	6,320	9,985,600	148

Table 1: Network sizes. The column "Pieces" indicates the number of pieces in each piecewise linear function resulting from the approximation.

The first cost structure models moderate economies of scale. We assigned to each edge $\{i, j\} \in E$ a cost function of the form $\phi_{ij}(\xi_{ij}) = a + b(\xi_{ij})^c$, with a, b, and c randomly generated from uniform distributions over $[0.1, 10], [0.33, 33.4]$, and $[0.8, 0.99]$. For an average cost function from this family, the marginal cost decreases by approximately 30% as the flow on an edge increases from 25 to 1,000. The second cost structure models strong economies of scale. The cost functions are as in the first case, except that c is sampled from a uniform distribution over $[0.0099, 0.99]$. In this case, for an average cost function, the marginal cost decreases by approximately 84% as the flow on an edge increases from 25 to 1,000. Note that on an undirected network with n nodes, there is an optimal solution with the flow on each edge in $\{0, 2, \ldots, n(n-1)\}$.

Table 1 specifies the problem sizes. Note that although the individual dimensions of the problems are moderate, the resulting number of variables is large, since a problem with n nodes and m edges yields $n(n-1)m$ flow variables. The largest problems we solved have 80 nodes, 1,580 edges, and 6,320 commodities. To approach them with an MIP solver, these problems would require 1,580 binary variables, 9,985,600 continuous variables and 10,491,200 constraints, even if we replaced the concave functions by fixed charge costs.

We chose $\epsilon = 0.01 = 1\%$ for the piecewise linear approximation. Here, we have been able to reduce the number of pieces significantly by using the tight approximation guarantee of Theorem 1 and the problem-specific bound of Lemma 9. After applying our piecewise linear approximation approach, we have reduced the number of pieces further by noting that for low argument values, our approach introduced tangents on a grid denser than the uniform grid $2, 4, 6, \ldots$ For each problem, we have reduced the number of pieces per cost function by approximately 47 by using the uniform grid for low argument values, and the grid generated by our approach elsewhere.

We used an improved version of the dual ascent method described by Balakrishnan

#	Moderate economies of scale				Strong economies of scale			
	Time	Sol. Edges	$\epsilon_{\mathrm{DA}}\%$	$\epsilon_{\mathrm{ALL}}\%$	Time	Sol. Edges	$\epsilon_{\mathrm{DA}}\%$	$\epsilon_{\mathrm{ALL}}\%$
1	0.13s	14	0.41	1.41	0.19s	9	0.35	1.35
2	3.17s	31	1.45	2.46	3.99s	19	1.06	2.07
3	3.50s	25.3	1.20	2.21	5.37s	19	3.38	4.42
4	18.5s	43.7	1.94	2.96	11.7s	29	1.18	2.20
5	31.3s	44	2.16	3.19	21.1s	29	3.50	4.54
6	1m23s	61.7	2.47	3.49	27.1s	39	2.20	3.22
7	2m6s	59	3.24	4.28	1m11s	39	3.17	4.21
8	3m45s	79	2.22	3.24	1m11s	49	3.42	4.46
9	6m19s	74.7	3.10	4.13	2m48s	49	4.22	5.26
10	8m45s	95	2.58	3.61	2m1s	59	3.27	4.30
11	18m52s	95.7	3.64	4.68	5m59s	59	4.25	5.29
12	16m44s	101.7	2.85	3.87	2m35s	69	3.77	4.81
13	39m18s	115.7	4.19	5.24	9m43s	69	4.98	6.03
14	32m46s	127.7	2.82	3.84	4m25s	79	4.10	5.14
15	1h24m	143	5.24	6.29	15m50s	79	5.67	6.73
Average			2.63	**3.66**			3.24	**4.27**

Table 2: Computational results. The values in column "Sol. Edges" represent the number of edges with positive flow in the obtained solutions.

et al. [BMW89] (also known as the primal-dual method [see e.g. GW97]) to solve the resulting fixed charge multicommodity flow problems. The method produces a feasible solution, whose cost we denote by Z_{26}^{DA}, to problem (26) and a lower bound Z_{26}^{LB} on the optimal value of problem (26). As a result, for this solution, we obtain an optimality gap $\epsilon_{\mathrm{DA}} = \frac{Z_{26}^{\mathrm{DA}}}{Z_{26}^{\mathrm{LB}}} - 1$ with respect to the piecewise linear problem, and an optimality gap $\epsilon_{\mathrm{ALL}} = (1+\epsilon)(1+\epsilon_{\mathrm{DA}}) - 1$ with respect to the original problem.

