A Natural Randomization Strategy for Multicommodity Flow and Related Algorithms

by

Andrew V. Goldberg

Department of Computer Science
Stanford University
Stanford, California 94305

This document has been approved for public release and sale; its distribution is unlimited.
We consider the approximation algorithm of Leighton et. al. for the multicommodity flow problem. We give a more natural randomization strategy that is simpler than the one in the original algorithm and results in a better running time. This strategy also applies to several related algorithms.
A Natural Randomization Strategy
for Multicommodity Flow and Related Algorithms

Andrew V. Goldberg*
Department of Computer Science
Stanford University
Stanford, CA 94305

July 1991

*Research partially supported by NSF Presidential Young Investigator Grant CCR-8858097 with matching funds from AT&T and DEC, and ONR Young Investigator Award N00014-91-J-1855.
Abstract

We consider the approximation algorithm of Leighton et. al. [8] for the multicommodity flow problem. We give a more natural randomization strategy that is simpler than the one in [8] and results in a better running time. This strategy also applies to several related algorithms.

1 Introduction

The multicommodity flow problem is as follows: given a network with capacity constraints and commodity supplies and demands, find a flow that satisfies the demands without violating capacity constraints. The multicommodity flow problem is a classical problem that has numerous applications. The concurrent flow problem is an optimization version of the multicommodity flow problem, where the goal is to maximize the fraction of the satisfied demands, i.e., to maximize \( z \) such that if the demands are multiplied by \( z \), the resulting multicommodity flow problem is feasible.

In this paper we work with the concurrent flow problem. We denote the number of nodes in the network by \( n \), the number of arcs by \( m \), the number of commodities by \( k \), the largest capacity by \( U \), the largest demand by \( D \), and assume that capacities and demands are integral.

The only known polynomial time algorithms for the problem are based on polynomial-time methods for linear programming, either the ellipsoid method [6] or the interior-point method [5]. The fastest currently known algorithm, due to Vaidya [12], takes advantage of the matrix structure to achieve improvement over the underlying interior-point method. This algorithm runs in \( O(k^{3.5}n^3m^{0.5}\log(nDU)) \) time.

For many applications it is sufficient to find an approximate solution to the problem, i.e., a feasible solution that ships \( (1-\varepsilon)z^* \) fraction of the demands, where \( z^* \) is the optimal value. In these applications, \( \varepsilon \) is a constant or a slowly decreasing function of \( n \) (e.g. \( \frac{1}{n} \)). Shahrokhi and Matula [11] gave an approximation algorithm for a special case of the concurrent flow problem with uniform capacities. Their work motivated several other papers. A faster algorithm for the uniform capacity case was given by Klein et. al. [7]. An algorithm for the general version of the concurrent flow problem was given by Leighton et. al. [8]. The deterministic version of this algorithm runs in \( O(k^2(\varepsilon^{-2}\log \frac{n}{\varepsilon} + \log n \log k)) \) minimum-cost flow computations and the randomized version in \( O(k(\varepsilon^{-3}\log \frac{n}{\varepsilon} + \log n \log k)) \) minimum-cost flow computations. (For a survey of minimum-cost flow algorithms, see e.g. [2].)
In this paper we give a modification of the relaxed optimality conditions that leads to a slightly simpler analysis of the algorithm and allows us to show that the simplest randomization strategy, which selects a commodity to be updated during the next iteration uniformly at random, works better than the more complicated strategy used in [8]. We show an $O(k(\varepsilon^{-2}\log \frac{n}{\varepsilon} + \log n \log k))$ bound on the number of minimum-cost flow computations.

From the theoretical viewpoint, the dependence on $\varepsilon$ is a major drawback of the algorithm of [8] since exponential precision is required to solve the multicommodity flow problem exactly. We reduce this dependence by a factor of $\varepsilon^{-1}$. Our randomization strategy appears to be a better choice in practice as well.

Our results can also be applied in a straightforward way to improve the results of [7] on the uniform multicommodity flow problem, as well as in the generalizations of the multicommodity flow algorithm to other linear programming problems [10].

