Easier derivation of bounded pitch inequalities for set covering problems +

Daniel Bienstock

Columbia University

Tel Aviv April 2018

Set covering

Set covering

$$\begin{aligned} & \min \; c^T x \\ \text{s.t.} \quad & Ax \geq \mathbf{e}, \quad x \text{ binary} \end{aligned}$$

A is a $0/1$ matrix, $\mathbf{e} = (1, \ldots, 1)^T$

Set covering

$$\begin{array}{ll} & \min \ c^T x \\ \text{s.t.} & Ax \geq \mathbf{e}, \quad x \ \text{binary} \end{array}$$

A is a $0/1$ matrix, $\mathbf{e} = (1, \ldots, 1)^T$

Starting point: Balas and Ng (1989), *All facets with coefficients 0,1,2*

Set covering

$$\begin{array}{ll} & \min \; c^T x \\ \text{s.t.} & Ax \geq \mathbf{e}, \quad x \;\text{binary} \end{array}$$

A is a $0/1$ matrix, $\mathbf{e} = (1, \ldots, 1)^T$

Starting point: Balas and Ng (1989), *All facets with coefficients 0,1,2*

$\rightarrow$ There are examples with exponentially many such facets

Set covering

$$\begin{aligned} \min \;\; & c^T x \\ \text{s.t.} \;\; & Ax \geq \mathbf{e}, \quad x \text{ binary} \end{aligned}$$

A is a $0/1$ matrix, $\mathbf{e} = (1, \ldots, 1)^T$

Starting point: Balas and Ng (1989), *All facets with coefficients 0,1,2*

$\rightarrow$ There are examples with exponentially many such facets

Can we account for all valid inequalities with small coefficients?

Theorem (B. and Mark Zuckerberg, 2004)

For any fixed integer $\boldsymbol{k} \geq \mathbf{1}$ there exists a *compact, extended* formulation whose solutions satisfy all valid inequalities with coefficients in $\{\mathbf{0}, \mathbf{1}, \ldots, \boldsymbol{k}\}$.

"compact:" of polynomial size (for fixed k)

"extended:" uses additional variables, a **lifted** formulation

Theorem (B. and Mark Zuckerberg, 2004)

For any fixed integer $k \geq 1$ there exists a *compact, extended* formulation whose solutions satisfy all valid inequalities with coefficients in $\{0, 1, \ldots, k\}$.

"compact:" of polynomial size (for fixed k)

"extended:" uses additional variables, a **lifted** formulation

Definition: An inequality $\alpha^T x \geq b$ for **valid** has **pitch** $\leq k$ if:

the sum of the smallest k positive α_j is at least b

Theorem (B. and Mark Zuckerberg, 2004)

For any fixed integer $k \geq 1$ there exists a *compact, extended* formulation whose solutions satisfy all valid inequalities with coefficients in $\{0, 1, \ldots, k\}$.

"compact:" of polynomial size (for fixed k)

"extended:" uses additional variables, a **lifted** formulation

Definition: An inequality $\alpha^T x \geq b$ for **valid** has **pitch** $\leq k$ if:

the sum of the smallest k positive α_j is at least b

Hence, inequalities with coefficients in $\{0, 1, \ldots, k\}$ have pitch $\leq k$

Theorem (B. and Mark Zuckerberg, 2004)

For any fixed integer $\boldsymbol{k} \geq \mathbf{1}$ there exists a *compact, extended* formulation whose solutions satisfy all valid inequalities with pitch $\leq \boldsymbol{k}$.

"compact:" of polynomial size (for fixed k)

"extended:" uses additional variables, a **lifted** formulation

Theorem (B. and Mark Zuckerberg, 2004)

For any fixed integer $\boldsymbol{k} \geq \mathbf{1}$ there exists a *compact, extended* formulation whose solutions satisfy all valid inequalities with pitch $\leq \boldsymbol{k}$.

"compact:" of polynomial size (for fixed k)

"extended:" uses additional variables, a **lifted** formulation

Definition: An inequality $\alpha^T \boldsymbol{x} \geq \boldsymbol{b}$ for **valid** has **pitch** $\leq \boldsymbol{k}$ if:

the sum of the smallest $\boldsymbol{k}$ positive α_j is at least $\boldsymbol{b}$

Theorem (B. and Mark Zuckerberg, 2004)

For any fixed integer $\boldsymbol{k} \geq \mathbf{1}$ there exists a *compact, extended* formulation whose solutions satisfy all valid inequalities with pitch $\leq \boldsymbol{k}$.

"compact:" of polynomial size (for fixed k)

"extended:" uses additional variables, a **lifted** formulation

Definition: An inequality $\alpha^T \boldsymbol{x} \geq \boldsymbol{b}$ for **valid** has **pitch** $\leq \boldsymbol{k}$ if:

the sum of the smallest $\boldsymbol{k}$ positive α_j is at least $\boldsymbol{b}$

Hence, inequalities with coefficients in $\{\mathbf{0}, \mathbf{1}, \ldots, \boldsymbol{k}\}$ have pitch $\leq \boldsymbol{k}$

Theorem (B. and Mark Zuckerberg, 2004)

For any fixed integer $\boldsymbol{k} \geq \mathbf{1}$ there exists a *compact, extended* formulation whose solutions satisfy all valid inequalities with pitch $\leq \boldsymbol{k}$.

"compact:" of polynomial size (for fixed k)

"extended:" uses additional variables, a **lifted** formulation

Theorem (B. and Mark Zuckerberg, 2004)

For any fixed integer $\boldsymbol{k} \geq \mathbf{1}$ there exists a *compact, extended* formulation whose solutions satisfy all valid inequalities with pitch $\leq \boldsymbol{k}$.

"compact:" of polynomial size (for fixed k)

"extended:" uses additional variables, a **lifted** formulation

Corollary: For any fixed positive integer $\boldsymbol{r} \geq \mathbf{1}$ and $\mathbf{0} < \boldsymbol{\epsilon} < \mathbf{1}$,

there is a compact extended formulation for set-covering whose solutions satisfy the **rank-r Gomory** closure within multiplicative error $\boldsymbol{\epsilon}$

Theorem (B. and Mark Zuckerberg, 2004)

For any fixed integer $\boldsymbol{k} \geq \mathbf{1}$ there exists a *compact, extended* formulation whose solutions satisfy all valid inequalities with pitch $\leq \boldsymbol{k}$.

"compact:" of polynomial size (for fixed k)

"extended:" uses additional variables, a **lifted** formulation

Corollary: For any fixed positive integer $\boldsymbol{r} \geq \mathbf{1}$ and $\mathbf{0} < \boldsymbol{\epsilon} < \mathbf{1}$,

there is a compact extended formulation for set-covering whose solutions satisfy the **rank-r Gomory** closure within multiplicative error $\boldsymbol{\epsilon}$

$\forall c \in \mathbb{R}^n$:

$$\min c^T x \quad \text{s.t. } x \in \text{projected formulation} \geq$$
$$(1-\epsilon)\left(\min c^T x \quad \text{s.t. } x \in \text{rank-r Gomory closure}\right)$$

Two recent, related papers:

- M. Mastrolilli (sum-of-squares mod 2)
- S. Fiorini, T. Huynh and S. Weltge (circuit complexity)

Two recent, related papers:

- M. Mastrolilli (sum-of-squares mod 2)
- S. Fiorini, T. Huynh and S. Weltge (circuit complexity)
- They point out that the B-Z formulation is 'complex'

Two recent, related papers:

- M. Mastrolilli (sum-of-squares mod 2)
- S. Fiorini, T. Huynh and S. Weltge (circuit complexity)
- They point out that the B-Z formulation is 'complex'

- Today, a shorter proof +

Vector Branching (from Z's PhD thesis)

Consider a (known) valid inequality

$$\sum_{j \in S} a_j x_j \;\geq\; a_0 \quad (\, > 0\,)$$

for a **binary** optimization problem.

Vector Branching (from Z's PhD thesis)

Consider a (known) valid inequality

$$\sum_{j \in S} a_j x_j \;\geq\; a_0 \quad (\,> 0\,)$$

for a **binary** optimization problem.
Let $S \;=\; \{j_1, j_2, \ldots, j_t\}$.