Table 2 summarizes the computational results. We performed all computations on an Intel Xeon 2.66 GHz. For each problem size and cost structure, we have averaged the optimality gap, computational time, and number of edges in the computed solution over 3 randomly-generated instances.

We obtained average optimality gaps of 3.66% for problems with moderate economies of scale, and 4.27% for problems with strong economies of scale. This difference in average optimality gap is consistent with computational experiments in the literature that analyze the difficulty of fixed charge problems as a function of the ratio of fixed costs to variable costs [BMW89, HS89]. Note that the solutions to problems with moderate economies of scale have more edges than those to problems with strong economies of scale; in fact, in the latter case, the edges always form a tree.

To the best of our knowledge, the literature does not contain exact or approximate computational results for concave cost multicommodity flow problems of this size. Bell and Lamar [BL97] introduce an exact branch-and-bound approach for *single-commodity* flows, and perform computational experiments on networks with up to 20 nodes and 96 edges. Fontes et al. [FHC03] propose a heuristic approach based on local search for single-source single-commodity flows, and present computational results on networks with up to 50 nodes and 200 edges. They obtain average optimality gaps of at most 13.81%, and conjecture that the actual gap between the obtained solutions and the optimal ones is much smaller. Also for single-source single-commodity flows, Fontes and Gonçalves [FG07] propose a heuristic approach that combines local search with a genetic algorithm, and present computational results on networks with up to 50 nodes and 200 edges. They obtain optimal solutions for problems with 19 nodes or less, but do not provide optimality gaps for larger problems.

6 Concave Cost Facility Location

In the concave cost facility location problem, there are m customers and n facilities. Each customer i has a demand of $d_i \geq 0$, and needs to be connected to a facility to satisfy it. Connecting customer i to facility j incurs a connection cost of $c_{ij}d_i$; the connection costs c_{ij} are nonnegative and satisfy the metric inequality.

Let $x_{ij} = 1$ if customer i is connected to facility j, and $x_{ij} = 0$ otherwise. Then the total demand satisfied by facility j is $\sum_{i=1}^m d_i x_{ij}$. Each facility j has an associated nondecreasing concave cost function $\phi_j : \mathbb{R}_+ \to \mathbb{R}_+$. We assume without loss of generality that $\phi_j(0) = 0$ for $j \in [n]$. At each facility j we incur a cost of $\phi_j\left(\sum_{i=1}^m d_i c_{ij}\right)$. The goal is to assign each customer to a facility, while minimizing the total connection and facility cost.

The concave cost facility location problem can be written as a mathematical program:

$$Z^*_{27} = \min \sum_{j=1}^n \phi_j\left(\sum_{i=1}^m d_i x_{ij}\right) + \sum_{j=1}^n \sum_{i=1}^m c_{ij} d_i x_{ij}, \tag{27a}$$

$$\text{s.t.} \sum_{j=1}^n x_{ij} = 1, \qquad i \in [n], \tag{27b}$$

$$x_{ij} \geq 0, \qquad i \in [m], j \in [n]. \tag{27c}$$

Let $D = \sum_{i=1}^m d_i$. We assume that the coefficients c_{ij} and d_i are integral.

Concave cost facility location has been studied since at least the 1960s [KH63, FLR66]. Mahdian and Pal [MP03] developed a $3 + \epsilon'$ approximation algorithm for this problem, for any $\epsilon' > 0$. When the problem has unit demands, that is $d_1 = \cdots = d_m = 1$, a wider variety of results become available. In particular, Hajiaghayi et al. [HMM03] obtained a 1.861-approximation algorithm. Hajiaghayi et al. [HMM03] and Mahdian et al. [MYZ06] described a 1.52-approximation algorithm.

Concerning hardness results, concave cost facility location contains the classical facility location problem as a special case, and therefore does not have a polynomial-time approximation scheme, unless P = NP, and does not have a 1.463-approximation algorithm, unless $\text{NP} \subseteq \text{DTIME}\big(n^{O(\log\log n)}\big)$ [GK99].