A recent paper of Grigoriadis and Khachiyan [4] extends the results of [8] to block-structured convex linear programs. This paper uses a randomization strategy similar to ours in a somewhat different framework.

This paper is organized as follows. Section 2 gives definitions and notation used in the paper. Section 3 describes relaxed optimality conditions. Section 4 discusses the exponential length function and its properties. Section 5 describes the algorithm and Section 6 analyses it. Section 7 discusses some aspects of the algorithm and general applicability of our techniques.

2 Definitions and Notation

In this paper we consider a directed version of the multicommodity flow problem; however, the results also hold for the undirected version with straightforward modifications. An input to the multicommodity flow problem is a graph $G = (V, E)$, a capacity function $u : V \rightarrow R^+,^1$ and a demand specification for commodities. For each commodity $i : 1 \leq i \leq k$, the specification contains the source $s_i$, the sink $t_i$, and a nonnegative demand $d_i$. Without loss of generality we assume that $m \geq n$ and that $G$ has no parallel arcs. We assume that capacities and demands are reals and denote the biggest capacity by $U$ and the biggest demand by $D$.

$^1R^+$ denotes the set of nonnegative reals.
A multicommodity flow $f$ is given by a set of functions $f_i$, $1 \leq i \leq k$, $f_i : E \to \mathbb{R}^+$. Each function $f_i$ must satisfy conservation constraints

$$\forall v \in V, \quad \sum_{(u,v) \in E} f_i(u,v) - \sum_{(v,w) \in E} f_i(v,w) = \begin{cases} 
  d_i & \text{if } v = t_i, \\
  -d_i & \text{if } v = s_i, \\
  0 & \text{otherwise.}
\end{cases}$$

We define $f(v,w) = \sum_{1 \leq i \leq k} f_i(v,w)$. A multicommodity flow is feasible if the capacity constraints are satisfied:

$$\forall (v,w) \in E, \quad f(v,w) \leq u(v,w).$$

The concurrent flow problem [11] is an optimization version of the multicommodity flow problem where the objective is to maximize $z$ such that the problem with demands $zd_i$ is feasible. An equivalent problem is to minimize $\lambda$ such that the problem with demands $d_i$ and capacities $\lambda u$ is feasible. Let $\lambda^*$ denote the optimal value of $\lambda$. We define the congestion by $\lambda(v,w) = f(v,w)/u(v,w)$, and let $\lambda = \max_{(v,w) \in E} \lambda(v,w)$. A multicommodity flow $f$ is $\epsilon$-optimal if $\lambda \leq (1 + \epsilon)\lambda^*$. In this paper we consider the problem of finding an $\epsilon$-optimal solution to the problem; throughout the paper we assume that $0 < \epsilon \leq 1$.

Next we introduce the length function, which has its roots in linear programming duality.

The length function $\ell : E \to \mathbb{R}^+$ is a nonzero, nonnegative function. Let $\text{dist}_\ell(v,w)$ denote the distance in $G$ from $v$ to $w$ with respect to $\ell$.

To simplify notation, we sometimes view length, capacity, and flow functions as vectors indexed by arcs. Given two vectors $a$ and $b$, let $a \circ b = \sum_{(v,w) \in E} a(v,w)b(v,w)$.

**Theorem 2.1** For a multicommodity flow $f$ and a length function $\ell$ we have

$$\lambda \ell(v,w) \circ u(v,w) \geq \sum_{i=1}^{k} \text{dist}_\ell(s_i, t_i) d_i.$$ 

Given a length function $\ell$, we define the cost of the flow of commodity $i$ by $C_i = f_i \circ \ell$. Let $C^*_i(\lambda)$ be the value of the minimum-cost flow $f_i$ satisfying the demands of commodity $i$ with costs $\ell$ and capacities $\lambda u$.

**Lemma 2.2** [8] For $f$, $C_i$, and $C^*_i(\lambda)$ as above,

$$\lambda \ell \circ u \geq \sum_{i=1}^{k} C_i \geq \sum_{i=1}^{k} C^*_i(\lambda).$$
A multicommodity flow $f$ minimizes $\lambda$ iff there is a length function $\ell$ for which the above terms are equal.