Vector Branching (from Z's PhD thesis)

Consider a (known) valid inequality

$$\sum_{j \in S} a_j x_j \;\geq\; a_0 \quad (\,>0\,)$$

for a **binary** optimization problem.
Let $S \;=\; \{j_1, j_2, \ldots, j_t\}$. Then

- $\boldsymbol{x_{j_1} = 1}$, **or**

Vector Branching (from Z's PhD thesis)

Consider a (known) valid inequality

$$\sum_{j \in S} a_j x_j \;\geq\; a_0 \quad (\,> 0\,)$$

for a **binary** optimization problem.
Let $S \;=\; \{j_1, j_2, \ldots, j_t\}$. Then

- $\boldsymbol{x_{j_1} = 1}$, **or**
- $x_{j_1} = 0$ **and** $\boldsymbol{x_{j_2} = 1}$, **or**

Vector Branching (from Z's PhD thesis)

Consider a (known) valid inequality

$$\sum_{j \in S} a_j x_j \;\geq\; a_0 \quad (>0)$$

for a **binary** optimization problem.
Let $S \;=\; \{j_1, j_2, \ldots, j_t\}$. Then

- $\boldsymbol{x_{j_1} = 1}$, **or**
- $x_{j_1} = 0$ **and** $\boldsymbol{x_{j_2} = 1}$, **or**
- $x_{j_1} = x_{j_2} = 0$ **and** $\boldsymbol{x_{j_3} = 1}$, **or**

Vector Branching (from Z's PhD thesis)

Consider a (known) valid inequality

$$\sum_{j \in S} a_j x_j \;\geq\; a_0 \quad (\, > 0\,)$$

for a **binary** optimization problem.
Let $S = \{j_1, j_2, \ldots, j_t\}$. Then

- $\boldsymbol{x_{j_1} = 1}$, **or**
- $x_{j_1} = 0$ **and** $\boldsymbol{x_{j_2} = 1}$, **or**
- $x_{j_1} = x_{j_2} = 0$ **and** $\boldsymbol{x_{j_3} = 1}$, **or**
- $\ldots$
- $x_{j_1} = \ldots = x_{j_{t-1}} = 0$ **and** $\boldsymbol{x_{j_t} = 1}$,

is a **valid** disjunction

Vector Branching (from Z's PhD thesis)

Consider a (known) valid inequality

$$\sum_{j \in S} a_j x_j \;\geq\; a_0 \quad (\,>0\,)$$

for a **binary** optimization problem.
Let $S \;=\; \{j_1, j_2, \ldots, j_t\}$. Then

- $\boldsymbol{x_{j_1} = 1}$, **or**
- $x_{j_1} = 0$ **and** $\boldsymbol{x_{j_2} = 1}$, **or**
- $x_{j_1} = x_{j_2} = 0$ **and** $\boldsymbol{x_{j_3} = 1}$, **or**
- $\ldots$
- $x_{j_1} = \ldots = x_{j_{t-1}} = 0$ **and** $\boldsymbol{x_{j_t} = 1}$,

is a **valid** disjunction
Gives rise to an alternate scheme for branch-and-bound

Theorem

Given a set-covering problem, suppose we apply vector branching to a given constraint

$$\sum_{j \in H} x_j \;\geq\; 1$$

Theorem

Given a set-covering problem, suppose we apply vector branching to a given constraint

$$\sum_{j \in H} x_j \;\geq\; 1$$

Then, the solution to any **node** of the branch-and-bound (sub)tree thus created satisfies every valid inequality

$$\boldsymbol{\alpha}^{\boldsymbol{T}} \boldsymbol{x} \;\geq\; \mathbf{2}$$

where

- $\alpha_j \;\in\; \{0, 1, 2\}$ for $j = 1, \ldots, n$
- H contained in the **support** of α

Example

Consider a valid inequality

$$\sum_{j \in S} x_j \;\geq\; 2 \tag{1}$$

and suppose we vector-branch on a set covering constraint

$$\sum_{j \in H} x_j \;\geq\; 1, \qquad \text{with } H \subseteq S$$

And now consider a node where $\boldsymbol{x_{j_t} = 1}$ with $j_t \in H$. **But**:

Example

Consider a valid inequality

$$\sum_{j \in S} x_j \;\geq\; 2 \tag{1}$$

and suppose we vector-branch on a set covering constraint

$$\sum_{j \in H} x_j \;\geq\; 1, \qquad \text{with } H \subseteq S$$

And now consider a node where $\boldsymbol{x_{j_t} = 1}$ with $j_t \in H$. **But**:
Since (1) is valid, **so is**:

$$\textstyle\sum_{j \in S - j_t} x_j \;\geq\; 1 \tag{2}$$

But, set-covering,

Example

Consider a valid inequality

$$\sum_{j \in S} x_j \;\geq\; 2 \tag{1}$$

and suppose we vector-branch on a set covering constraint

$$\sum_{j \in H} x_j \;\geq\; 1, \qquad \text{with } H \subseteq S$$

And now consider a node where $\boldsymbol{x_{j_t} = 1}$ with $j_t \in H$. **But**:
Since (1) is valid, **so is**:

$$\textstyle\sum_{j \in S - j_t} x_j \;\geq\; 1 \tag{2}$$

But, set-covering, so (2) **must** be implied by a set-covering constraint.

Example

Consider a valid inequality

$$\sum_{j \in S} x_j \ \geq \ 2 \tag{1}$$

and suppose we vector-branch on a set covering constraint

$$\sum_{j \in H} x_j \ \geq \ 1, \qquad \text{with } H \subseteq S$$

And now consider a node where $\boldsymbol{x_{j_t} = 1}$ with $j_t \in H$. **But**:
Since (1) is valid, **so is**:

$$\textstyle\sum_{j \in S - j_t} x_j \ \geq \ 1 \tag{2}$$

But, set-covering, so (2) **must** be implied by a set-covering constraint.
So the solution to the node must satisfy (1).

Example

Consider a valid inequality

$$\sum_{j \in S} x_j \;\geq\; 2 \tag{1}$$

and suppose we vector-branch on a set covering constraint

$$\sum_{j \in H} x_j \;\geq\; 1, \qquad \text{with } H \subseteq S$$

And now consider a node where $\boldsymbol{x_{j_t} = 1}$ with $j_t \in H$. **But**:
Since (1) is valid, **so is**:

$$\textstyle\sum_{j \in S - j_t} x_j \;\geq\; 1 \tag{2}$$

But, set-covering, so (2) **must** be implied by a set-covering constraint.
So the solution to the node must satisfy (1). **Related: Letchford 2001**

Pitch k

Consider a valid inequality of pitch k:

$$\sum_{j \in S} \alpha_j x_j \;\geq\; \alpha_0 \qquad (3)$$

and suppose we vector-branch on a set covering constraint

$$\sum_{j \in H} x_j \;\geq\; 1, \qquad \text{with } H \subseteq S$$

And now consider a node where $\boldsymbol{x_{j_t} = 1}$ with $j_t \in H$. **But**:

Pitch k

Consider a valid inequality of pitch k:

$$\sum_{j \in S} \alpha_j x_j \ \geq \ \alpha_0 \tag{3}$$

and suppose we vector-branch on a set covering constraint

$$\sum_{j \in H} x_j \ \geq \ 1, \qquad \text{with } H \subseteq S$$

And now consider a node where $\mathbf{x_{j_t} = 1}$ with $j_t \in H$. **But**:
Since (3) is valid, **so is**:

$$\textstyle\sum_{j \in S - j_t} \alpha_j x_j \ \geq \ \alpha_0 - \alpha_{j_t} \tag{4}$$

But,

Pitch k

Consider a valid inequality of pitch k:

$$\sum_{j \in S} \alpha_j x_j \;\geq\; \alpha_0 \tag{3}$$

and suppose we vector-branch on a set covering constraint

$$\sum_{j \in H} x_j \;\geq\; 1, \qquad \text{with } H \subseteq S$$

And now consider a node where $\boldsymbol{x_{j_t} = 1}$ with $j_t \in H$. **But**:
Since (3) is valid, **so is**:

$$\textstyle\sum_{j \in S - j_t} \alpha_j x_j \;\geq\; \alpha_0 - \alpha_{j_t} \tag{4}$$