As before, problem (27) satisfies the assumptions needed by Lemma 6. We apply the approach of equations (17)–(19) to approximate it to within a factor of $1+\epsilon$, use formulation (24) to write the resulting problem as an integer program, and disaggregate the integer program:

$$Z^*_{28} = \min \sum_{j=1}^{n}\sum_{p=0}^{P} f_j^p y_j^p + \sum_{j=1}^{n}\sum_{p=0}^{P}\sum_{i=1}^{m}(s_j^p + c_{ij}) d_i x_{ij}^p, \tag{28a}$$

$$\text{s.t.} \sum_{j=1}^{n}\sum_{p=0}^{P} x_{ij}^p = 1, \qquad i \in [m], \tag{28b}$$

$$0 \le x_{ij}^p \le y_j^p, \qquad i \in [m], j \in [n], p \in \{0, \dots, P\}, \tag{28c}$$

$$y_j^p \in \{0, 1\}, \qquad j \in [n], p \in \{0, \dots, P\}. \tag{28d}$$

We have obtained a classical facility location problem that has m customers and Pn facilities, with each facility j in the old problem corresponding to P facilities (j, p) in the new problem. Each coefficient f_j^p can be viewed as the cost of opening facility (j, p), while $s_j^p + c_{ij}$ can be viewed as the unit cost of connecting customer i to facility (j, p). Note that the new connection costs $s_j^p + c_{ij}$ satisfy the metric inequality. The binary variable y_j^p indicates whether facility (j, p) is open.

Problem (28) is one of the fundamental problems in operations research [CNW90, NW99]. Hochbaum [Hoc82] showed that the greedy algorithm is a $O(\log n)$ approximation algorithm for it, even when the connection costs c_{ij} are non-metric. Shmoys et al. [STA97] gave the first constant-factor approximation algorithm for this problem, with a factor of 3.16. More recently, Mahdian et al. [MYZ06] developed a 1.52-approximation algorithm, and Byrka [Byr07] obtained a 1.4991-approximation algorithm. The above hardness results of Guha and Khuller [GK99] also apply to this problem.

Bound (19) yields $P \le \left\lceil \frac{2U(A,b)}{\log_2(1+4\epsilon+4\epsilon^2)} \right\rceil$, with A and b being the constraint matrix and right-hand side vector of problem (27). We can obtain a lower value for P by taking $l_i = 1$ and $u_i = D$, which yields $P \le \left\lceil \frac{\log_2 D}{\log_2(1+4\epsilon+4\epsilon^2)} \right\rceil$. The proof of the following lemma is similar to that of Lemma 9.

Lemma 10. $Z^*_{27} \le Z^*_{28} \le (1+\epsilon) Z^*_{27}$.

Combining our piecewise-linear approximation approach with the 1.4991-approximation algorithm of Byrka [Byr07], we obtain the following result.

Theorem 4. *There exists a* $1.4991 + \epsilon'$ *approximation algorithm for the concave cost facility location problem, for any* $\epsilon' > 0$.

We can similarly combine our approach with other approximation algorithms for classical facility location. For example, by combining it with the 1.52-approximation algorithm of Mahdian et al. [MYZ06], we obtain a $1.52 + \epsilon'$ approximation algorithm for concave cost facility location.

Acknowledgments

We are grateful to Professor Anant Balakrishnan for providing us with the original version of the dual ascent code. This research was supported in part by the Air Force Office of Scientific Research, Amazon.com, and the Singapore-MIT Alliance.

References

[ALM+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, 1998.

[AMOR95] Ravindra K. Ahuja, Thomas L. Magnanti, James B. Orlin, and M. R. Reddy. Applications of network optimization. In *Network models*, volume 7 of *Handbooks Oper. Res. Management Sci.*, pages 1–83. North-Holland, Amsterdam, 1995.

[And04] Matthew Andrews. Hardness of buy-at-bulk network design. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 115–124, Washington, DC, USA, 2004. IEEE Computer Society.

[Ata01] Alper Atamtürk. Flow pack facets of the single node fixed-charge flow polytope. *Oper. Res. Lett.*, 29(3):107–114, 2001.

[BL97] Gavin J. Bell and Bruce W. Lamar. Solution methods for nonconvex network flow problems. In *Network optimization (Gainesville, FL, 1996)*, volume 450 of *Lecture Notes in Econom. and Math. Systems*, pages 32–50. Springer, Berlin, 1997.