Lemma 2.3 [8] For $f$, $\ell$ as above,
\[ \frac{\sum_{i=1}^{k} C_i^*(\lambda)}{u \circ \ell} \]
is a lower bound on $\lambda^*$.

3 Relaxed Optimality

The relaxed optimality conditions are as follows:

\[ \forall (v, w) \in E \quad (1 + \epsilon)f(v, w) \geq \lambda u(v, w) \text{ or } u(v, w)\ell(v, w) \leq \frac{\epsilon}{m}(u \circ \ell). \] (1)

\[ (1 - 2\epsilon) \sum C_i \leq \sum C_i^*(\lambda). \] (2)

The first condition is the same as the one in [8]; it states that either an arc is close to being saturated, or its "volume" is small compared to the total "volume". The second condition states that the total cost of $f$ with respect to $\ell$ is close to the optimal.

Theorem 3.1 If $f$, $\ell$, and $\epsilon$ satisfy the relaxed optimality conditions, then $\lambda$ is at most $(1 + O(\epsilon))\lambda^*$.

Proof. From condition (1), we have
\[ \sum_{(v, w) \in E} ((1 + \epsilon)f(v, w)\ell(v, w) + \lambda \frac{\epsilon}{m} u \circ \ell) \geq \lambda u \circ \ell. \]

Rewriting, we get
\[ (1 + \epsilon)f \circ \ell \geq \lambda u \circ \ell(1 - \epsilon) \]
or
\[ \frac{1 - \epsilon}{1 + \epsilon} \lambda u \circ \ell \leq f \circ \ell = \sum C_i \leq \frac{1}{1 - 2\epsilon} \sum C_i^*(\lambda). \] (3)
By Lemma 2.3 and the previous inequality,
\[ \lambda^* \geq \frac{1}{u \circ l} \sum_i C_i^*(\lambda) \geq \frac{(1 - \epsilon)(1 - 2\epsilon)}{1 + \epsilon}, \]
and the theorem follows. ■

4 Exponential Length Function

Shahrokhi and Matula [11] were the first to use the length function that exponentially depends on flow in the context of the uniform capacities. Leighton et al. introduced the following generalization of this idea.

Define
\[ \ell(v, w) = \frac{e^{\alpha \lambda(v,w)}}{u(v,w)}. \]

Lemma 4.1 [8] Suppose \( f \) is defined by (4) and \( \alpha \geq (1 + \epsilon) \lambda^{-1} \epsilon^{-1} \ln(m \epsilon^{-1}) \). Then the first relaxed optimality condition (1) is satisfied.

Proof. Suppose \( (1 + \epsilon) f(v, w) < \lambda u(v, w) \). Then \( \lambda(v, w) < \lambda/(1 + \epsilon) \). We need to show that \( u(v, w) \ell(v, w) \leq \frac{\epsilon}{m} u \circ \ell \).

Note that
\[ \frac{\epsilon}{m} u \circ \ell = \frac{\epsilon}{m} \sum_{E} e^{\alpha \lambda(v,w)} \geq \frac{\epsilon}{m} e^{\alpha \lambda}. \]

To obtain the last inequality, we bounded the sum of nonnegative terms by the largest one.

On the other hand,
\[ u(v, w) \ell(v, w) = e^{\alpha \lambda(v,w)} < e^{\alpha \frac{\lambda}{1 + \epsilon}} = e^{\alpha \lambda} e^{-\alpha \lambda \epsilon}. \]

By the assumption of the lemma, the second term is at most \( e^{-\ln(m \epsilon)} = \frac{\epsilon}{m} \), and the lemma follows. ■

5 Algorithm Description

The algorithm maintains a multicommodity flow \( f \) and the corresponding exponential length function \( \ell \) such that \( f \) satisfies the demands. By Lemma 4.1, the first relaxed optimality
condition is always satisfied. While the second condition is not satisfied, the algorithm picks a random commodity and, if the potential function \( u \circ \ell \) decreases, replaces a fraction of its flow by the minimum-cost flow with respect to the cost function \( \ell \). The expected decrease in the potential function is large, ensuring quick termination of the algorithm with high probability.