But, (4) has **pitch** $\leq \boldsymbol{k - 1}$

Pitch k

Consider a valid inequality of pitch k:

$$\sum_{j \in S} \alpha_j x_j \;\geq\; \alpha_0 \qquad (3)$$

and suppose we vector-branch on a set covering constraint

$$\sum_{j \in H} x_j \;\geq\; 1, \qquad \text{with } H \subseteq S$$

And now consider a node where $x_{j_t} = 1$ with $j_t \in H$. **But**:
Since (3) is valid, **so is**:

$$\sum_{j \in S - j_t} \alpha_j x_j \;\geq\; \alpha_0 - \alpha_{j_t} \qquad (4)$$

But, (4) has **pitch** $\leq k - 1$
So all we need is a **recursive** construction

Construction

Construction – a few corners are cut

- Set-covering system $Ax \geq \boldsymbol{e}$.
- Pitch $\boldsymbol{p} \geq \mathbf{2}$
- $\mathscr{Z}^{\boldsymbol{p}-\mathbf{1}}$: recursively constructed formulation whose solutions satisfy all valid inequalities of pitch $\leq \boldsymbol{p} - \mathbf{1}$.
- For $\boldsymbol{p} = \mathbf{2}$,

Construction – a few corners are cut

- Set-covering system $Ax \geq \boldsymbol{e}$.
- Pitch $\boldsymbol{p} \geq \mathbf{2}$
- $\mathcal{Z}^{\boldsymbol{p}-\mathbf{1}}$: recursively constructed formulation whose solutions satisfy all valid inequalities of pitch $\leq \boldsymbol{p} - \mathbf{1}$.
- For $\boldsymbol{p} = \mathbf{2}$, $\mathcal{Z}^{\boldsymbol{p}-\mathbf{1}}$ is the original formulation $\boldsymbol{Ax} \geq \boldsymbol{e}$
- Now we will consider a row $\boldsymbol{i}$ of $\boldsymbol{Ax} \geq \boldsymbol{e}$ and, effectively, vector-branch on it
- Actually we will write the corresponding **disjunction**

Let the row be

$$\sum_{j \in S^i} x_j \ \geq \ 1$$

where $S^i \ = \ \{j_1, j_2, \ldots, j_{|S^i|}\}$.

Row $\boldsymbol{i}$ of $\boldsymbol{Ax} \geq \boldsymbol{e}$: $\sum_{j \in S^i} x_j \geq 1$, where $S^i = \{j_1, \ldots, j_{|S^i|}\}$.

(a) For $\mathbf{1} \leq \boldsymbol{t} \leq |\boldsymbol{S^i}|$, polyhedron $D_i^p(t) \subseteq \mathbb{R}^n$ given by

$$x_{j_t} = 1 \tag{5}$$

$$x_{j_h} = 0 \quad \forall\, 1 \leq h < t, \quad \text{and} \tag{6}$$

$$x \in \mathcal{Z}^{p-1} \tag{7}$$

(b) Polyhedron $\boldsymbol{D_i^p} \doteq \mathrm{conv}\{\boldsymbol{D_i^p(t)} : \mathbf{1} \leq \boldsymbol{t} \leq |\boldsymbol{S^i}|\}$

Row $\boldsymbol{i}$ of $\boldsymbol{Ax} \geq \boldsymbol{e}$: $\sum_{j \in S^i} x_j \geq 1$, where $S^i = \{j_1, \ldots, j_{|S^i|}\}$.

(a) For $\mathbf{1} \leq \boldsymbol{t} \leq |\boldsymbol{S^i}|$, polyhedron $D_i^p(t) \subseteq \mathbb{R}^n$ given by

$$x_{j_t} = 1 \tag{5}$$

$$x_{j_h} = 0 \quad \forall\, 1 \leq h < t, \quad \text{and} \tag{6}$$

$$x \in \mathcal{Z}^{p-1} \tag{7}$$

(b) Polyhedron $\boldsymbol{D_i^p} \doteq \mathbf{conv}\{\boldsymbol{D_i^p(t)} : \mathbf{1} \leq \boldsymbol{t} \leq |\boldsymbol{S^i}|\}$

Finally: $\boldsymbol{Z^p} \doteq \bigcap_i \boldsymbol{D_i^{p-1}}$

Row $\boldsymbol{i}$ of $\boldsymbol{Ax} \geq \boldsymbol{e}$: $\sum_{j \in S^i} x_j \geq 1$, where $S^i = \{j_1, \ldots, j_{|S^i|}\}$.

(a) For $\mathbf{1} \leq \boldsymbol{t} \leq |\boldsymbol{S^i}|$, polyhedron $D_i^p(t) \subseteq \mathbb{R}^n$ given by

$$
\begin{aligned}
x_{j_t} &= 1 && (5)\\
x_{j_h} &= 0 \quad \forall\, 1 \leq h < t, \quad \text{and} && (6)\\
x &\in \mathcal{Z}^{p-1} && (7)
\end{aligned}
$$

(b) Polyhedron $\boldsymbol{D_i^p} \doteq \mathrm{conv}\{\boldsymbol{D_i^p(t)} : \mathbf{1} \leq \boldsymbol{t} \leq |\boldsymbol{S^i}|\}$

Finally: $\boldsymbol{Z^p} \doteq \bigcap_i \boldsymbol{D_i^{p-1}}$

Lemma:

$\boldsymbol{Z^p}$ can be described by a polynomial-size formulation for fixed **p**, and its feasible solutions satisfy all valid inequalities of pitch $\leq \boldsymbol{p}$.

Subapplication 1a: minimum knapsack

Subapplication 1a: minimum knapsack

$$\min \; c^T x$$
$$\text{s.t.} \quad \sum_j w_j x_j \geq \; \boldsymbol{b}, \qquad x \text{ binary}$$

$\boldsymbol{w} \geq \mathbf{0}, \;\; \boldsymbol{b} > \mathbf{0}$

- "FPTAS" exists

Subapplication 1a: minimum knapsack

$$\begin{aligned} & \min \; c^T x \\ \text{s.t.} \quad & \sum_j w_j x_j \geq \; \boldsymbol{b}, \qquad x \text{ binary} \end{aligned}$$

$\boldsymbol{w} \geq \mathbf{0}, \;\; \boldsymbol{b} > \mathbf{0}$

- "FPTAS" exists (the one I know requires a disjunction)

Subapplication 1a: minimum knapsack

$$\min \; c^T x$$

$$\text{s.t.} \qquad \sum_j w_j x_j \geq \; \boldsymbol{b}, \qquad x \;\text{ binary}$$

$\boldsymbol{w} \geq \mathbf{0}, \;\; \boldsymbol{b} > \mathbf{0}$

- "FPTAS" exists (the one I know requires a disjunction)
- Problem not well understood

Subapplication 1a: minimum knapsack

$$\min \; c^T x$$

$$\text{s.t.} \quad \sum_j w_j x_j \geq \; \boldsymbol{b}, \qquad x \; \text{binary}$$

$\boldsymbol{w} \geq \mathbf{0}, \;\; \boldsymbol{b} > \mathbf{0}$

- "FPTAS" exists (the one I know requires a disjunction)
- Problem not well understood

Open question:

Subapplication 1a: minimum knapsack

$$\min \; c^T x$$
$$\text{s.t.} \quad \sum_j w_j x_j \geq \boldsymbol{b}, \qquad x \text{ binary}$$

$\boldsymbol{w} \geq \mathbf{0}, \;\; \boldsymbol{b} > \mathbf{0}$

- "FPTAS" exists (the one I know requires a disjunction)
- Problem not well understood

Open question: Given $\boldsymbol{w}, \boldsymbol{b}$ is there a compact extended formulation that yields a constant factor approximation, $\forall\, \boldsymbol{c}$?