[BMM97] A. Balakrishnan, Thomas L. Magnanti, and P. Mirchandani. Network design. In Mauro Dell'Amico and Francesco Maffioli, editors, *Annotated bibliographies in combinatorial optimization*, Wiley-Interscience Series in Discrete Mathematics and Optimization, chapter 18, pages 311–334. John Wiley & Sons Ltd., Chichester, 1997. A Wiley-Interscience Publication.

[BMW89] A. Balakrishnan, T. L. Magnanti, and R. T. Wong. A dual-ascent procedure for large-scale uncapacitated network design. *Oper. Res.*, 37(5):716–740, 1989.

[BP89] Marshall Bern and Paul Plassmann. The Steiner problem with edge lengths 1 and 2. *Inform. Process. Lett.*, 32(4):171–176, 1989.

[Byr07] Jaroslaw Byrka. An optimal bifactor approximation algorithm for the metric uncapacitated facility location problem. In Moses Charikar, Klaus Jansen, Omer Reingold, and Jos Rolim, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 4627 of *Lecture Notes in Computer Science*, pages 29–43. Springer Berlin / Heidelberg, 2007. 10.1007/978-3-540-74208-1-3.

[CGM03] Keely L. Croxton, Bernard Gendron, and Thomas L. Magnanti. A comparison of mixed-integer programming models for nonconvex piecewise linear cost minimization problems. *Management Sci.*, 49:1268–1273, September 2003.

[CHKS06] C. Chekuri, M. T. Hajiaghayi, G. Kortsarz, and M. R. Salavatipour. Approximation algorithms for non-uniform buy-at-bulk network design. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, pages 677–686, Washington, DC, USA, 2006. IEEE Computer Society.

[CK05] Moses Charikar and Adriana Karagiozova. On non-uniform multicommodity buy-at-bulk network design. In *STOC'05: Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 176–182. ACM, New York, 2005.

[CNW90] Gérard Cornuéjols, George L. Nemhauser, and Laurence A. Wolsey. The uncapacitated facility location problem. In *Discrete location theory*, Wiley-Intersci. Ser. Discrete Math. Optim., pages 119–171. Wiley, New York, 1990.

[FG07] Dalila B. M. M. Fontes and José Fernando Gonçalves. Heuristic solutions for general concave minimum cost network flow problems. *Networks*, 50(1):67–76, 2007.

[FHC03] Dalila B. M. M. Fontes, Eleni Hadjiconstantinou, and Nicos Christofides. Upper bounds for single-source uncapacitated concave minimum-cost network flow problems. *Networks*, 41(4):221–228, 2003. Special issue in memory of Ernesto Q. V. Martins.

[FLR66] E. Feldman, F. A. Lehrer, and T. L. Ray. Warehouse location under continuous economies of scale. *Management Sci.*, 12(9):670–684, 1966.

[Geo77] Arthur M. Geoffrion. Objective function approximations in mathematical programming. *Math. Programming*, 13(1):23–37, 1977.

[GK99] Sudipto Guha and Samir Khuller. Greedy strikes back: improved facility location algorithms. *J. Algorithms*, 31(1):228–248, 1999.

[GLS93] Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, second edition, 1993.

[GM94] F. Güder and J. G. Morris. Optimal objective function approximation for separable convex quadratic programming. *Math. Programming*, 67(1, Ser. A):133–142, 1994.

[GP90] G. M. Guisewite and P. M. Pardalos. Minimum concave-cost network flow problems: applications, complexity, and algorithms. *Ann. Oper. Res.*, 25(1-4):75–99, 1990. Computational methods in global optimization.

[GW97] M.X. Goemans and D.P. Williamson. The primal-dual method for approximation algorithms and its application to network design problems. In Dorit S. Hochbaum, editor, *Approximation algorithms for NP-hard problems*, chapter 4, pages 144–191. PWS Pub. Co., Boston, 1997.

[HH61] Warren M. Hirsch and Alan J. Hoffman. Extreme varieties, concave functions, and the fixed charge problem. *Comm. Pure Appl. Math.*, 14:355–369, 1961.

[HH98] Kaj Holmberg and Johan Hellstrand. Solving the uncapacitated network design problem by a Lagrangean heuristic and branch-and-bound. *Oper. Res.*, 46(2):247–259, 1998.

[HMM03] M. T. Hajiaghayi, M. Mahdian, and V. S. Mirrokni. The facility location problem with general cost functions. *Networks*, 42(1):42–47, 2003.

