Randomized Rounding: A Technique for Provably Good Algorithms and Algorithmic Proofs

Prabhakar Raghavan. Clark D. Thompson

Computer Science Division 573 Evans Hall U.C. Berkeley, CA 94720

ABSTRACT

We study the relation between a class of 0-1 integer linear programs and their rational relaxations. We show that the rational optimum to a problem instance can be used to construct a provably good 0-1 solution by means of a randomized algorithm. Our technique can be extended to provide bounds on the disparity between the rational and 0-1 optima for a given problem instance.

May 29, 1985

Supported by Semiconductor Research Corporation Grant SRC 1-44247-52055.

Randomized Rounding: A Technique for Provably Good Algorithms and Algorithmic Proofs

Prabhakar Raghavan Clark D. Thompson

Computer Science Division 573 Evans Hall U.C. Berkeley, CA 94720

1. General Outline

The relation of integer programs to their rational relaxations has been the subject of considerable interest [1,4,7]. Such efforts usually fall into two categories: (1) Showing existence results for feasible solutions to integer programs in terms of the solutions to their rational relaxations, and (2) Using the information derived from solutions to the relaxations to construct provably good solutions to the integer programs.

We present a technique here which we call randomized rounding. This technique is applicable to a class of 0-1 integer linear programs, and yields results in both the categories listed above. Our technique is probabilistic; for the existence results, we prove that the solution to an integer program satisfies a certain property by showing that a randomly generated solution satisfies that property with non-zero probability. In this random generation of solutions, we make use of the optimal solution to the rational relaxation linear program. By modifying the procedure used to derive the existence result, we can obtain an algorithm that is provably good in the following sense. We show that with high probability, our algorithm will provide an integer solution in which the objective function takes on a value close to the optimum of the rational relaxation (the optimal value of the objective function in the relaxed version is no worse than the optimal value of the objective function in the original 0-1 integer program).

We now give a general outline of the technique. Let Π_I be a 0-1 linear program, with variables $x_i \in \{0,1\}$. Let Π_R be its rational relaxation, with $x_i \in [0,1]$. The basic algorithm consists of the following two phases:

- (1) Solve Π_R ; let the variables take on values $\hat{x}_i \in [0,1]$.
- (2) In this phase, the variables x_i are randomly set to one or zero according to the following rule:

$$Prob. [x_i = 1] = \hat{x}_i \tag{1.1}$$

In proving results about the outcome of the two-phase process described above, we repeatedly make use of the following results from probability theory. Let B(m,p,N) denote the probability that there will be more than m successes in N Bernoulli trials each with success probability p.

THEOREM 1.1 (Hoeffding): If $\Psi_1, \Psi_2 \Psi_N$ are completely independent Bernoulli trials such that $E(\Psi_k) = p_k, \ \Psi = \Psi_1 + \Psi_2 + + \Psi_N$ we have

$$P(\Psi \ge m) \le B(m, p, N) \tag{1.2}$$

where

$$p = \frac{\sum\limits_{k=1}^{N} p_k}{N}$$

The other fact that we require is the well-known bound due to H.Chernoff [2]:

THEOREM 1.2: If $m = (1+\beta) \cdot Np$, then for $0 \le \beta \le 1$,

$$B(m,p,N) < exp(-\frac{\beta^2 Np}{2})$$
 (1.3)

In the next two sections, we provide some direct applications of the technique: section 2 deals with a routing problem that arises in the design of VLSI circuits, and the following section treats the 0-1 multicommodity flow problem. Section 4 provides an extension to the basic technique in order to deal with some situations that cannot be directly handled. The problem of simple k-matching is used to illustrate this extension, which we call scaling. Section 5 concludes with remarks on whether our bounds can be improved.

2. A Routing Problem in VLSI

In this section we illustrate the basic principles of the randomized rounding technique by means of a routing problem that arises in the design of a certain class of VLSI circuits. The problem is that of global routing in gate-arrays [10], and is defined as follows.