Subapplication 1a: minimum knapsack

$$\min \; c^T x$$

$$\text{s.t.} \quad \sum_j w_j x_j \geq \boldsymbol{b}, \qquad x \text{ binary}$$

$\boldsymbol{w} \geq \mathbf{0}, \;\; \boldsymbol{b} > \mathbf{0}$

- "FPTAS" exists (the one I know requires a disjunction)
- Problem not well understood

Open question: Given $\boldsymbol{w}, \boldsymbol{b}$ is there a compact extended formulation that yields a constant factor approximation, $\forall \, \boldsymbol{c}$?

ANY constant whatsoever?

Application 1a: minimum knapsack

$$\begin{aligned} &\min \ c^T x \\ \text{s.t.} \quad &\sum_j w_j x_j \geq \ \boldsymbol{b} \end{aligned}$$

$\boldsymbol{w} \geq \mathbf{0}$, $\boldsymbol{b} > \mathbf{0}$, **integral**

Well-known result: equivalent to set-covering problem,

Application 1a: minimum knapsack

$$\begin{aligned} &\min \; c^T x \\ \text{s.t.} \quad &\sum_j w_j x_j \geq \; \boldsymbol{b} \end{aligned}$$

$\boldsymbol{w} \geq \mathbf{0}, \;\; \boldsymbol{b} > \mathbf{0}, \;\;$ **integral**

Well-known result: equivalent to set-covering problem, with constraints

$$\sum_{j \in S} x_j \; \geq \; 1, \quad \forall\, S \;\; \textbf{with} \; \sum_{j \in S} w_j \; \geq \; w^* \; \doteq \; \sum_j w_j - b + 1$$

Application 1a: minimum knapsack

$$\min \; c^T x$$
$$\text{s.t.} \quad \sum_j w_j x_j \geq \; \boldsymbol{b}$$

$\boldsymbol{w} \geq \mathbf{0}, \;\; \boldsymbol{b} > \mathbf{0}, \;\;$ **integral**

Well-known result: equivalent to set-covering problem, with constraints

$$\sum_{j \in S} x_j \; \geq \; 1, \quad \forall\, S \;\; \textbf{with} \; \sum_{j \in S} w_j \; \geq \; w^* \; \doteq \; \sum_j w_j - b + 1$$

But exponentially many constraints

Minimum knapsack

Using equivalence with set-covering,

Minimum knapsack

Using equivalence with set-covering,

- Compact, extended formulation that yields valid inequalities of **pitch** $\leq \boldsymbol{k}$, for fixed $\boldsymbol{k}$?

Minimum knapsack

Using equivalence with set-covering,

- Compact, extended formulation that yields valid inequalities of **pitch** $\leq \boldsymbol{k}$, for fixed $\boldsymbol{k}$? **X**

Minimum knapsack

Using equivalence with set-covering,

- Compact, extended formulation that yields valid inequalities of **pitch** $\leq k$, for fixed k? X
- Compact, extended formulation that yields valid inequalities with **coefficients** in $0, 1, \ldots, k$, for fixed k?

Minimum knapsack

Using equivalence with set-covering,

- Compact, extended formulation that yields valid inequalities of **pitch** $\leq \boldsymbol{k}$, for fixed $\boldsymbol{k}$? X
- Compact, extended formulation that yields valid inequalities with **coefficients** in $\mathbf{0}, \mathbf{1}, \ldots, \boldsymbol{k}$, for fixed $\boldsymbol{k}$? X

Minimum knapsack

Using equivalence with set-covering,

- Compact, extended formulation that yields valid inequalities of **pitch** $\leq k$, for fixed k? X
- Compact, extended formulation that yields valid inequalities with **coefficients** in $0, 1, \ldots, k$, for fixed k? X
- Polynomial-time separation over valid inequalities with **coefficients** in $0, 1, \ldots, k$, for fixed k?

Minimum knapsack

Using equivalence with set-covering,

- Compact, extended formulation that yields valid inequalities of **pitch** $\leq \boldsymbol{k}$, for fixed $\boldsymbol{k}$? **X**
- Compact, extended formulation that yields valid inequalities with **coefficients** in $\mathbf{0}, \mathbf{1}, \ldots, \boldsymbol{k}$, for fixed $\boldsymbol{k}$? **X**
- Polynomial-time separation over valid inequalities with **coefficients** in $\mathbf{0}, \mathbf{1}, \ldots, \boldsymbol{k}$, for fixed $\boldsymbol{k}$? **(implied)**

Minimum knapsack

Using equivalence with set-covering,

- Compact, extended formulation that yields valid inequalities of **pitch** $\leq \boldsymbol{k}$, for fixed $\boldsymbol{k}$? **X**
- Compact, extended formulation that yields valid inequalities with **coefficients** in $\mathbf{0}, \mathbf{1}, \ldots, \boldsymbol{k}$, for fixed $\boldsymbol{k}$? **X**
- Polynomial-time separation over valid inequalities with **coefficients** in $\mathbf{0}, \mathbf{1}, \ldots, \boldsymbol{k}$, for fixed $\boldsymbol{k}$? **(implied)**
- Polynomial-time **near** separation over valid inequalities with **coefficients** in $\mathbf{0}, \mathbf{1}, \ldots, \boldsymbol{k}$, for fixed $\boldsymbol{k}$.

Minimum knapsack

Using equivalence with set-covering,

- Compact, extended formulation that yields valid inequalities of **pitch** $\leq \boldsymbol{k}$, for fixed $\boldsymbol{k}$? **X**
- Compact, extended formulation that yields valid inequalities with **coefficients** in $\mathbf{0, 1, \ldots, \boldsymbol{k}}$, for fixed $\boldsymbol{k}$? **X**
- Polynomial-time separation over valid inequalities with **coefficients** in $\mathbf{0, 1, \ldots, \boldsymbol{k}}$, for fixed $\boldsymbol{k}$? **(implied)**
- Polynomial-time **near** separation over valid inequalities with **coefficients** in $\mathbf{0, 1, \ldots, \boldsymbol{k}}$, for fixed $\boldsymbol{k}$.

Given $\boldsymbol{y}$, either

- Find a valid inequality with coefficients in $\mathbf{0, 1, \ldots, \boldsymbol{k}}$, violated by $\boldsymbol{y}$, or
- Certify that $\boldsymbol{\alpha^T y \geq \alpha_0} - \mathrm{o}(\mathbf{1})$ for all valid $\alpha^T x \geq \alpha_0$ with $\alpha_j \in \{0, 1, \ldots, k\}$ for all j.

Minimum knapsack

Using equivalence with set-covering,

- Compact, extended formulation that yields valid inequalities of **pitch** $\leq \boldsymbol{k}$, for fixed $\boldsymbol{k}$? X
- Compact, extended formulation that yields valid inequalities with **coefficients** in $\mathbf{0, 1, \ldots,} \boldsymbol{k}$, for fixed $\boldsymbol{k}$? X
- Polynomial-time separation over valid inequalities with **coefficients** in $\mathbf{0, 1, \ldots,} \boldsymbol{k}$, for fixed $\boldsymbol{k}$? **(implied)**
- Polynomial-time **near** separation over valid inequalities with **coefficients** in $\mathbf{0, 1, \ldots,} \boldsymbol{k}$, for fixed $\boldsymbol{k}$.

Given $\boldsymbol{y}$, either

- Find a valid inequality with coefficients in $\mathbf{0, 1, \ldots,} \boldsymbol{k}$, violated by $\boldsymbol{y}$, or
- Certify that $\boldsymbol{\alpha^T y \geq \alpha_0} - \mathrm{o}(\mathbf{1})$ for all valid $\alpha^T x \geq \alpha_0$ with $\alpha_j \in \{0, 1, \ldots, k\}$ for all j. e.g. $\mathrm{o}(1) = \boldsymbol{O(1/n)}$

knapsack: $\sum_j w_j x_j \geq \mathrm{b}$, $\boldsymbol{w^*} \doteq \sum_j w_j - b + 1$

knapsack: $\sum_j w_j x_j \geq \mathrm{b}$, $\boldsymbol{w^*} \doteq \sum_j w_j - b + 1$

Warmup

Given $\boldsymbol{y}$, does it satisfy every valid inequality $\sum_{\boldsymbol{j \in S}} \boldsymbol{x_j} \geq \mathbf{2}$?

knapsack: $\sum_j w_j x_j \geq \mathrm{b}, \;\; \boldsymbol{w^*} \; \doteq \; \sum_j w_j - b + 1$

Warmup

Given $\boldsymbol{y}$, does it satisfy every valid inequality $\sum_{\boldsymbol{j \in S}} \boldsymbol{x_j} \geq \mathbf{2}$?
What is $\boldsymbol{S}$ here?