Finding initial solution The algorithm starts with an initial solution \( f \) such that \( \lambda \leq k\lambda^* \). Such a solution is obtained by finding, for each commodity \( i \), a maximum flow \( g_i \) in the network with source \( s_i \), sink \( t_i \), and capacity function \( u \), and setting \( f_i(v, w) = g_i(v, w) \frac{d_i}{|g_i|} \), where \( |g_i| \) denotes the value of \( g_i \). It is easy to see that for the resulting flow \( f \), \( \lambda \leq k\lambda^* \). The length function \( \ell \) defined by the initial flow is also computed.

Improving the current solution At each iteration, the algorithm iteratively improves the current flow, until the desired approximation precision is obtained. An iteration (rerouting step) works as follows.

1. Select commodity \( i \) from the set \( \{1, \ldots, k\} \) uniformly at random.

2. Consider the network with capacity function \( \lambda u \) and cost function \( \ell \). Compute the minimum cost flow \( f_i^* \) in this network satisfying demands for the commodity \( i \).

3. Define \( f'_i = (1 - \sigma)f_i + \sigma f_i^* \), and let \( f' \) be the multicommodity flow obtained from \( f \) by replacing \( f_i \) by \( f'_i \).

4. Compute the length function \( \ell' \) for \( f' \).

5. If \( u \circ \ell > u \circ \ell' \) then replace \( f \) by \( f' \).

The parameters \( \alpha \) and \( \sigma \) are set as follows: \( \alpha = 2(1 + \epsilon)\lambda^{-1}\epsilon^{-1} \ln(m\epsilon^{-1}) \) and \( \sigma = \frac{\epsilon}{4\alpha \lambda} \). The values of these parameters are not updated at every iteration, but the values are updated when \( \lambda \) decreases by at least a factor of two from its value during the time of the last update. Intuitively, an iteration of the algorithm replaces a \( \sigma \) fraction of a random commodity by the same fraction of the optimal flow of this commodity. The flow is updated only if the potential function \( u \circ \ell \) decreases.
Nonscaling and scaling algorithms A simple implementation of the algorithm is to set $\epsilon$ to the desired value at the very beginning, and improve the flow until it becomes $\epsilon$-optimal.

An alternative is to scale $\epsilon$. Note, however, that we assumed that $\epsilon \leq 1$. Thus the scaling algorithm starts with a constant $\epsilon$ (say 1) and finds an $\epsilon$-optimal solution $f^{(0)}$. Then at $j$th scaling iteration, $\epsilon$ is reduced by a factor of two and a new $\epsilon$-optimal solution $f^{(j)}$ is computed starting from $f^{(j-1)}$, until the desired precision is reached.

As we shall see later, the use of scaling improves the running time of the algorithm by a factor of $\epsilon^{-1}$ if $\epsilon$ is small enough.

Termination detection The above description assumes that we know when the current flow becomes $\epsilon$-optimal. This is not the case; however, we can test for $\epsilon$-optimality by computing minimum-cost flows of all commodities. This can be done every $k$ iterations while increasing the worst-case running time bound by a constant factor.

6 Algorithm Analysis

For the purpose of the analysis, we define the potential function $\Phi$ by $\Phi = u \circ \ell$.

The next lemma shows that when $f_i$ is rerouted, $\Phi$ decreases by almost $\alpha \sigma (C_i - C^*_i(\lambda))$. The proof of this lemma is similar to the proof of [8] showing that $\Phi$ decreases significantly if a “bad” commodity is rerouted.

Let $\Phi$ and $\Phi_a$ be the values of the potential function before and after the rerouting, respectively.

