11 LP-Based Approximation

15 July 2025

Prof. Dr. Sebastian Wild

CS627 (Summer 2025)
Philipps-Universität Marburg
version 2025-07-15 00:42 H

Outline

11 LP-Based Approximation

11.1 (Integer) Linear Optimization Recap

LPs in Standard Form

Definition 11.1 (LP)

A linear program (LP) in *standard form* with n *variables* and m *constraints* is characterized by a matrix $\boldsymbol{A} \in \mathbb{Z}^{m \times n}$, a vector $\boldsymbol{b} \in \mathbb{Z}^m$, and a vector $\boldsymbol{c} \in \mathbb{Z}^n$ and is written as

$$\begin{array}{rl} \min & \boldsymbol{c}^T \boldsymbol{x} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b} \\ & \boldsymbol{x} \geq 0 \end{array} \qquad \begin{array}{rll} \min & \sum_{j=1}^n c_j \cdot x_j & \\ \text{s.t.} & \sum_{j=1}^n a_{ij} \cdot x_j \geq b_i & \text{for all } i \in [m] \\ & x_j \geq 0 \quad \text{for all } j \in [n] & \end{array}$$

(Inequalities on vectors apply componentwise.)

Any vector $x \in \mathbb{R}^n$ with $Ax \geq b$ and $x \geq 0$ is called a *feasible solution* for the LP, and $c^T x$ is its objective value. An *optimal solution* is a feasible vector x^* with **min**imal objective value. ◀

Remark 11.2 (Rational coefficients)

We can in general allow $A \in \mathbb{Q}^{m \times x}$, $b \in \mathbb{Q}^m$ and $c \in \mathbb{Q}^n$; by multiplying constraints and scaling objective function with the common denominator we obtain an equivalent LP. ◀

Example LP

$$\begin{aligned} \min \quad & 7x_1 + x_2 + 5x_3 \\ \text{s.t.} \quad & x_1 - x_2 + 3x_3 \geq 10 \\ & 5x_1 + 2x_2 - x_3 \geq 6 \\ & x_1, x_2, x_3 \geq 0 \end{aligned}$$

$\rightsquigarrow$ Optimal solution $x^* = (1.75, 0, 2.75)$ with $c^T x^* = 26$.

Extreme point: feasible point that is *not* a convex combination of two distinct feasible solutions.

Remark 11.3 (Facts on LPs)

1. More general versions of LP possible:
   $=$ constraints, unrestricted variables, max instead of min . . .
   $\rightsquigarrow$ can all be transformed into equivalent one in standard form.
2. LP can be *infeasible* (no solution), *unbounded* (no optimal solution) or *finite*.
3. If LP has optimal solution, there is an optimal extreme point $\rightsquigarrow$ finite problem!
4. Optimal solutions can be computed in polytime (ellipsoid method).

Integer Linear Program in Standard Form

Definition 11.4 (ILP)

An *integer linear program* in standard form is an LP with the additional integrality constraints $x_j \in \mathbb{N}_0$:

$$\begin{aligned} \min \quad & \boldsymbol{c}^T \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b} \\ & \boldsymbol{x} \in \mathbb{N}_0^n \end{aligned}$$

◀

Remark 11.5 (Facts on ILPs)

1. Generalized versions can again be transformed into standard form.
2. Decision version of the problem NP-complete.

◀

11.2 LP Relaxations & Rounding

LP Relaxation Approximations

Since ILPs are NP-complete, any NP problem can be written as an ILP

well, for decision versions . . . but often very natural to write optimization problems as ILP

$\rightsquigarrow$ A natural idea to obtain approximately optimal solutions for NPO problems:

1. Formulate problem as ILP (I)
2. Drop integrality constraints from (I) $\rightsquigarrow$ LP (P)
3. Obtain optimal fractional solution x^* for (P)
4. . . . ?
   Somehow get back to feasible solution for (I)
   Simplest version: Round to nearest integer!

