The Randomization Repertoire

Rajmohan Rajaraman
Northeastern University, Boston

May 2012
Randomization in Network Optimization

• Very important toolkit:
  – Often simple routines yielding the best known approximations to hard optimization problems
  – Often efficient in practice
• However, no separation known between randomization and determinism
  – Quite likely that “P = RP”
  – Most randomized algorithms have been derandomized
• In distributed computing environments, randomization can provably help
Outline

• Basic tools from probability theory
  – Chernoff-type bounds

• Randomized rounding
  – Set cover
    – Unsplittable multi-commodity flow

• Dependent rounding
  – Multi-path multi-commodity flow

• Randomization in distributed computing
  – Maximal independent set
Basic Probability

• Linearity of expectation: For any random variables $X_1, X_2, ..., X_n$, we have
  $- E[\Sigma_i X_i] = \Sigma_i E[X_i]$

• Markov’s inequality: For any random variable $X$
  $- \Pr[X \geq c] \leq E[X]/c$

• Union bound: For any sequence of events $E_1, E_2, ..., E_n$, we have
  $- \Pr[U_i E_i] \leq \Sigma_i \Pr[E_i]$
Basic Probability: Large Deviations

• Chebyshev inequality: For any random variable $X$ with mean $\mu$ and standard deviation $\sigma$

$$\Pr[|X - \mu| \geq c] \leq \frac{\sigma^2}{c^2}$$

• Applies to any random variable

• Can be used to effectively bound large deviation for sum of pairwise independent random variables
Chernoff Bound

• Let $X_1, X_2, \ldots, X_n$ be $n$ independent random variables in $\{0,1\}$
• For any nonnegative $\delta$
  \[ \Pr[X \geq (1 + \delta)\mu] \leq \left( \frac{e^{\delta}}{(1 + \delta)^{1+\delta}} \right)^{\mu} \]
• For any $\delta$ in $[0,1]$
  \[ \Pr[|X - \mu| \geq \delta\mu] \leq 2e^{-\mu\delta^2/3} \]
Chernoff Bound Extensions

- Chernoff-Hoeffding bounds
  - For real-valued variables
- Bounds extend to negatively correlated variables
- Partial and almost independence [Schmidt-Siegel-Srinivasan 95,...]
  - $k$-wise and almost $k$-wise independence
- Azuma’s inequality for martingales
- [Alon-Spencer 91, Motwani-Raghavan, Mitzenmacher-Upfal 04]
- Matrix Chernoff bounds for random matrices [Ahlswede-Winter 02, Rudelson-Vershynin 07, Tropp 11]
Randomized Rounding

• Write integer program for optimization problem
• Relax integrality constraint: often from \{0,1\} to \[0,1\]
• View relaxed variables as probabilities
• Round the variables to \{0,1\} according to corresponding probability
  – In most basic form, do this independently for each variable
  – Many rounding algorithms are much more sophisticated
  – Formulating the appropriate math program is also often a major contribution
• [Raghavan-Thompson 87]
Randomized Rounding Algorithms

- Multi-commodity flow [Raghavan-Thompson 87]
- MAXSAT [Goemans-Williamson 94]
- Group Steiner tree [Garg-Konjevod-Ravi 98]
- MAXCUT [Goemans-Williamson 95]
Set Cover

• Given a universe $U = \{e_1, e_2, \ldots, e_n\}$ of elements, a collection of $m$ subsets of $U$, and a cost $c(S)$ for each subset $S$, determine the minimum-cost collection of subsets that covers $U$

$$\min \sum_{S} c(S)x(S)$$

$$\sum_{S : e \in S} x(S) \geq 1 \quad \text{for all } e$$

$$x(S) \in \{0, 1\} \quad \text{for all } S$$
Set Cover LP Relaxation

• Given a universe \( U = \{e_1, e_2, \ldots, e_n\} \) of elements, a collection of \( m \) subsets of \( U \), and a cost \( c(S) \) for each subset \( S \), determine the minimum-cost collection of subsets that covers \( U \)

\[
\begin{align*}
\min & \sum_{S} c(S)x(S) \\
\sum_{S : e \in S} x(S) & \geq 1 \quad \text{for all } e \\
x(S) & \in [0,1] \quad \text{for all } S
\end{align*}
\]
Randomized Rounding for Set Cover

• Let OPT be the value of the LP
  – Optimal set cover cost ≥ OPT
• ROUND: Select each S with probability x(S)
• Claim: Expected cost of solution is OPT
  – Linearity of expectation
• Claim: Pr[element f is covered] ≥ 1-1/e
• Repeat ROUND Θ(log(n)) times
  – With probability at least 1-1/n, cost is O(OPT log(n)) and every element is covered
• [Vazirani 03]
Unsplittable Multi-commodity Flow

• Given:
  – Directed graph $G = (V,E)$ with capacity nonnegative capacity $c(f)$ for each edge $f$
  – Demand pairs $(s_i,t_i)$ with demand $d_i$

• Goal: Choose one path for each pair so as to minimize relative congestion

$$\max_{f \in E} \frac{\text{Total demand routed on } f}{c(f)}$$

• Basic randomized rounding achieves near-optimal approximation

• [Raghavan-Thompson 87]
Multi-path Multi-commodity Flow

• Same setup as the original multi-commodity flow problem, except
  – Instead of 1 path, need $r_i$ paths for $i$th pair
• Basic randomized rounding does not ensure the above constraint
• Can refine it to achieve a reasonably good approximation
• Dependent rounding achieves near-optimal result
  [Gandhi-Khuller-Parthasarthy-Srinivasan 05]
Randomization in Distributed Computing

• A number of distributed computing tasks need “local algorithms” to compute globally good solutions
  – Maximal independent set
  – Minimum dominating set

• Easy best-possible greedy algorithms
  – Inherently sequential
  – Place an ordering on the nodes

• How to compute in a distributed setting?
  – At the cost of a small factor, can work with an approximate order
  – Challenge: Break symmetry among competing nodes [Luby 86]