Lemma 6.1

$$\Phi - \Phi_a \geq \alpha \sigma (C_i - C^*_i(\lambda)) - \epsilon C_i).$$

Proof. Let $\ell$ and $\ell_a$ be the length functions before and after the rerouting. By Taylor’s theorem, for $|t| \leq \epsilon/4 \leq 1/4$ we have

$$e^{x+t} \leq e^x + te^x + \frac{\epsilon}{2} |t| e^x.$$

Thus

$$\ell_a(v, w) \leq \ell(v, w) + \frac{\alpha \sigma}{u(v, w)} (f^*_i(v, w) - f_i(v, w)) \ell(v, w) + \frac{\epsilon \alpha \sigma}{2u(v, w)} |f^*_i(v, w) - f_i(v, w)| \ell(v, w).$$
(note that $f_i^*(v, w) - f_i(v, w) \leq \lambda u(v, w)$, and recall the choice of $\sigma$). We have

$$\Phi - \Phi_a = (\ell - \ell_a) \circ u$$

$$\geq \alpha \sigma (f_i^* - f_i) \circ \ell - \frac{\epsilon \alpha \sigma}{2} |f_i^* - f_i| \circ \ell$$

$$\geq \alpha \sigma (C_i - C_i^*(\lambda)) - \epsilon \alpha \sigma C_i.$$  

The last line follows by definition of $C_i$ and $C_i^*(\lambda)$ and the fact that $C_i \geq C_i^*(\lambda) \geq 0$. 

Note that if the current flow does not satisfy the second relaxed optimality condition, then the expected value of $C_i - C_i^*(\lambda)$ is large and $\Phi$ decreases significantly. The following lemma formalizes this statement.

**Lemma 6.2** Suppose $\frac{\epsilon}{8\alpha \lambda} \leq \sigma \leq \frac{\epsilon}{4\alpha \lambda}$ and $f$ does not satisfy the second relaxed optimality condition. Then the expected decrease in $\Phi$ due to a rerouting step is $\Omega(\frac{\epsilon^2}{k})$.

**Proof.** Since the algorithm selects a commodity uniformly at random, we have

$$E[\Phi - \Phi_a] \geq \frac{1}{k} \alpha \sigma \sum_i (C_i - C_i^*(\lambda) - \epsilon C_i)$$

$$= \frac{\alpha \sigma}{k} \left( \sum_i ((1 - 2\epsilon)C_i - C_i^*) + \epsilon \sum_i C_i \right)$$

$$\geq \frac{\alpha \sigma}{k} \frac{1 - \epsilon}{1 + \epsilon} \epsilon \lambda \Phi$$

$$= \Omega(\frac{\epsilon^2}{k}).$$

The third line follows from the assumption that $f$ does not satisfy the second relaxed optimality condition and since by (3)

$$\frac{1 - \epsilon}{1 + \epsilon} \lambda \Phi \leq \sum_i C_i.$$  

The last line follows from the assumption on $\sigma$. 

Note that since $\alpha$ and $\sigma$ are updated every time $\lambda$ decreases by a factor of two, conditions of the lemmas 4.1 and 6.2 are always satisfied.

Next we analyze the nonscaling algorithm.

8
Lemma 6.3 Suppose the parameters $\alpha$ and $\sigma$ are set when the congestion is $\lambda_0$. In expected $O(\varepsilon^{-3} k \log \frac{n}{\varepsilon})$ iterations, either the algorithm terminates or $\lambda$ decreases to $\lambda_0/2$ or less.

Proof. Initially, $\Phi \leq m e^{\alpha \lambda_0}$ since the congestion on any arc is at most $\lambda_0$. Note that until $\lambda$ decreases by a factor of two, the congestion on some arc is at least $\lambda_0/2$, so $\Phi \geq e^{\alpha \lambda_0/2}$. Thus $\Phi$ cannot decrease by more than a factor of $m e^{\alpha \lambda_0/2}$ without $\lambda$ getting below $\lambda_0/2$.