- Inequality is valid iff $\forall \boldsymbol{k} \in \boldsymbol{S}, \;\; \sum_{\boldsymbol{j \in S - k}} \boldsymbol{w_j} \;\geq\; \boldsymbol{w^*}$

knapsack: $\sum_j w_j x_j \geq \mathrm{b}$, $\boldsymbol{w^*} \doteq \sum_j w_j - b + 1$

Warmup

Given $\boldsymbol{y}$, does it satisfy every valid inequality $\sum_{\boldsymbol{j \in S}} \boldsymbol{x_j} \geq \mathbf{2}$?
What is $\boldsymbol{S}$ here?

- Inequality is valid iff $\forall \boldsymbol{k} \in \boldsymbol{S}$, $\sum_{\boldsymbol{j \in S - k}} \boldsymbol{w_j} \geq \boldsymbol{w^*}$
- Same as: $\sum_{\boldsymbol{j \in S - k}} \boldsymbol{w_j} \geq \boldsymbol{w^*}$ for specific $\boldsymbol{k}$ **:** $\mathrm{argmax}_{\boldsymbol{j \in S}}\{\boldsymbol{w_j}\}$

knapsack: $\sum_j w_j x_j \geq \mathrm{b}$, $\boldsymbol{w^*} \doteq \sum_j w_j - b + 1$

Warmup

Given $\boldsymbol{y}$, does it satisfy every valid inequality $\sum_{\boldsymbol{j \in S}} \boldsymbol{x_j} \geq \mathbf{2}$?
What is $\boldsymbol{S}$ here?

- Inequality is valid iff $\forall \boldsymbol{k} \in \boldsymbol{S}$, $\sum_{\boldsymbol{j \in S - k}} \boldsymbol{w_j} \geq \boldsymbol{w^*}$
- Same as: $\sum_{\boldsymbol{j \in S - k}} \boldsymbol{w_j} \geq \boldsymbol{w^*}$ for specific $\boldsymbol{k}$: $\mathbf{argmax}_{\boldsymbol{j \in S}}\{\boldsymbol{w_j}\}$
- For $\boldsymbol{k} = \mathbf{1, 2, \ldots,} \boldsymbol{n}$, solve minimum-knapsack problem

$$\min \sum_j y_j z_j \tag{8}$$

$$\text{s.t.} \sum_{j \neq k} w_j z_j \geq w^*, \qquad z \text{ binary} \tag{9}$$

$$\boldsymbol{z_k} = \mathbf{1},\ \boldsymbol{z_j} = \mathbf{0}\ \forall\, j \text{ with } w_j > w_k \tag{10}$$

knapsack: $\sum_j w_j x_j \geq \text{b}$, $\boldsymbol{w^*} \doteq \sum_j w_j - b + 1$

Warmup

Given $\boldsymbol{y}$, does it satisfy every valid inequality $\sum_{\boldsymbol{j \in S}} \boldsymbol{x_j} \geq \mathbf{2}$?
What is $\boldsymbol{S}$ here?

- Inequality is valid iff $\forall \boldsymbol{k} \in \boldsymbol{S}$, $\sum_{\boldsymbol{j \in S - k}} \boldsymbol{w_j} \geq \boldsymbol{w^*}$
- Same as: $\sum_{\boldsymbol{j \in S - k}} \boldsymbol{w_j} \geq \boldsymbol{w^*}$ for specific $\boldsymbol{k} : \operatorname{argmax}_{\boldsymbol{j \in S}} \{\boldsymbol{w_j}\}$
- For $\boldsymbol{k = 1, 2, \ldots, n}$, solve minimum-knapsack problem

$$\min \sum_j y_j z_j \tag{8}$$

$$\text{s.t.} \sum_{j \neq k} w_j z_j \geq w^*, \qquad z \text{ binary} \tag{9}$$

$$\boldsymbol{z_k = 1}, \ \boldsymbol{z_j = 0} \ \forall j \text{ with } w_j > w_k \tag{10}$$

Wait, how do we solve?

knapsack: $\sum_j w_j x_j \geq \text{b}$, $\boldsymbol{w^*} \doteq \sum_j w_j - b + 1$

Warmup

Given $\boldsymbol{y}$, does it satisfy every valid inequality $\sum_{j \in S} x_j \geq 2$?
What is $\boldsymbol{S}$ here?

- Inequality is valid iff $\forall \boldsymbol{k} \in \boldsymbol{S}$, $\sum_{j \in S-k} w_j \geq \boldsymbol{w^*}$
- Same as: $\sum_{j \in S-k} w_j \geq w^*$ for specific $\boldsymbol{k} : \operatorname{argmax}_{j \in S}\{w_j\}$
- For $k = 1, 2, \ldots, n$, solve minimum-knapsack problem

$$\min \sum_j y_j z_j \tag{8}$$

$$\text{s.t.} \sum_{j \neq k} w_j z_j \geq w^*, \qquad z \text{ binary} \tag{9}$$

$$z_k = 1,\; z_j = 0 \;\forall\, j \text{ with } w_j > w_k \tag{10}$$

Wait, how do we solve?
In objective round up y_j, to next multiple of $1/n^2$

knapsack: $\sum_j w_j x_j \geq \text{b}$, $\boldsymbol{w^*} \doteq \sum_j w_j - b + 1$

Warmup

Given $\boldsymbol{y}$, does it satisfy every valid inequality $\sum_{j \in S} x_j \geq 2$?
What is $\boldsymbol{S}$ here?

- Inequality is valid iff $\forall k \in S$, $\sum_{j \in S-k} w_j \geq w^*$
- Same as: $\sum_{j \in S-k} w_j \geq w^*$ for specific k : $\mathrm{argmax}_{j \in S}\{w_j\}$
- For $k = 1, 2, \ldots, n$, solve minimum-knapsack problem

$$\min \sum_j y_j z_j \tag{8}$$

$$\text{s.t.} \sum_{j \neq k} w_j z_j \geq w^*, \qquad z \text{ binary} \tag{9}$$

$$z_k = 1,\ z_j = 0\ \forall\, j \text{ with } w_j > w_k \tag{10}$$

Wait, how do we solve?
In objective round up y_j, to next multiple of $1/n^2$
So, get approximate separation, with violation if objective < 2

General case? First, coefficients in 0, 1, 2, 3

Example: $8x_1 + 5x_2 + 4x_3 + 4x_4 + 4x_5 \ \geq\ 13$ (the knapsack)

Valid: $x_1 + 2x_2 + x_3 + x_4 + x_5 \ \geq\ 3$ (non-monotone)

Not valid: $x_1 + x_2 + x_3 + x_4 + x_5 \ \geq\ 3$

A non-monotone "transposition" or "error"

General case? First, coefficients in 0, 1, 2, 3

$$\begin{aligned}
\textit{Example:}\quad & 8x_1 + 5x_2 + 4x_3 + 4x_4 + 4x_5 \ \geq\ 13 \quad \text{(the knapsack)} \\
\textit{Valid:}\quad & x_1 + 2x_2 + x_3 + x_4 + x_5 \ \geq\ 3 \quad \text{(non-monotone)} \\
\textbf{Not valid:}\quad & x_1 + x_2 + x_3 + x_4 + x_5 \ \geq\ 3
\end{aligned}$$

A non-monotone "transposition" or "error"

$$\begin{aligned}
\textit{Example:}\quad & 6x_1 + 6x_2 + 5x_3 + 4x_4 + 4x_5 \ \geq\ 13 \quad \text{(the knapsack)} \\
\textit{Valid:}\quad & x_1 + x_2 + 2x_3 + x_4 + x_5 \ \geq\ 3 \quad \text{(non-monotone, 2 errors)} \\
\textbf{Yes valid:}\quad & x_1 + x_2 + x_3 + x_4 + x_5 \ \geq\ 3
\end{aligned}$$