Note: Integrality gap of (I)LP is key barrier in this approach

Set Cover as ILP

The Set Cover ILP

Idea $x_j = 1$ iff S_j in cover.
Notation: For $e \in U = [n]$ set $V(e) = \{j : e \in S_j\}$.

$$\begin{aligned} \min \quad & \sum_{j=1}^{k} c(S_j) \cdot x_j \\ \text{s.t.} \quad & \sum_{j \in V(e)} x_j \ge 1 \quad \forall e \in U \\ & x \in \mathbb{N}_0^k \end{aligned} \qquad \text{(I)}$$

$$\begin{aligned} \min \quad & \sum_{j=1}^{k} c(S_j) \cdot x_j \\ \text{s.t.} \quad & \sum_{j \in V(e)} x_j \ge 1 \quad \forall e \in U \\ & x \ge 0 \end{aligned} \qquad \text{(P)}$$

Observation: Any optimal solution fulfills $x \in \{0,1\}^k$

LP Relaxation: replace $x \in \mathbb{N}_0^k$ by $x \ge 0$.

$\leadsto$ efficiently solvable, but might get fractional solutions x^*.

Write $OPT_{(I)}$ resp. $OPT_{(P)}$ for the optimal objective value $\leadsto$ $OPT_{(I)} \le OPT_{(P)}$

Simple Rounding

```
procedure frequencyCutoffSetCover(n,S,c)
    f := global frequency of S
    x* := optimal solution of relaxed set cover LP.
    𝒞 := ∅
    for j := 1,...,k
        if x*_j ≥ 1/f then add j to 𝒞
    return 𝒞
```

Theorem 11.6

frequencyCutoffSetCover is an f-approximation for SetCover.

Corollary 11.7

frequencyCutoffSetCover is a 2-approximation for VertexCover.

Proof:

(1) $\mathcal{C}$ is a set cover

Let $e \in U$ be arbitrary. Since x^* is feasible, we have $\sum_{j \in V(e)} \geq 1$.

$|V(e)| = f_e \leq f \quad \rightsquigarrow \quad$ one x_j with $j \in V(e)$ must be $x_j \geq 1/f$.

$\rightsquigarrow \quad j \in \mathcal{C}$ and e is covered.

(2) f-approximation.

x^* optimal for $(P) \quad \rightsquigarrow \quad c^T x^* = OPT_{(P)} \overset{\text{min-problem}}{\leq} OPT_{(I)}$

Simple Rounding [2]

Proof (cont.):

For every $j \in \mathcal{C}$, $x_j^* \geq 1/f$.

$$\rightsquigarrow \quad c(\mathcal{C}) = \sum_{j \in \mathcal{C}} c(S_j) \leq \sum_{j \in \mathcal{C}} f \cdot x_j^* \cdot c(S_j) = f \cdot \sum_{j \in \mathcal{C}} \cdot x_j^* \cdot c(S_j) \leq f \cdot \sum_{j \in [k]} \cdot x_j^* \cdot c(S_j) = f \cdot OPT_{(P)} \leq f \cdot OPT_{(I)}$$

■

Simple Rounding – Analysis is tight

In the worst case, the above threshold method cannot be better than an f-approximation

Consider the **“Fully Symmetric instance:”**

Suppose $f \mid n$
$U = [0..n)$ with $S_j = \{j, j+1, \ldots, j+f-1\} \bmod n$, for all $j \in [0..n)$
All sets of equal cost, $c(S_j) = 1$

$\rightsquigarrow$ n/f sets suffice;
but $x^* = (\frac{1}{f}, \ldots, \frac{1}{f})$ is optimal for (P) $\rightsquigarrow$ frequencyCutoffSetCover outputs $\mathcal{C} = [0..n)$

11.3 Randomized Rounding

Fractions as probabilities

Another intuitive use of fractional solutions $x_j^* \in (0,1)$: include S_j with probability x_j^* in $\mathcal{C}$

$$\rightsquigarrow \quad \mathbb{E}[c(\mathcal{C})] \;=\; \sum_{j=1}^{k} x_j^* \cdot c(S_j) \;=\; OPT_{(P)} \text{ (!)}$$

Too good to be true? Yeah, mostly not a feasible solution.