We are given a two-dimensional rectilinear $n \times n$ lattice L_n ; in the context of gate-arrays, lattice-nodes represent logic circuit elements and lattice-edges represent channels in which wires used to connect the nodes can be routed. In an instance of the problem, we are given a collection of nets, where a net a_i is a set of nodes to be connected by means of a Steiner tree in L_n . In addition, for each net a_i , we are given a set of possible trees b_{ij} that can be used for connecting the nodes in that net. A solution to the problem consists of choosing one tree for each net in the instance, from the allowed possibilities for that net. The number of trees in a solution that contain a given edge is termed the width of that edge in that solution. The width of a solution is the maximum width of an edge taken over all edges in the lattice. Our objective is to find a solution of minimum width.

This problem is readily formulated as a 0-1 integer program by assigning a variable for each configuration of each net: thus, let x_{ij} be an indicator variable denoting whether or not the j^{th} tree b_{ij} is chosen for net a_i . Constraints of the form

$$\sum_{i} x_{ij} = 1, \forall i \qquad (2.1)$$

ensure that a choice is made for each net. The number of trees in the solution that contain a given edge e is bounded above by an unknown quantity W which we seek to minimize as objective function. We express these notions by means of

$$\sum_{b_{ij} \text{ contains } e} x_{ij} \leq W , \forall e$$
 (2.2)

and

Minimize
$$W$$
, s.t. (2.1), (2.2) and $x_{ij} \in \{0,1\} \ \forall \ i,j$ (2.3)

Consider a linear programming relaxation of (2.3) in which fractional solutions are allowed:

Minimize
$$W$$
, s.t. (2.1), (2.2) and $x_{ij} \in [0,1] \ \forall i,j$ (2.3a)

The optimum solution to (2.3a) can be found in polynomial time [5]. Let the optimum (fractional) value of x_{ij} be \hat{x}_{ij} . Furthermore, let W_1 be the optimum width obtained from the linear program solution; W_1 is a lower bound on the best possible integer optimum width. We now seek to use these fractional solutions to obtain integer solutions to (2.3). We do this by means of randomization: for each i, set x_{ij} to one with probability \hat{x}_{ij} . The choice is done in an exclusive manner, i.e. for each i exactly one of the x_{ij} is set to one; the rest are set to zero. (Constraint (2.1) ensures that we can do this). This random choice is made independently for all i.

THEOREM 2.1: For any ϵ such that $0 < \epsilon < 1$, the width of the solution produced by the above procedure does not exceed

$$W_1 + \left[2 \cdot W_1 \cdot \ln \frac{2n(n-1)}{\epsilon} \right]^{1/2} \tag{2.4}$$

with probability at least $1-\epsilon$.

PROOF: The proof follows from the observation that the width of a lattice-edge e is the sum of independent Bernoulli trials. The expected value of this sum is no more than W_1 , since the biases used for the coin-flipping were the \hat{x}_{ij} determined by the LP. Hoeffding's lemma is thus applicable with $p = \frac{W_1}{N}$. Chernoff's bound is now applied with

$$\beta \ge \left\lceil \frac{2 \ln \frac{2n(n-1)}{\epsilon}}{W_1} \right\rceil^{1/2}$$

This ensures that the rounded width of any edge does not exceed the figure in (2.4) with probability at least $1 - \frac{\epsilon}{2n(n-1)}$; then the maximum of the widths of the 2n(n-1) edges in the lattice does not exceed (2.4) with probability $1-\epsilon$.

The second (randomization) stage can be repeated to improve the solution; we thus have a Las Vegas procedure. In $n \times n$ gate-arrays, W_1 grows as n^c [9] for some $c \in \{0.5,1\}$; the approximation of Theorem 2.1 is thus asymptotically a good one.

Theorem 2.1 gives (probabilistically) a provably good solution to the routing problem. Viewed in a slightly different light, it is also a proof that there exists an integer solution to (2.3) whose objective function value is related to W_1 by the following relation.

THEOREM 2.2: Let W_1 be the optimum objective function value of the linear programming relaxation of (2.3). Then, there exists an integer solution of width not exceeding

$$W_1 + \left[2 \cdot W_1 \cdot \ln \ 2n(n-1)\right]^{1/2} \tag{2.5}$$

PROOF: Similar to the proof of Theorem 2.1. Using the Chernoff/Hoeffding bounds, we show that the probability that an edge has width exceeding the quantity in (2.5) is less than 1/2n(n-1). It follows that the probability is non-zero that none of the 2n(n-1) edges has width exceeding (2.5), in a randomly generated solution. This proves the existence of such an integer solution.