General case? First, coefficients in 0, 1, 2, 3

Example: $8x_1 + 5x_2 + 4x_3 + 4x_4 + 4x_5 \ \geq\ 13$ (the knapsack)

Valid: $x_1 + 2x_2 + x_3 + x_4 + x_5 \ \geq\ 3$ (non-monotone)

Not **valid:** $x_1 + x_2 + x_3 + x_4 + x_5 \ \geq\ 3$

A non-monotone "transposition" or "error"

Example: $6x_1 + 6x_2 + 5x_3 + 4x_4 + 4x_5 \ \geq\ 13$ (the knapsack)

Valid: $x_1 + x_2 + 2x_3 + x_4 + x_5 \ \geq\ 3$ (non-monotone, 2 errors)

Yes **valid:** $x_1 + x_2 + x_3 + x_4 + x_5 \ \geq\ 3$

$\rightarrow$ When right-hand side $= \mathbf{3}$, at most **one** error

General case? First, coefficients in 0, 1, 2, 3

Example: $8x_1 + 5x_2 + 4x_3 + 4x_4 + 4x_5 \;\geq\; 13$ (the knapsack)

Valid: $x_1 + 2x_2 + x_3 + x_4 + x_5 \;\geq\; 3$ (non-monotone)

Not valid: $x_1 + x_2 + x_3 + x_4 + x_5 \;\geq\; 3$

A non-monotone "transposition" or "error"

Example: $6x_1 + 6x_2 + 5x_3 + 4x_4 + 4x_5 \;\geq\; 13$ (the knapsack)

Valid: $x_1 + x_2 + 2x_3 + x_4 + x_5 \;\geq\; 3$ (non-monotone, 2 errors)

Yes valid: $x_1 + x_2 + x_3 + x_4 + x_5 \;\geq\; 3$

$\rightarrow$ When right-hand side $= \mathbf{3}$, at most **one** error

Separation by **enumeration** of errors; each case is a knapsack;

General case? First, coefficients in 0, 1, 2, 3

Example: $8x_1 + 5x_2 + 4x_3 + 4x_4 + 4x_5 \;\geq\; 13$ (the knapsack)

Valid: $x_1 + 2x_2 + x_3 + x_4 + x_5 \;\geq\; 3$ (non-monotone)

Not valid: $x_1 + x_2 + x_3 + x_4 + x_5 \;\geq\; 3$

A non-monotone "transposition" or "error"

Example: $6x_1 + 6x_2 + 5x_3 + 4x_4 + 4x_5 \;\geq\; 13$ (the knapsack)

Valid: $x_1 + x_2 + 2x_3 + x_4 + x_5 \;\geq\; 3$ (non-monotone, 2 errors)

Yes valid: $x_1 + x_2 + x_3 + x_4 + x_5 \;\geq\; 3$

$\rightarrow$ When right-hand side $\mathbf{= 3}$, at most **one** error
Separation by **enumeration** of errors; each case is a knapsack; $\boldsymbol{O(n^2)}$ cases

General case? (coefficients in 0, 1, 2, ..., k)

Basic principle: an inequality

$$k\,x(S_k) + (k-1)\,x(S_{k-1}) + \ldots + x(S_1) \;\geq\; k \tag{11}$$

is **equivalent** to its set of covers –

General case? (coefficients in 0, 1, 2, ..., k)

Basic principle: an inequality

$$k\, x(S_k) + (k-1)\, x(S_{k-1}) + \ldots + x(S_1) \ \geq \ k \tag{11}$$

is **equivalent** to its set of covers –

so (11) is valid **iff**

General case? (coefficients in 0, 1, 2, ..., k)

Basic principle: an inequality

$$k\,x(S_k) + (k-1)\,x(S_{k-1}) + \ldots + x(S_1) \ \geq \ k \tag{11}$$

is **equivalent** to its set of covers –

so (11) is valid **iff** its covers are **also** covers for the original knapsack

General case? (coefficients in 0, 1, 2, ..., k)

Basic principle: an inequality

$$k\,x(S_k) \,+\, (k-1)\,x(S_{k-1}) \,+ \ldots +\, x(S_1) \;\geq\; k \tag{11}$$

is **equivalent** to its set of covers **–**

so (11) is valid **iff** its covers are **also** covers for the original knapsack

Corollary: can show that (11) can have at most $<$ **k** errors

General case? (coefficients in 0, 1, 2, ..., k)

Basic principle: an inequality

$$k\,x(S_k) + (k-1)\,x(S_{k-1}) + \ldots + x(S_1) \;\geq\; k \tag{11}$$

is **equivalent** to its set of covers –

so (11) is valid **iff** its covers are **also** covers for the original knapsack

Corollary: can show that (11) can have at most $<$ **k** errors

or else it is dominated, or invalid

Separation by **enumeration** of errors; each case is a knapsack;

General case? (coefficients in 0, 1, 2, ..., k)

Basic principle: an inequality

$$k\, x(S_k) \,+\, (k-1)\, x(S_{k-1}) \,+ \ldots +\, x(S_1) \;\geq\; k \tag{11}$$

is **equivalent** to its set of covers –

so (11) is valid **iff** its covers are **also** covers for the original knapsack

Corollary: can show that (11) can have at most $<$ **k** errors

or else it is dominated, or invalid

Separation by **enumeration** of errors; each case is a knapsack;

$O(n^{F(k)})$ cases

Application 2: polynomial optimization problems and NN training

Application 2: polynomial optimization problems and NN training

Polynomial optimization:

$$
\begin{aligned}
& \min \; c^T x \\
\text{s.t.} \quad & f_i(x) \le 0, \qquad i = 1, \ldots, m \qquad \text{(polynomial ineq.)} \\
& 0 \le x_j \le 1, \qquad \text{all } j
\end{aligned}
\tag{12}
$$

- **Intersection graph**

Application 2: polynomial optimization problems and NN training

Polynomial optimization:

$$
\begin{aligned}
& \min \ c^T x \\
\text{s.t.} \quad & f_i(x) \le 0, \qquad i = 1, \ldots, m \qquad \text{(polynomial ineq.)} \\
& 0 \le x_j \le 1, \qquad \text{all } j
\end{aligned} \tag{12}
$$

- **Intersection graph**
  A vertex for each variable and an edge anytime two variables appear in the same $\boldsymbol{f_i}$
- **Tree-width**

Polynomial optimization:

$$\begin{aligned} & \min \; c^T x \\ \text{s.t.} \quad & f_i(x) \le 0, \qquad i = 1, \ldots, m \qquad \text{(polynomial ineq.)} \\ & 0 \le x_j \le 1, \qquad \text{all } j \end{aligned}$$

- **Intersection graph**

Polynomial optimization:

$$
\begin{array}{lll}
 & \min \; c^T x & \\
\text{s.t.} & f_i(x) \leq 0, \qquad i = 1, \ldots, m & \text{(polynomial ineq.)} \\
 & 0 \leq x_j \leq 1, \qquad \text{all } j &
\end{array}
$$

- **Intersection graph**
  A vertex for each variable and an edge anytime two variables appear in the same $\boldsymbol{f_i}$
- **Tree-width** of a graph **G**
  Minimum clique number (minus one) over all chordal supergraphs of G

Polynomial optimization:

$$\begin{array}{lll} & \min \; c^T x & & \\ \text{s.t.} & f_i(x) \le 0, & i = 1, \ldots, m & \text{(polynomial ineq.)} \\ & 0 \le x_j \le 1, & \text{all } j & \end{array}$$

Theorem (B. and Muñoz 2015, SIOPT 2018).

Suppose:

the intersection graph has tree-width $\boldsymbol{\omega}$ and the $\boldsymbol{f_i}$ of degree $\le \boldsymbol{\rho}$.

Polynomial optimization:

$$\begin{array}{llll} & \min \ c^T x & & \\ \text{s.t.} & f_i(x) \le 0, & i = 1, \ldots, m & \text{(polynomial ineq.)} \\ & 0 \le x_j \le 1, & \text{all } j & \end{array}$$

Theorem (B. and Muñoz 2015, SIOPT 2018).