But the idea can be rescued.

Intuition: If e occurs in f_e sets, we have

$$\mathbb{P}[e \text{ covered}] \;=\; 1 - \mathbb{P}\left[\bigcap_{j \in V(e)} S_j \notin \mathcal{C}\right] \;=\; 1 - \prod_{j \in V(e)} \left(1 - x_j^*\right) \;\geq\; 1 - \left(1 - \frac{1}{f_e}\right)^{f_e} \;\geq\; 1 - \frac{1}{e}$$

$\rightsquigarrow$ Coupon collector with n coupons $\rightsquigarrow$ $\approx H_n$ repetitions

Assuming we keep trying and collect all sets ever chosen

Probably not better than greedy in worst case, but technique is general & tweakable

Randomized Rounding

```
procedure randomizedRoundingSet(n, S, c, r)
    x* := optimal solution of relaxed set cover LP.
    for i := 1, ..., r
        C_i := ∅
        for j := 1, ..., k
            b := coin flip with prob x*_j
            if b == 1 then C_i := C_i ∪ {j}
    return C := ⋃_{i=1}^r C_i
```

For simplicity, always set $\boxed{r = \lceil \ln(4n) \rceil}$

Lemma 11.8

randomizedRoundingSet computes a feasible set-cover with probability $\ge \frac{3}{4}$.

Proof:

Recall from calculation above that for $e \in U$ and a single iteration of the outer loop:

$$\mathbb{P}[e \text{ not covered by } \mathcal{C}_i] \le \left(1 - \frac{1}{f_e}\right)^{f_e} \le \frac{1}{e}$$

$$\rightsquigarrow \quad \mathbb{P}[e \text{ not covered by } \mathcal{C}] = \prod_{i=1}^{r} \mathbb{P}[e \text{ not covered by } \mathcal{C}_i] \le \left(\frac{1}{e}\right)^r$$

With the union bound over all n elements and $r = \ln(4n)$, we obtain

$\mathbb{P}[\mathcal{C} \text{ not a set cover}] \le n e^{-r} = \frac{1}{4}$. ∎

Randomized Rounding – Analysis

Lemma 11.9 (Expected quality)

Let $\mathcal{C}$ by computed by randomizedRoundingSet with r repetitions.
The *expected* cost are $\mathbb{E}[c(\mathcal{C})] \le r \cdot OPT_{(P)}$.

$\rightsquigarrow$ For $r = \ln(4n)$ we have by Markov's inequality: $\mathbb{P}\left[c(\mathcal{C}) \ge 4\ln(4n) \cdot OPT_{(P)}\right] \le \frac{1}{4}$

Proof:
We choose $\mathcal{C} = \mathcal{C}_1 \cup \cdots \cup \mathcal{C}_r$.
For the cost we get

$$\mathbb{E}[c(\mathcal{C})] \le \mathbb{E}\left[\sum_{i=1}^{r} c(\mathcal{C}_i)\right] = \sum_{i=1}^{r} \mathbb{E}[c(\mathcal{C}_i)] = r \cdot OPT_{(P)}$$

■

Randomized Rounding Approximation for Set Cover

```
procedure randomizedRoundingSetCover(n, S, c)
    𝒞 = randomizedRoundingSet(n, S, c, ⌈ln(4n)⌉)
    if 𝒞 is a set cover
        return 𝒞
    else
        return S
```

Theorem 11.10 (randomizedRoundingSetCover randomized approx)

randomizedRoundingSetCover is a randomized $4\ln(4n)$-approximation for SetCover.