3. Undirected Multicommodity Flow Problems

In undirected multicommodity flow problems, we are given an undirected graph G(V,E). In an instance of the problem, various vertices are the sites of as sources and sinks (sources are denoted by s_i and sinks by t_i , $1 \le i \le k$). A vertex $v \in V$ may be the location of more than one source (sink). One unit of flow is to be conveyed from each source s_i to its corresponding sink t_i through the edges in E. Each edge $e \in E$ has a capacity c(e) which is an upper limit on the total amount of flow in e. We insist that the flow of any commodity in any edge be either zero or one. Note that an edge could have flow going in both directions; for instance, the flow from s_i to t_i (hereafter referred to as the flow of commodity i) could be in a direction opposite to that of the flow of commodity j in some edge e. Each of these commodities uses up one unit of the capacity of that edge, regardless of their direction.

We consider two types of such multicommodity flow problems. In the first kind, we try to maximize the total flow subject to meeting the capacity constraints (as well as conservation constraints for each commodity at each vertex). In a second variant of the problem, we require that all edges must have the same capacity; we try to minimize this common capacity while

realizing unit flows for all k commodities. In this section, we focus on the second variant; the techniques used in its solution, together with some new ones to be introduced in the next section, can be used to solve the first variant. The general integral problem is known to be NP-Complete [3], although the non-integral version can be solved using linear programming methods [6] in polynomial time.

The algorithm consists of the following three major phases:

- (1) Solving a non-integral multi-commodity flow problem.
- (2) Path stripping.
- (3) Randomized path selection.

Non-integral Multicommodity Flow: As in the previous section, we relax the requirement of 0-1 flows to allow fractional flows in the interval [0,1]. The relaxed capacity-minimization problem can be solved, for instance, by linear programming. Let us then assume that we have solved the non-integral problem and assigned to each edge $e \in E$ a flow $f_i(e) \in [0,1]$ for each commodity i. A capacity constraint of the form

$$\sum_{i=0}^{k} f_i(e) \leq C \tag{3.1}$$

is then satisfied for each $e \in E$, where C is the optimal solution to our non-integral edge-capacity optimization problem. As before, C is a lower bound on the best possible integral solution.

Path stripping: The main idea of this phase is to convert the edge flows for each commodity i into a set Γ_i of possible paths which could be used to realize the flow of that commodity. Initially, Γ_i is empty.

For each i:

(1) Form a directed graph $G_i(V, E_i)$ where E_i is a set of directed edges derived from E as follows: For each $e \in E$, assign a direction to e which is the direction of flow of commodity i in e. If $f_i(e) = 0$, e is excluded from E_i .

(2) Discover a directed path $\{e_1, \dots, e_p\}$ in G_i from s_i to t_i using a depth-first search, discarding loops. Let

$$f_m = \min \{ f_i(e_i), 1 \le j \le p \}$$
 (3.2)

For $1 \le j \le p$, replace $f_i(e_j)$ by $f_i(e_j) - f_m$. Add the path $\{e_1, \dots, e_p\}$ to Γ_i along with its weight f_m .

(3) Remove any edges with zero flow from E_i. If there is non-zero flow leaving s_i, repeat step (2). Otherwise, next i.

It is clear that the above process terminates, since at each execution of step (2), at least one edge (the one with minimum flow in the path) is deleted from E_i . Thus the number of times it is executed is upper bounded by $|E_i|$. It is also evident that on termination, the sum of the weights of the paths in Γ_i is one.

Randomization: For each i:

Cast a $|\Gamma_i|$ -faced die with face-probabilities equal to the weights of the paths in Γ_i . Assign to net N_i the path whose face comes up. Next i.

We can then prove a theorem similar to Theorem 2.1:

THEOREM 3.1: For any ϵ such that $0 < \epsilon < 1$, the integer capacity of the solution produced by the above procedure does not exceed

$$C + \left[2 \cdot C \cdot \ln \frac{|E|}{\epsilon}\right]^{1/2} \tag{3.3}$$

with probability at least $1-\epsilon$.

PROOF: The proof is similar to that of Theorem 2.1, invoking Hoeffding's and Chernoff's inequalities. The expected number of unit flows through edge e is given by (3.1).

