A General Framework for Convex Relaxation of Polynomial Optimization Problems over Cones

Masakazu Kojima, Sunyoung Kim and Hayato Waki

March 2003

Contents

1. Convex relaxation of global optimization problems

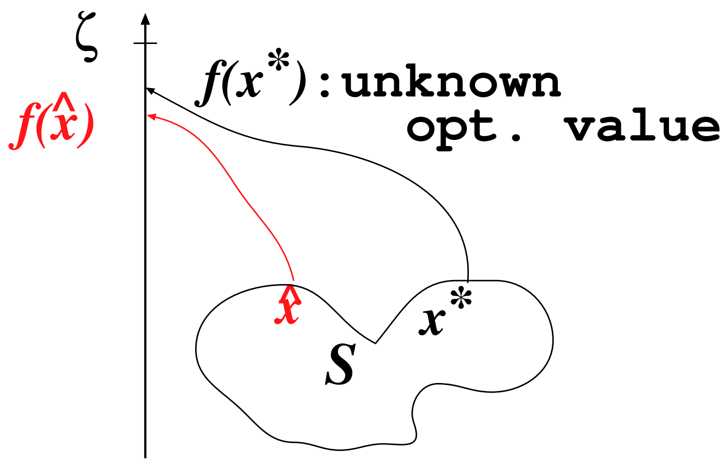
(1) max. $f(x)$ sub.to $x \in S$, where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $S \subset \mathbb{R}^n$.

(a) a feasible solution $\hat{x} \in S$ with a larger objective value $f(\hat{x})$

(b) a smaller upper bound ζ for the unknown optimal value $f(x^*)$

$\Longrightarrow$ a main role of convex relaxation

If $\zeta - f(\hat{x})$ is smaller, we can accept $\hat{x}$ as a higher quality approximate optimal solution.



- SDP relaxation is very powerful in theory.
  - (a) Lovász-Schrijver'91 for 0-1 IPs
  - (b) Goemans-Willianson'95 for max-cut problems
  - (c) Some special QOPs can be solved approximately or exactly by SDP relaxation, Nesterov'88, Ye'99, Zhang'00, Ye-Zhang'01
  - (d) Hierarchical SDP relaxation by Lasserre'01, Parrilo for polynomial programs — theoretically very powerful: the optimal value can be approximated in arbitrary accuracy by solving a finite number of SDP relaxations.
  - (e) . . .

- **Can SDP (or convex) relaxation, without combining any technique on (b), be powerful enough to solve practical large scale problems?**

???, mainly because solving large scale SDPs is expensive .

- **Incorporate convex relaxation into branch-and-bound method.**
- **How to combine them effectively.**
- **Exploration of effective and inexpensive convex relaxations.**

Besides SDP and LP relaxation, we explore various convex relaxations towards practically effective and efficient methods.

The purpose of this talk is to present

a general framework for convex relaxation methods

The main ingredients are:

(a) Polynomial Optimization Problems $\supset$ QOPs and 0-1 IPs

$\Downarrow$(b) Add valid constraints and reformulate

(c) Polynomial Optimization Problems over Cones

$\Downarrow$ (d) Linearization

(e) Linear Optimization Problems over Cones

I will talk about

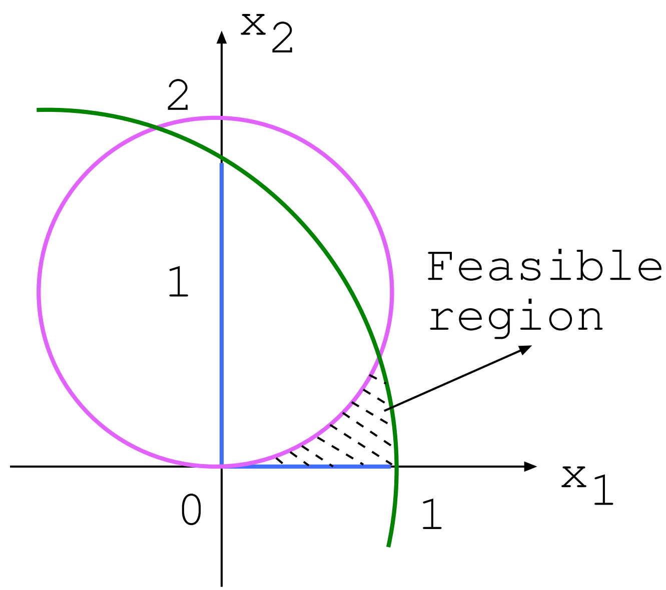
- **An illustrative example**
- **(c) $\Rightarrow$ (d) $\Rightarrow$ (e)**
- **(b)**