Proof:

$$\begin{aligned}\mathbb{P}[\mathcal{C} \text{ not SC} \;\vee\; c(\mathcal{C}) > 4\ln(4n)\cdot OPT_{(P)}] &\le \mathbb{P}[\mathcal{C} \text{ not SC}] + \mathbb{P}[c(\mathcal{C}) > 4\ln(4n)\cdot OPT_{(P)}] \\ &\underset{\text{Lemma 11.8, Lemma 11.9}}{\le} \frac{1}{4} + \frac{1}{4} \\ &= \frac{1}{2}.\end{aligned}$$

■

11.4 LP Duality

Bounding optimal values of LPs

Starting with an original ("primal") LP, how can we bound on its optimal objective value?

$$\begin{array}{ll} \min & 7x_1 + x_2 + 5x_3 \\ \text{s.t.} & x_1 - x_2 + 3x_3 \geq 10 \\ & 5x_1 + 2x_2 - x_3 \geq 6 \\ & x_1, x_2, x_3 \geq 0 \end{array}$$

Optimal solution:
$x^* = (1.75, 0, 2.75)$ with $c^T x^* = 26$.

Dual LPs

$$\begin{aligned} \min\ & c^T x \\ \text{s.t.}\ & Ax \geq b \qquad (P) \\ & x \geq 0 \end{aligned} \qquad\qquad \begin{aligned} \max\ & b^T y \\ \text{s.t.}\ & A^T y \leq c \qquad (D) \\ & y \geq 0 \end{aligned}$$

Generalizations:

- ith constraint in primal with '$\geq$' $\leftrightsquigarrow$ $y_i \geq 0$
- ith constraint in primal with '$=$' $\leftrightsquigarrow$ y_i unconstrained

Lemma 11.11 (Weak Duality)

If x and y are *feasible* solutions for the primal resp. dual LP, it holds that $\boxed{c^T x \geq b^T y.}$

Proof:

Dual constraint $A^T y \leq c$ implies $c^T \geq (A^T y)^T = y^T A$.

$$\rightsquigarrow \quad c^T x \;\geq\; (y^T A)x \;=\; y^T(Ax) \underset{\text{prim. constr.}}{\geq} y^T b \;=\; b^T y \qquad \blacksquare$$

Duality Theory

Indeed, one can show by a closer study that the optimal objective values *always coincide*.

Theorem 11.12 (Strong duality)

The primal LP has a finite optimal objective if and only if the dual has. If x^* resp. y^* are two optimal solutions to the primal resp. dual LP then $c^T x^* = b^T y^*$ holds. ◀

Theorem 11.13 (Complementary Slackness Conditions (CSC))

Let x and y be feasible solutions to the primal and dual LP.
The pair (x, y) is optimal *if and only if*

1. $\forall j \in [n] \ : \ x_j = 0 \ \vee \ \sum_{1 \le i \le m} a_{i,j} \cdot y_i = c_j$ and
2. $\forall i \in [m] \ : \ y_i = 0 \ \vee \ \sum_{1 \le j \le n} a_{i,j} \cdot x_j = b_i$. ◀

Remark 11.14

1. Strong duality implies that the LP threshold decision problem is in NP ∩ co-NP:
   Yes-certificate: feasible solution; No-certificate: feasible solution *for the dual*.
   (We know it actually lies in P)
2. For ILPs, we only get weak duality. ◀

11.5 Vertex Cover and Matching Revisited

Vertex Cover & Maximum Matching

Vertex Cover

$$\min \sum_{v\in V} x_v$$

$$\text{s. t. } x_v + x_w \geq 1 \quad \forall vw \in E$$

$$x_v \in \{0,1\} \quad \forall v \in V$$

Maximum Matching

$$\max \sum_{e\in E} y_e$$

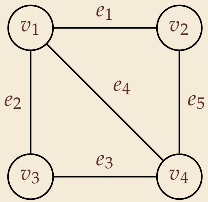
$$\text{s. t. } \sum_{vw\in E} y_{vw} \leq 1 \quad \forall v \in V$$

$$y_e \in \{0,1\} \quad \forall e \in E$$

⇝ Consider the LP relaxations

Vertex Cover & Maximum Matching – Example

Graph G



Minimum Vertex Cover

$$
\begin{array}{ll}
\min & x_1 + x_2 + x_3 + x_4 \\
\text{s.t.} & x_1 + x_2 \geq 1 \\
& x_1 + x_3 \geq 1 \\
& x_3 + x_4 \geq 1 \\
& x_1 + x_4 \geq 1 \\
& x_2 + x_4 \geq 1 \\
& x_1, x_2, x_3, x_4 \geq 0
\end{array}
$$

$$
\begin{pmatrix}
1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 \\
0 & 0 & 1 & 1 \\
1 & 0 & 0 & 1 \\
0 & 1 & 0 & 1
\end{pmatrix}
$$