Suppose:

the intersection graph has tree-width ω and the f_i of degree $\le \rho$.

Then, for every $0 < \epsilon < 1$ there is a **disjunctive LP** relaxation with

Polynomial optimization:

$$\begin{array}{llll} & \min \ c^T x & & \\ \text{s.t.} & f_i(x) \le 0, & i = 1, \ldots, m & \text{(polynomial ineq.)} \\ & 0 \le x_j \le 1, & \text{all } j & \end{array}$$

Theorem (B. and Muñoz 2015, SIOPT 2018).

Suppose:

the intersection graph has tree-width $\boldsymbol{\omega}$ and the $\boldsymbol{f_i}$ of degree $\le \boldsymbol{\rho}$.

Then, for every $\mathbf{0} < \boldsymbol{\epsilon} < \mathbf{1}$ there is a **disjunctive LP** relaxation with

$\boldsymbol{O\left((2\rho/\epsilon)^{\omega+1}\, n \log(\rho/\epsilon)\right)}$ variables and constraints

Polynomial optimization:

$$\begin{array}{llll} & \min \ c^T x & & \\ \text{s.t.} & f_i(x) \le 0, & i = 1, \ldots, m & \text{(polynomial ineq.)} \\ & 0 \le x_j \le 1, & \text{all } j & \end{array}$$

Theorem (B. and Muñoz 2015, SIOPT 2018).

Suppose:

the intersection graph has tree-width $\boldsymbol{\omega}$ and the $\boldsymbol{f_i}$ of degree $\le \boldsymbol{\rho}$.

Then, for every $\mathbf{0} < \boldsymbol{\epsilon} < \mathbf{1}$ there is a **disjunctive LP** relaxation with

$\boldsymbol{O}\left(\mathbf{(2\rho/\epsilon)^{\omega+1}\, n \log(\rho/\epsilon)}\right)$ variables and constraints

Optimality and feasibility errors $\boldsymbol{O(\epsilon)}$ (additive)

Subapplication 2a: training of deep neural networks with RLUs

Subapplication 2a: training of deep neural networks with RLUs

As per Arora Basu Mianjy Mukherjee ICLR '18

The setup:

- $\boldsymbol{D}$ data points $(\boldsymbol{x_i}, \boldsymbol{y_i}),\ \mathbf{1} \leq \boldsymbol{i} \leq \boldsymbol{D}$, $\boldsymbol{x_i} \in \mathbb{R}^{\boldsymbol{n}}$, $\boldsymbol{y_i} \in \mathbb{R}$

Subapplication 2a: training of deep neural networks with RLUs

As per Arora Basu Mianjy Mukherjee ICLR '18

The setup:

- $\boldsymbol{D}$ data points $(\boldsymbol{x_i}, \boldsymbol{y_i}),\ \mathbf{1} \leq \boldsymbol{i} \leq \boldsymbol{D},\ \ \boldsymbol{x_i} \in \mathbb{R}^{\boldsymbol{n}},\ \ \boldsymbol{y_i} \in \mathbb{R}$
- **Task:** **compute** a function $\boldsymbol{f} : \mathbb{R}^{\boldsymbol{n}} \to \mathbb{R}$ to minimize

$$\frac{1}{D} \sum_{i=1} \left(y_i - f(x_i)\right)^2$$

Subapplication 2a: training of deep neural networks with RLUs

As per Arora Basu Mianjy Mukherjee ICLR '18

The setup:

- $\boldsymbol{D}$ data points $(\boldsymbol{x_i}, \boldsymbol{y_i})$, $\boldsymbol{1 \le i \le D}$, $\boldsymbol{x_i} \in \mathbb{R}^{\boldsymbol{n}}$, $\boldsymbol{y_i} \in \mathbb{R}$
- **Task:** **compute** a function $\boldsymbol{f} : \mathbb{R}^{\boldsymbol{n}} \to \mathbb{R}$ to minimize

$$\frac{1}{D} \sum_{i=1} (y_i - f(x_i))^2$$

- $\boldsymbol{f = T_{k+1} \circ \sigma \circ T_k \circ \sigma \ldots \circ \sigma \circ T_1}$ ("$\circ$" $=$ composition)

Subapplication 2a: training of deep neural networks with RLUs

As per Arora Basu Mianjy Mukherjee ICLR '18

The setup:

- D data points (x_i, y_i), $1 \le i \le D$, $x_i \in \mathbb{R}^n$, $y_i \in \mathbb{R}$
- **Task:** **compute** a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ to minimize

$$\frac{1}{D} \sum_{i=1} (y_i - f(x_i))^2$$

- $f = T_{k+1} \circ \sigma \circ T_k \circ \sigma \ldots \circ \sigma \circ T_1$ ("$\circ$" $=$ composition)
- $\sigma(t) = \max\{0, t\}$

Subapplication 2a: training of deep neural networks with RLUs

As per Arora Basu Mianjy Mukherjee ICLR '18

The setup:

- D data points (x_i, y_i), $1 \le i \le D$, $x_i \in \mathbb{R}^n$, $y_i \in \mathbb{R}$
- **Task:** **compute** a function $f : \mathbb{R}^n \to \mathbb{R}$ to minimize

$$\frac{1}{D} \sum_{i=1} (y_i - f(x_i))^2$$

- $f = T_{k+1} \circ \sigma \circ T_k \circ \sigma \ldots \circ \sigma \circ T_1$ ("$\circ$" = composition)
- $\sigma(t) = \max\{0, t\}$
- Each T_h affine: $T_h(y) = A_h y + b_h$,

Subapplication 2a: training of deep neural networks with RLUs

As per Arora Basu Mianjy Mukherjee ICLR '18

The setup:

- D data points (x_i, y_i), $1 \le i \le D$, $x_i \in \mathbb{R}^n$, $y_i \in \mathbb{R}$
- **Task:** **compute** a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ to minimize

$$\frac{1}{D} \sum_{i=1} (y_i - f(x_i))^2$$

- $f = T_{k+1} \circ \sigma \circ T_k \circ \sigma \ldots \circ \sigma \circ T_1$ ("$\circ$" = composition)
- $\sigma(t) = \max\{0, t\}$
- Each T_h affine: $T_h(y) = A_h y + b_h$,
- For some w, A_1 is $n \times w$, A_{k+1} is $w \times 1$, A_h is $w \times w$ otherwise.

Subapplication 2a: training of deep neural networks with RLUs

As per Arora Basu Mianjy Mukherjee ICLR '18

The setup:

- D data points (x_i, y_i), $1 \leq i \leq D$, $x_i \in \mathbb{R}^n$, $y_i \in \mathbb{R}$
- **Task:** **compute** a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ to minimize

$$\frac{1}{D} \sum_{i=1} (y_i - f(x_i))^2$$

- $f = T_{k+1} \circ \sigma \circ T_k \circ \sigma \ldots \circ \sigma \circ T_1$ ("$\circ$" = composition)
- $\sigma(t) = \max\{0, t\}$
- Each T_h affine: $T_h(y) = A_h y + b_h$,
- For some w, A_1 is $n \times w$, A_{k+1} is $w \times 1$, A_h is $w \times w$ otherwise. Similarly with the b_h.

- D data points (x_i, y_i), $1 \le i \le D$, $x_i \in \mathbb{R}^n$, $y_i \in \mathbb{R}$
- **Task:** **compute** a function $f : \mathbb{R}^n \to \mathbb{R}$ to minimize

$$\frac{1}{D} \sum_{i=1} (y_i - f(x_i))^2$$

- $f = T_{k+1} \circ \sigma \circ T_k \circ \sigma \ldots \circ \sigma \circ T_1$ ("$\circ$" = composition)
- $\sigma(t) = \max\{0, t\}$
- Each T_h affine: $T_h(y) = A_h y + b_h$,
- For some w, A_1 is $n \times w$, A_{k+1} is $w \times 1$, A_h is $w \times w$ otherise. Similarly with the b_h.