2. An illustrative example

Original problem: max. $-2x_1 + x_2$

sub.to $x_1 \geq 0,\ x_2 \geq 0,\ x_1^2 + x_2^2 - 2x_2 \geq 0,$

$$\left\| \begin{pmatrix} x_1 + 1 \\ x_2 \end{pmatrix} \right\| \leq 2 \text{ (SOCP constraint)}$$


x2
2
1
Feasible
region
0
1
x1

$\Downarrow$ **Valid constraints and/or reformulation**

$$\begin{array}{ll} \textbf{max.} & -2x_1 + x_2 \\ \textbf{sub.to} & x_1 \geq 0,\ x_2 \geq 0,\ x_1^2 \geq 0,\ x_1x_2 \geq 0,\ x_2^2 \geq 0,\ x_1^2 + x_2^2 - 2x_2 \geq 0, \\ & \left\| \begin{pmatrix} x_1 + 1 \\ x_2 \end{pmatrix} \right\| \leq 2,\ \left\| \begin{pmatrix} x_1^2 + x_1 \\ x_1x_2 \end{pmatrix} \right\| \leq 2x_1,\ \left\| \begin{pmatrix} x_1x_2 + x_2 \\ x_2^2 \end{pmatrix} \right\| \leq 2x_2. \end{array}$$

$\Downarrow$ **Linearization: Keep the linear terms, but replace each nonlinear term by a single independent variable**

$$
\begin{array}{ll}
\text{max.} & -2x_1 + x_2 \\
\text{sub.to} & x_1 \geq 0,\ x_2 \geq 0,\ X_{11} \geq 0,\ X_{12} \geq 0,\ X_{22} \geq 0, \\
& X_{11} + X_{22} - 2x_2 \geq 0, \\
& \left\| \begin{pmatrix} x_1 + 1 \\ x_2 \end{pmatrix} \right\| \leq 2, \left\| \begin{pmatrix} X_{11} + x_1 \\ X_{12} \end{pmatrix} \right\| \leq 2x_1, \left\| \begin{pmatrix} X_{12} + x_2 \\ X_{22} \end{pmatrix} \right\| \leq 2x_2.
\end{array}
$$

$$\Uparrow \quad X_{11} = x_1 x_1,\, X_{12} = x_1 x_2,\, X_{22} = x_2 x_2$$

$$\begin{array}{ll}
\textbf{max.} & -2x_1 + x_2 \\
\textbf{sub.to} & x_1 \geq 0,\ x_2 \geq 0,\ X_{11} \geq 0,\ X_{12} \geq 0,\ X_{22} \geq 0, \\
 & X_{11} + X_{22} - 2x_2 \geq 0, \\
 & \left\| \begin{pmatrix} x_1 + 1 \\ x_2 \end{pmatrix} \right\| \leq 2,\ \left\| \begin{pmatrix} X_{11} + x_1 \\ X_{12} \end{pmatrix} \right\| \leq 2x_1,\ \left\| \begin{pmatrix} X_{12} + x_2 \\ X_{22} \end{pmatrix} \right\| \leq 2x_2.
\end{array}$$

3. Polynomial opt. problems over cones and their linearization

max. $f_0(x)$ **sub.to** $f(x) \in \mathcal{K}$, **where**

$\mathcal{K}$ **: a closed convex cone in** $\mathbb{R}^m$,

$x = (x_1, \ldots, x_n)$ **: a variable vector,** $f(x) \equiv (f_1(x), \ldots, f_m(x))$,

$f_j(x)$ **: a polynomial in** $x_1, \ldots, x_n$ $(j = 0, 1, \ldots, m)$.

Typical examples of $\mathcal{K}$: $\mathbb{R}^m_+$ **: the nonnegative orthant of** $\mathbb{R}^m$**.**

$\mathbb{S}^\ell_+$ **: the cone of $\ell \times \ell$ psd symmetric matrices, where we identify each $\ell \times \ell$ matrix as an $\ell \times \ell$ dim vector.**

$$\mathbb{N}^{1+\ell}_p \equiv \left\{ v = (v_0, v_1, \ldots, v_\ell) \in \mathbb{R}^{1+\ell} : \sum_{i=1}^{\ell} |v_i|^p \leq v_0^p \right\}$$

: the pth order cone ($p \geq 1$).