By Lemma 6.2, $O(\varepsilon^{-2} k)$ iterations reduce $\Phi$ by a factor of two. $\Phi$ can be halved at most $O(\log(m e^{\alpha \lambda_0/2})) = O(\varepsilon^{-1} \log \frac{n}{\varepsilon})$ times before $\lambda$ is halved. Therefore in $O(\varepsilon^{-3} k \log \frac{n}{\varepsilon})$ iterations (expected), the algorithm terminates or $\lambda$ is halved. $\blacksquare$

Note that the work done by an iteration of the algorithm is dominated by a minimum-cost flow computation. Combined with the above lemma and the fact that initially $\lambda \leq k \lambda^*$, this yields the following result.

Theorem 6.4 The nonscaling algorithm runs in expected $O(\varepsilon^{-3} k \log \log \frac{n}{\varepsilon})$ minimum-cost flow computations.

To analyze the scaling algorithm, we need the following version of Lemma 6.3.

Lemma 6.5 Suppose the parameters $\alpha$ and $\sigma$ are set when the congestion is $\lambda_0$, and the initial flow is $O(\varepsilon)$-optimal. Then in expected $O(\varepsilon^{-2} k \log \frac{n}{\varepsilon})$ iterations, either the algorithm terminates or $\lambda$ decreases by at least a factor of two.

Proof. Recall that $\varepsilon \leq 1$, so $\lambda_0 = O(\lambda^*)$. The proof is similar to that of Lemma 6.3, except that the upper bound on $\Phi$ is $m e^{\alpha(1+O(\varepsilon)\lambda^*)}$ and the lower bound is $e^{\alpha \lambda^*}$, so $\Phi$ needs to decrease by a factor of $m e^{\alpha O(\varepsilon)\lambda}$. $\blacksquare$

The scaling algorithm starts with $\varepsilon = O(1)$, and obtains an $O(1)$-optimal solution in $O(k \log k \log m)$ iterations by Theorem 6.4. Then the scaling process starts. Since the number of rerouting steps needed to reduce $\varepsilon$ by a factor of two is proportional to $\varepsilon^{-2}$, the last scaling iteration dominates. This iteration terminates in $O(\varepsilon^{-2} k \log \frac{n}{\varepsilon})$ rerouting steps.

We have

Theorem 6.6 The scaling algorithm runs in expected $O(k(\varepsilon^{-2} \log \frac{n}{\varepsilon} + \log k \log n))$ minimum-cost flow computations.
7 Remarks

First, we would like to address the selection of a minimum-cost flow subroutine. Since the costs in the minimum-cost subproblems can be big due to the use of the exponential length function, Orlin's strongly polynomial algorithm [9] appears to be the best choice. However, it is enough to solve the subproblems approximately, so the costs can be rounded to small integers. When this is done, the cost scaling algorithm of Goldberg and Tarjan [3] or the double scaling algorithm of Ahuja et. al. [1] become a better choice. The selection of the subroutine is discussed in more detail in [8].

The deterministic version of the algorithm of Leighton et. al. [8] finds a commodity $i$ with the biggest $C_i - C_i^*(\lambda)$ and reroutes this commodity. Our variant of the relaxed optimality conditions also can be used to analyze this algorithm. Since the maximum of a set of numbers is at least as big as the average, the deterministic choice of $i$ gives an improvement in $\Phi$ that is at least as big as the expected improvement. Thus the bound on the number of rerouting steps apply for the deterministic version of the algorithm. Each deterministic rerouting step, however, requires $k$ minimum-cost flow computations (to compute $C_j^*(\lambda)$ for every commodity $j$); in contrast, a randomized rerouting step requires a single minimum-cost flow computation.

Our randomization strategy can also be applied to the algorithm of Klein et. al. [7]. This improves the running time bound of the randomized algorithm of [7] by a factor of $\epsilon^{-1}$. We omit the details, which are straightforward given those of [7], [8], and this paper. A recent paper of Plotkin, Shmoys, and Tardos [10], that extends the results of [8] to a more general class of linear programming problems, also extends our randomization strategy to the more general framework.

Acknowledgements

The author would like to thank Éva Tardos and Serge Plotkin for helpful discussions, and Robert Kennedy and Tomasz Radzik for comments on a draft of this paper.

References