- D data points (x_i, y_i), $1 \le i \le D$, $x_i \in \mathbb{R}^n$, $y_i \in \mathbb{R}$
- **Task:** **compute** a function $f : \mathbb{R}^n \to \mathbb{R}$ to minimize

$$\frac{1}{D} \sum_{i=1} (y_i - f(x_i))^2$$

- $f = T_{k+1} \circ \sigma \circ T_k \circ \sigma \ldots \circ \sigma \circ T_1$ ("$\circ$" $=$ composition)
- $\sigma(t) = \max\{0, t\}$
- Each T_h affine: $T_h(y) = A_h y + b_h$,
- For some w, A_1 is $n \times w$, A_{k+1} is $w \times 1$, A_h is $w \times w$ otherise. Similarly with the b_h.

Theorem (Arora et al 2018).
If k = 1 (one "hidden layer") there is an exact algorithm of complexity

- D data points (x_i, y_i), $1 \le i \le D$, $x_i \in \mathbb{R}^n$, $y_i \in \mathbb{R}$
- **Task:** **compute** a function $f : \mathbb{R}^n \to \mathbb{R}$ to minimize

$$\frac{1}{D} \sum_{i=1} (y_i - f(x_i))^2$$

- $f = T_{k+1} \circ \sigma \circ T_k \circ \sigma \ldots \circ \sigma \circ T_1$ ("$\circ$" $=$ composition)
- $\sigma(t) = \max\{0, t\}$
- Each T_h affine: $T_h(y) = A_h y + b_h$,
- For some w, A_1 is $n \times w$, A_{k+1} is $w \times 1$, A_h is $w \times w$ otherise. Similarly with the b_h.

Theorem (Arora et al 2018).
If $k = 1$ (one "hidden layer") there is an exact algorithm of complexity

$$O\left(2^w D^{nw} \mathrm{poly}(D, n, w) \right)$$

Polynomial in the size of the data set, for fixed n, w

- D data points (x_i, y_i), $1 \leq i \leq D$, $x_i \in \mathbb{R}^n$, $y_i \in \mathbb{R}$
- **Task:** compute $f = T_{k+1} \circ \sigma \circ T_k \circ \sigma \ldots \circ \sigma \circ T_1$ to minimize $\frac{1}{D}\sum_{i=1}(y_i - f(x_i))^2$
- A_1 is $n \times w$, A_{k+1} is $w \times 1$, A_h is $w \times w$ otherise.

Theorem (Arora et al 2018).
If $k = 1$ (one "hidden layer") there is an exact algorithm of complexity

$$O\left(2^w D^{nw} \mathrm{poly}(D, n, w)\right)$$

- D data points (x_i, y_i), $1 \le i \le D$, $x_i \in \mathbb{R}^n$, $y_i \in \mathbb{R}$
- **Task:** compute $f = T_{k+1} \circ \sigma \circ T_k \circ \sigma \ldots \circ \sigma \circ T_1$ to minimize $\frac{1}{D} \sum_{i=1} (y_i - f(x_i))^2$
- A_1 is $n \times w$, A_{k+1} is $w \times 1$, A_h is $w \times w$ otherise.

Theorem (Arora et al 2018).
If $k = 1$ (one "hidden layer") there is an exact algorithm of complexity

$$O\left(2^w D^{nw} \mathrm{poly}(D, n, w)\right)$$

Application of B. and Muñoz poly-opt result:

- $\boldsymbol{D}$ data points $(\boldsymbol{x_i}, \boldsymbol{y_i})$, $1 \leq \boldsymbol{i} \leq \boldsymbol{D}$, $\boldsymbol{x_i} \in \mathbb{R}^{\boldsymbol{n}}$, $\boldsymbol{y_i} \in \mathbb{R}$
- **Task:** compute $\boldsymbol{f} = \boldsymbol{T_{k+1}} \circ \boldsymbol{\sigma} \circ \boldsymbol{T_k} \circ \boldsymbol{\sigma} \ldots \circ \boldsymbol{\sigma} \circ \boldsymbol{T_1}$ to minimize $\frac{1}{D} \sum_{i=1} (y_i - f(x_i))^2$
- $\boldsymbol{A_1}$ is $\boldsymbol{n} \times \boldsymbol{w}$, $\boldsymbol{A_{k+1}}$ is $\boldsymbol{w} \times \boldsymbol{1}$, $\boldsymbol{A_h}$ is $\boldsymbol{w} \times \boldsymbol{w}$ otherise.

Theorem (Arora et al 2018).
If $\mathbf{k = 1}$ (one "hidden layer") there is an exact algorithm of complexity

$$O\,(\,2^w D^{nw} \mathrm{poly}(D, n, w)\,)$$

Application of B. and Muñoz poly-opt result:

- **Weakening:** Assume that a bound on the absolute value of the entries in the $\boldsymbol{A_h}$, $\boldsymbol{b_h}$ is known

- $\boldsymbol{D}$ data points $(\boldsymbol{x_i}, \boldsymbol{y_i})$, $\boldsymbol{1} \leq \boldsymbol{i} \leq \boldsymbol{D}$, $\boldsymbol{x_i} \in \mathbb{R}^{\boldsymbol{n}}$, $\boldsymbol{y_i} \in \mathbb{R}$
- **Task:** compute $\boldsymbol{f} = \boldsymbol{T_{k+1}} \circ \boldsymbol{\sigma} \circ \boldsymbol{T_k} \circ \boldsymbol{\sigma} \ldots \circ \boldsymbol{\sigma} \circ \boldsymbol{T_1}$ to minimize $\frac{1}{D} \sum_{i=1} (y_i - f(x_i))^2$
- $\boldsymbol{A_1}$ is $\boldsymbol{n} \times \boldsymbol{w}$, $\boldsymbol{A_{k+1}}$ is $\boldsymbol{w} \times \boldsymbol{1}$, $\boldsymbol{A_h}$ is $\boldsymbol{w} \times \boldsymbol{w}$ otherise.

Theorem (Arora et al 2018).
If $\mathbf{k = 1}$ (one "hidden layer") there is an exact algorithm of complexity

$$O\left(2^w D^{nw} \mathrm{poly}(D, n, w)\right)$$

Application of B. and Muñoz poly-opt result:

- **Weakening:** Assume that a bound on the absolute value of the entries in the $\boldsymbol{A_h}$, $\boldsymbol{b_h}$ is known
- **Weakening:** For any $\mathbf{0} < \epsilon < \mathbf{1}$, **additive errors** $\boldsymbol{O(\epsilon)}$

- $\boldsymbol{D}$ data points $(\boldsymbol{x_i}, \boldsymbol{y_i})$, $\boldsymbol{1 \le i \le D}$, $\boldsymbol{x_i} \in \mathbb{R}^{\boldsymbol{n}}$, $\boldsymbol{y_i} \in \mathbb{R}$
- **Task:** compute $\boldsymbol{f} = \boldsymbol{T_{k+1} \circ \sigma \circ T_k \circ \sigma \ldots \circ \sigma \circ T_1}$ to minimize $\frac{1}{D} \sum_{i=1} (y_i - f(x_i))^2$
- $\boldsymbol{A_1}$ is $\boldsymbol{n \times w}$, $\boldsymbol{A_{k+1}}$ is $\boldsymbol{w \times 1}$, $\boldsymbol{A_h}$ is $\boldsymbol{w \times w}$ otherise.

Theorem (Arora et al 2018).
If k = 1 (one "hidden layer") there is an exact algorithm of complexity

$$O\left(2^w D^{nw} \text{poly}(D, n, w)\right)$$

Application of B. and Muñoz poly-opt result:

- **Weakening:** Assume that a bound on the absolute value of the entries in the $\boldsymbol{A_h}$, $\boldsymbol{b_h}$ is known
- **Weakening:** For any $\boldsymbol{0 < \epsilon < 1}$, **additive errors** $\boldsymbol{O(\epsilon)}$

Theorem. For any $\boldsymbol{k, n, w, \epsilon}$ **approximate LP** of size

$$\boldsymbol{O}\left(\left(\frac{4}{\epsilon}\right)^{\boldsymbol{O((k-1)w^2+nw)}} \text{poly}(\boldsymbol{D, n, w, k}) \right)$$