[Hoc82] Dorit S. Hochbaum. Heuristics for the fixed cost median problem. *Math. Programming*, 22(2):148–162, 1982.

[HS89] Dorit S. Hochbaum and Arie Segev. Analysis of a flow problem with fixed charges. *Networks*, 19(3):291–312, 1989.

[JMM+03] Kamal Jain, Mohammad Mahdian, Evangelos Markakis, Amin Saberi, and Vijay V. Vazirani. Greedy facility location algorithms analyzed using dual fitting with factor-revealing LP. *J. ACM*, 50(6):795–824 (electronic), 2003.

[KH63] Alfred A. Kuehn and Michael J. Hamburger. A heuristic program for locating warehouses. *Management Sci.*, 9(4):643–666, 1963.

[Kon00] Spyros Kontogiorgis. Practical piecewise-linear approximation for monotropic optimization. *INFORMS J. Comput.*, 12(4):324–340, 2000.

[KV02] Bernhard Korte and Jens Vygen. *Combinatorial optimization*, volume 21 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, second edition, 2002. Theory and algorithms.

[MMP00] Adam Meyerson, Kamesh Munagala, and Serge Plotkin. Cost-distance: two metric network design. In *41st Annual Symposium on Foundations of Computer Science (Redondo Beach, CA, 2000)*, pages 624–630. IEEE Comput. Soc. Press, Los Alamitos, CA, 2000.

[MP03] Mohammad Mahdian and Martin Pál. Universal facility location. In *Algorithms—ESA 2003*, volume 2832 of *Lecture Notes in Comput. Sci.*, pages 409–421. Springer, Berlin, 2003.

[MS04] Thomas L. Magnanti and Dan Stratila. Separable concave optimization approximately equals piecewise linear optimization. In *Integer programming and combinatorial optimization*, volume 3064 of *Lecture Notes in Comput. Sci.*, pages 234–243. Springer, Berlin, 2004.

[Mun03] Kamesh Munagala. *Approximation algorithms for concave cost network flow problems*. PhD thesis, Stanford University, Department of Computer Science, March 2003.

[MYZ06] Mohammad Mahdian, Yinyu Ye, and Jiawei Zhang. Approximation algorithms for metric facility location problems. *SIAM J. Comput.*, 36(2):411–432 (electronic), 2006.

[NW99] George Nemhauser and Laurence Wolsey. *Integer and combinatorial optimization*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., New York, 1999. Reprint of the 1988 original, A Wiley-Interscience Publication.

[OW03] Francisco Ortega and Laurence A. Wolsey. A branch-and-cut algorithm for the single-commodity, uncapacitated, fixed-charge network flow problem. *Networks*, 41(3):143–158, 2003.

[PK90] P. M. Pardalos and N. Kovoor. An algorithm for a singly constrained class of quadratic programs subject to upper and lower bounds. *Math. Programming*, 46(3, (Ser. A)):321–328, 1990.

[RP86] J. B. Rosen and P. M. Pardalos. Global minimization of large-scale constrained concave quadratic problems by separable programming. *Math. Programming*, 34(2):163–174, 1986.

[RS97] Ran Raz and Shmuel Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *STOC '97 (El Paso, TX)*, pages 475–484 (electronic). ACM, New York, 1997.

[RSSZ10] H. Edwin Romeijn, Thomas C. Sharkey, Zuo-Jun Max Shen, and Jiawei Zhang. Integrating facility location and production planning decisions. *Networks*, 55(2):78–89, 2010.

[Sch03] Alexander Schrijver. *Combinatorial optimization. Polyhedra and efficiency. Vol. A*, volume 24 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 2003. Paths, flows, matchings, Chapters 1–38.

[STA97] David B. Shmoys, Éva Tardos, and Karen Aardal. Approximation algorithms for facility location problems (extended abstract). In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, STOC '97, pages 265–274, New York, NY, USA, 1997. ACM.

[Str08] Dan Stratila. *Combinatorial optimization problems with concave costs*. PhD thesis, Massachusetts Institute of Technology, Operations Research Center, September 2008.

[Tha78] Lakshman S. Thakur. Error analysis for convex separable programs: the piecewise linear approximation and the bounds on the optimal objective value. *SIAM J. Appl. Math.*, 34(4):704–714, 1978.