Maximum Matching

$$
\begin{array}{ll}
\max & y_1 + y_2 + y_3 + y_4 + y_5 \\
\text{s.t.} & y_1 + y_2 + y_4 \leq 1 \\
& y_1 + y_5 \leq 1 \\
& y_2 + y_3 \leq 1 \\
& y_3 + y_4 + y_5 \leq 1 \\
& y_1, y_2, y_3, y_4, y_5 \geq 0
\end{array}
$$

$$
\begin{pmatrix}
1 & 1 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 \\
0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 1
\end{pmatrix}
$$

incidence matrix of G!

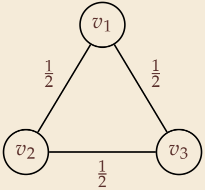
Vertex Cover & Maximum Matching – Dual Problems

Problems are *dual!*

- ⇝ Our earlier lemma "VC $\geq$ M" is just weak duality (on the ILPs)

- ⇝ Can generally try to build approximation algorithm by constructing pair of primally/dually feasible solutions

Note: Dual **LPs** have **equal** optimal objective value;
For dual **ILPs**, can have a ***duality gap***



- ⇝ For VertexCover/MaximumMatching, duality gap is 2.

Bipartite Graphs

Except for **bipartite graphs!**
Bipartite graph: $V(G) = L \dot{\cup} R$, $E(G) \subset L \times R$

Known:

every square submatrix has determinant 0, 1, or −1

- incidence matrix A of bipartite G is a *totally unimodular (TU)* matrix
- A TU $\rightsquigarrow$ LPs $\min\{c^T x : Ax \ge b, x \ge 0\}$ and $\max\{b^T y : A^T y \le c, y \ge 0\}$
  with integral b and c have **integral** optimal solutions x^* and y^*

$\rightsquigarrow$ No integrality gap and no duality gap!

Here, also easy to see directly:

- Maximum matching in bipartite graph must have one side (L or R) completely matched

$\rightsquigarrow$ Taking all of these vertices must be a VC

11.6 Set Cover Duality & Dual Fitting

Dual Fitting

Dual fitting uses (I)LPs for a minimization problem as follows:

- Simple algorithm maintains primally feasible and **integral** $c^T x$.
- In the analysis, we show that cost of x ist at most the cost of an implicitly computed (nonintegral) dual y.
  However, y is not in general dually feasible.
- By *scaling* y down by a factor $\delta > 1$, we obtain a feasible dual solution, with cost a factor δ larger

$\rightsquigarrow$ $c^T x \leq \delta \cdot OPT$

Set Cover LP and its dual

Recall: Input: $S = (S_1, \ldots, S_k)$ over universe U; define $V(e) = \{j : e \in S_j\}$.

$$\begin{aligned} \min \quad & \sum_{j=1}^{k} c(S_j) \cdot x_j \\ \text{s. t.} \quad & \sum_{j \in V(e)} x_j \geq 1 \quad \forall e \in U \\ & \boldsymbol{x} \geq \boldsymbol{0} \end{aligned}$$

$$\begin{aligned} \max \quad & \sum_{e \in U} y_e \\ \text{s. t.} \quad & \sum_{e \in S_j} y_e \leq c(S_j) \quad \forall j \in [k] \\ & \boldsymbol{y} \geq \boldsymbol{0} \end{aligned}$$

Intuition:
Pack as much (y_u) of good u as possible, so that for group S_j its capacity $c(S_j)$ is exceeded.