An existence result similar to Theorem 2.2 can be inferred readily from the above theorem. In [10] it is shown that the path-stripping and randomization phases described above can be replaced by a random-walk, with the same results.

4. Randomized Rounding with Scaling

The problems considered in the previous sections were similar in that the right-hand sides of the major constraints were the objective function itself (W in the routing problem and C in the flow problem). In this section we will consider the case when the right-hand sides of the constraints defining the problem are parameters independent of the objective function. A new technique which we call scaling is introduced in order to handle such problems.

Let k be a fixed quantity. Suppose we have a constraint of the form

$$\sum_{i} x_{i} \leq k \tag{4.1}$$

where the x_i are variables confined to the interval [0,1]. Moreover, suppose we have fractional values \hat{x}_i for these variables (derived from the solution of the appropriate relaxation), and the \hat{x}_i are then interpreted as probabilities for a randomized rounding phase as in the previous sections. The difficulty lies in the fact that there is a significant probability that the values of the x_i after rounding will not satisfy (4.1). Furthermore, it is not clear whether there is a non-zero probability that the randomized rounding will yield a solution in which none of the constraints is violated. We now present a device by which we can reduce the probability that a constraint is violated to less than 1/n we call this device scaling. In this manner, we reduce the probability that any constraint is violated to less than one.

The idea is to multiply each of the \hat{x}_i by some fraction less than one. The resultant value is used in the rounding stage as the probability that x_i is set to one. Intuitively, this reduces the number of variables that are set to one and thus the probability that a constraint is violated. The example below illustrates the scaling technique, together with the details of determining the fraction used. We consider the problem of simple k-matching defined below. We use the terminology of Lovász [7].

A hypergraph H is a finite set of edges, where an edge is a non-empty subset of an n-element set V. The elements of V are called vertices. A k-matching of H is a set M of edges such that each vertex in V belongs to at most k of the edges in M. The maximum number of edges in any

k-matching of H is denoted by $v_k(H)$. A k-matching is simple if no edge of H occurs more than once in M. The maximum number of edges in any simple k-matching of H is denoted by $\widetilde{v}_k(H)$. The problem of determining $\widetilde{v}_k(H)$ can be formulated as an integer program as follows.

Suppose H has n vertices and m edges. Let A be an $n \times m$ matrix in which all the entries are either zero or one; A represents the vertex-edge incidence matrix of H. Let x_i , $1 \le i \le m$ be 0-1 indicator variables that denote whether or not edge i is in M. Let x denote the m-vector of these variables. Let k be a fixed quantity. The constraints are represented by

$$\mathbf{A} \cdot \mathbf{x} < k \cdot \mathbf{u}_{\mathbf{n}} \tag{4.2}$$

where u_n is the n-vector of all ones. Consider the 0-1 integer linear program:

Maximize
$$\sum_{i=1}^{m} z_i$$
, s.t. (4.2), $z_i \in \{0,1\}$ (4.3)

As usual, we solve the LP relaxation with $x_i \in [0,1]$. Let \widetilde{v}_k^* be the optimum value of the objective function. Instead of directly proceeding to the randomization phase, we multiply the optimal values \widehat{x}_i for the variables by the quantity $1-\delta$; the computation of δ is described below. Let

$$x'_{i} = \hat{x}_{i} \cdot (1 - \delta) \tag{4.4}$$

In the randomization stage, we now use the values x'_i as the probabilities rather than the \hat{x}_i . After rounding, the expected sum of any row of (4.2) is no more than $k \cdot (1-\delta)$. The expected value of the objective function is $\tilde{v}_k^* \cdot (1-\delta)$. In proving the quality of the rounded solution, we require an additional result from probability theory:

THEOREM 4.1 (Bernstein [8]): Let A be one of the possible outcomes of an experiment, suppose p = P(A) > 0 and put q = 1-p. Let the random variable Φ_N denote the relative frequency of A in an experiment consisting of N independent trials. Then for $0 < \epsilon < pq$ we have

$$P(||\Phi_N - p|| \ge \epsilon) \le 2 \cdot \exp\left[-\frac{N\epsilon^2}{2pq(1 + \frac{\epsilon}{2pq})^2}\right]$$
 (4.5)