Analysis of greedySetCover by dual fitting

Recall greedySetCover from Unit 10:

```
procedure greedySetCover(n, 𝒮, c)
    𝒞 := ∅;  C := ∅
    // For analysis: i := 1
    while C ≠ [n]
        i* := arg min_{i∈[n]} c(S_i) / |S_i \ C|
        𝒞 := 𝒞 ∪ {i*}
        C := C ∪ S_{i*}
        // For analysis only:
        // α_i := c(S_{i*}) / |S_{i*} \ C|
        // for e ∈ S_{i*} \ C set price(e) := α_i
        // i := i + 1
    return 𝒞
```

Lemma 11.15

$y_e = price(e)/H_n$ is dual-feasible. ◀

Proof:

$price(e)$ essentially dual variable, but not directly feasible. (Recall $\sum_{e\in U} price(e) = c(\mathcal{C})$).

Consider the dual constraint for S_j:

$\sum_{e\in S_j} y_e \le c(S_j)$. Write $\ell = |S_j|$.

Let $e_1, \ldots, e_n$ be elements in order as covered by algorithm.

When e_i covered, S_j contains $\ge \ell - (i-1)$ *uncovered* elements.

$\rightsquigarrow$ S_j covers e_i at price $\le \dfrac{c(S_j)}{\ell - i + 1}$ per element.

$\rightsquigarrow$ $price(e_i) \le \dfrac{c(S_j)}{\ell - i + 1}$ $\rightsquigarrow$ $y_{e_i} \le \dfrac{1}{H_n}\dfrac{c(S_j)}{\ell - i + 1}$

Analysis of greedySetCover by dual fitting [2]

Proof (cont.):

Consider dual constraint for S_j:

$$\sum_{e \in S_j} y_e \;=\; \sum_{m=1}^{\ell} y_{e_{i_m}} \;\le\; \frac{c(S_j)}{H_n} \sum_{m=1}^{\ell} \frac{1}{m} \;=\; \frac{H_\ell}{H_n} c(S_j) \;\le\; c(S_j)$$

$$\rightsquigarrow \quad c(\mathcal{C}) \;\le\; H_n \cdot OPT_{(D)} \;=\; H_n \cdot OPT_{(P)}.$$

Also note: actually suffices to scale by H_ℓ for $\ell = \max |S_j|$.

Integrality Gap of Set Cover

Previous result shows that integrality gap $\frac{OPT}{OPT_{(P)}} \leq H_n$.

Can we give a lower bound?

Theorem 11.16 (Integrality Gap of Set Cover)

For the set cover ILP and its relaxation holds

$$\frac{OPT}{OPT_{(P)}} \geq \frac{\log_2(n+1)}{2\frac{n}{n+1}} \sim \frac{1}{2\ln 2}H_n \approx 0.721 H_n$$

◀

11.7 The Primal-Dual Schema

Primal-Dual Schema

So far:

- ad hoc methods, a posteriori justified by LP arguments
- rounding algorithms, must solve primal LP to optimality (expensive!)

Can we use duality more directly?

CSC for set cover

Complementary Slackness Conditions for Set Cover

$$x_j = 0 \lor \sum_{u \in S_j} y_u = c(S_j) \qquad \forall j \in [k]$$

$$y_u = 0 \lor \sum_{j \in V(u)} x_j = 1 \qquad \forall u \in U$$

Problem: In general only simultaneously fulfilled by fractional solutions
Relax dual constraints to

$$y_u = 0 \lor \sum_{j \in V(u)} x_j \leq f \qquad \forall u \in U$$

i. e., every element at most f times $\rightsquigarrow$ trivially fulfilled.

Primal Dual Set Cover

procedure primalDualSetCover(n,S,c)
  f = global frequency
  $\boldsymbol{x} = \mathbf{0}$, $\boldsymbol{y} = \mathbf{0}$, $T = [n]$
  while $T \neq \emptyset$
    Choose $u \in T$ arbitrary
    Increase y_u until CSC holds **for** (at least) one more set S_j
    for all S_j with $\sum_{u \in S_j} y_u = c(S_j)$
      $T = T \setminus S_j$
      $x_j = 1$
  return $\{j \in [k] : x_j = 1\}$

Theorem 11.17

`primalDualSetCover` is an f-approximation for SetCover. ◀