Using Bernstein's Theorem together with the Chernoff/Hoeffding results, we can now prove

THEOREM 4.2: Let δ_1 and δ_2 be positive constants such that $\delta_2 > n \cdot e^{-r/4}$ and $\delta_1 + \delta_2 < 1$. Let $a = \frac{2}{r} \ln \frac{n}{\delta_2}$ and

$$v'_{k} = \widetilde{v}_{k}^{*} \cdot (1 - \delta) = \widetilde{v}_{k}^{*} \cdot (1 - \frac{(a^{2} + 4a)^{1/2} - a}{2})$$
 (4.6)

Then there exists an integer solution to (4.3) satisfying

$$\widetilde{v}_{k} \geq v'_{k} - \left[(2bv'_{k})^{1/2} + b \right] \cdot \left[\frac{bm}{2v'_{k} (m - v'_{k})} + 1 \right]$$
(4.7)

where $b = \ln \frac{2}{\delta_1}$.

REMARK: In essence, Theorem 4.2 guarantees the existence of an integer solution of value $v'_k - O(\{v'_k\}^{1/2})$.

PROOF: By Lemma 4.3 below,

Prob. [A constraint is violated]
$$< \frac{\delta_2}{n}$$

Thus the probability that any of the n constraints is violated is less than δ_2 . By Lemma 4.4,

Prob. [(4.7) is violated]
$$\leq \delta_1$$

Since δ_1 + δ_2 < 1, the statement of the theorem follows.

LEMMA 4.3: For the choice of δ in equation (4.6),

Prob. [A constraint is violated]
$$< \frac{\delta_2}{n}$$

PROOF: Directly from the Chernoff/Hoeffding bounds, with $\beta = \frac{\delta}{1-\delta}$. The condition on δ_2 in the statement of Theorem 4.2 guarantees that a < 1/2 and thus that $\beta < 1$.

LEMMA 4.4:

Prob. [(4.7) is violated]
$$< \delta_1$$

PROOF: Since we are bounding the probability of deviations below the mean, we use Bernstein's Theorem here, with $p = \frac{v'_k}{m}$, to get an upper bound. Solving the quadratic equation that arises, we find that if we choose the following deviation from the mean v'_k , the tail probability is no more than δ_1 :

Deviation
$$= \frac{b + (2b \ v'_k \cdot \frac{(m - v'_k)}{m})^{1/2}}{1 - \frac{bm}{2 \ v'_k \ (m - v'_k)}}$$

$$\leq ((2bv'_k)^{1/2} + b) \cdot (1 + \frac{bm}{2v'_k (m - v'_k)})$$

Note that Theorem 4.1 applies only if $k > 4 \cdot \ln n$, for otherwise $\delta_2 \ge 1$. (Using a more accurate form of the Chernoff bound we can handle values of k somewhat less than $4 \cdot \ln n$, but we still require k to be $\Omega(\log n)$).

For the sake of variety, we have chosen to illustrate an existence result here rather than an algorithm as in the previous sections. By introducing a parameter ϵ representing the failure probability, we can modify the above theorem so that the probability that the procedure succeeds is $1-\epsilon$ rather than merely non-zero. This provides us with a provably good algorithm for simple k-matching.

In Theorem 4.2, we do not require all the RHS values in (4.3) to be the same. Consider the modification

Maximize
$$\sum_{i=1}^{m} x_i$$
 s.t. $\mathbf{A} \cdot \mathbf{x} \leq \mathbf{r}$ (4.8)

where r is an n-vector of RHS values r_i , $1 \le i \le n$. This may be thought of as a resource allocation problem where r_i units of resource i are available. Each of m jobs requires one unit of each of various resources; if all resources necessary for a job are available, it can be scheduled. We wish to maximize the number of jobs scheduled.

The following Theorem is analogous to Theorem 4.2.

Theorem 4.5: Let δ_1 and δ_2 be positive constants such that $\delta_1 + \delta_2 < 1$. Let

$$v'_{k} = \widetilde{v}_{k}^{*} (1-\delta)$$

where \widetilde{v}_k^* is the rational optimum of (4.8). If there exists a constant δ in the interval (0,1/2] such that

$$\sum_{i=1}^{n} \exp\left[-\frac{\delta^2 r_i}{2(1-\delta)}\right] < \delta_2 \tag{4.9}$$

then there exists an integer solution to (4.8) with objective function value at least

$$v'_{k} - \left[(2bv'_{k})^{1/2} + b \right] \cdot \left[\frac{bm}{2v'_{k}(m - v'_{k})} + 1 \right]$$
 (4.10)

where $b = \ln \frac{2}{\delta_1}$.

PROOF: Similar to Theorem 4.2, with Lemma 4.3 replaced by the condition (4.9).

5. Conclusions

Our results from the preceding sections deal with a class of 0-1 optimization problems. In sections 2 and 3, we developed solutions to routing and multicommodity flow problems that were close to the best possible solution. In section 4, we studied a matching problem and a resource allocation problem. In both cases, we were able to show the existence of solutions close to the rational optimum.

We have been able to apply randomized rounding only to 0-1 optimization problems with a very special structure. Furthermore, even for such structured problems, we require that the problem parameters lie in specific ranges in order that the technique be effective. For instance, in the k-matching problem in section 4, k had to be $\Omega(\log n)$.

It is worth examining the tightness of our results. In general, there are two main factors that make the bounds loose. Analysis of the sum of independent Bernoulli trials of success

probabilities p_1 , p_2 ,, p_N shows that Hoeffding's inequality is tightest when the probabilities are equal. If the probabilities in any problem instance Π_R span a wide range, Hoeffding's bound is weak. A second weakness of our bounds is that they relate a feasible 0-1 solution to the rational optimum, not to the 0-1 optimum. In some problem instances, the 0-1 optimum differs significantly from the rational optimum; our bounds would then be closer to the best possible than is suggested by our theorems.

It is not clear whether we can extend the idea of randomized rounding to other convex 0-1 optimization problems. This is because Chernoff-type bounds only apply to the sum of independent random variables; since we interpret optimal fractional values as probabilities, we must have linear constraints and a linear objective function in order to use the Chernoff bound. It would be interesting to explore the quality of the randomized rounding approximation in, say, convex quadratic 0-1 optimization problems.

6. Acknowledgements

The authors are grateful to Richard Karp for his assistance at various points in this work, and to Narendra Karmarkar for suggesting that a probabilistic technique could be used for the VLSI routing problem described in section 2.

References

- Ron Aharoni, Paul Erdos, and Nathan Linial, "Dual Integer Linear Programs and the Relationship between their Optima," Proceedings of the Seventeenth ACM Symposium on Theory of Computing, ACM, New York, May 1985.
- H. Chernoff, "A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the Sum of Observations," Ann. Math. Stat., vol. 23, pp. 493-509, 1952.
- 3. S. Even, A. Itai, and A. Shamir, "On the complexity of timetable and multi-commodity flow problems," SIAM Journal on Computing, vol. 5, no. 4, pp. 691-703, Dec. 1976.

- 4. Z. Füredi, "Maximum degree and fractional matchings in uniform hypergraphs,"

 Combinatorica, vol. 1, no. 2, pp. 155-162, 1981.
- 5. Narendra Karmarkar, "A new polynomial-time algorithm for linear programming,"

 Proceedings of the Sixteenth ACM Symposium on Theory of Computing, ACM, New York,

 1984.
- 6. R.M. Karp, "Reducibility among combinatorial problems," in Complexity of Computer Computations, ed. R.N. Miller, J.W. Thatcher, pp. 85-104, Plenum Press, New York, 1972.
- L. Lovász, "On the ratio of optimal and fractional covers," Discrete Mathematics, vol. 13, pp. 383-390, 1975.
- 8. A. Rényi, "Probability Theory," North-Holland Series in Applied Mathematics and Mechanics, vol. 10, North-Holland Publishing Company, Amsterdam, 1970.
- Prabhakar Raghavan and Clark D. Thompson, "Randomized Routing in Gate-Arrays,"
 CSD/84/202, Computer Science Division, UC Berkeley, Sep. 1984.
- 10. Prabhakar Raghavan and Clark D. Thompson, "Provably Good Routing in Graphs: Regular Arrays," Proceedings of the Seventeenth ACM Symposium on Theory of Computing, ACM, New York, May 1985.