$\mathbb{N}^{1+\ell}_2$ **: the second order cone.**

When $f_j(x)$ ($j = 0, 1, 2, \ldots, m$) are linear,

$$\mathcal{K} = \mathbb{S}^\ell_+ \Rightarrow \text{SDP (Semidefinite Program)},$$

$$\mathcal{K} = \mathbb{N}^{1+\ell}_2 \Rightarrow \text{SOCP (Second-Order Cone Program)}$$

Linearization — Keep the linear terms, but replace each nonlinear term by a single independent variable.

Example 1:

$$f(x_1, x_2) = \begin{pmatrix} 1 - 2x_1 + 3x_2 + 4x_1^2 + 5x_1x_2 + 6x_2^2 \\ 9 + 8x_1 + 7x_2 + 6x_1^2 - 5x_1x_2 - 4x_2^2 \end{pmatrix} \in \mathcal{K}$$

$\Downarrow$ **Linearization**

$$F(x_1, x_2, X_{11}, X_{12}, X_{22}) = \begin{pmatrix} 1 - 2x_1 + 3x_2 + 4X_{11} + 5X_{12} + 6X_{22} \\ 9 + 8x_1 + 7x_2 + 6X_{11} - 5X_{12} - 4X_{22} \end{pmatrix} \in \mathcal{K}$$

Here the three new variables X_{11}, X_{12} and X_{22} are introduced.

Linearization — **Keep the linear terms, but replace each nonlinear term by a single independent variable.**

Example 2:

$$f(x_1, x_2, x_3) = \begin{pmatrix} 1 - 2x_1 + 3x_2 + 4x_1^2 x_3 + 5x_1 x_2 x_3 + 6x_3^4 \\ 9 + 8x_1 + 7x_2 + 6x_1^2 x_3 - 5x_1 x_2 x_3 - 4x_3^4 \end{pmatrix} \in \mathcal{K}$$

$\Downarrow$ **Linearization**

$$F(x_1, x_2, U, V, W) = \begin{pmatrix} 1 - 2x_1 + 3x_2 + 4U + 5V + 6W \\ 9 + 8x_1 + 7x_2 + 6U - 5V - 4W \end{pmatrix} \in \mathcal{K}$$

Here the new variables U, V and W are introduced. In general, we need a systematic method of assigning a new variable to each nonlinear term.

Linearization — Keep the linear terms, but replace each nonlinear term by a single independent variable.

Systematic method of assigning a new variable to each nonlinear term:

$$\text{a nonlinear term } x_1^{\alpha} x_2^{\beta} \cdots x_n^{\zeta} \Rightarrow y_{(\alpha,\beta,\ldots,\zeta)} \in \mathbb{R} \text{ a new variable}$$

For example,

$$n = 5,\ x_1^2 x_2 x_3^3 x_5^4 = x_1^2 x_2^1 x_3^3 x_4^0 x_5^4 \Rightarrow y_{(2,1,3,0,4)}.$$

In theory, any method of assigning a new variable to each nonlinear term works. ⇒ This method is not essential.

4. General framework for convex relaxation

Original QOP, 0-1 IP, Polynomial programs to be solved

$\Downarrow$ **Valid constraints and/or reformulate**

POP: max. $f_0(x)$ **sub.to** $f(x) \in \mathcal{K}$**, where**

$\mathcal{K}$: **a closed convex cone in** $\mathbb{R}^m$,

$x = (x_1, \dots, x_n)$: **a variable vector,** $f(x) \equiv (f_1(x), \dots, f_m(x))$,

$f_j(x)$: **a polynomial in** $x_1, \dots, x_n$ $(j = 0, 1, \dots, m)$.

$\Downarrow$ **Linearization — Keep the linear terms, but replace each**
$\Downarrow$ **nonlinear term by a single independent variable.**

LOP: max. $F_0(x, y)$ **sub.to** $F(x, y) \in \mathcal{K}$**, where**

y **denotes a new variable vector whose elements correspond to nonlinear terms appeared in the polynomials** $f_j(x)$ $(j = 0, 1, \dots, m)$.

Illustrative example again — 1

Original problem: max. $-2x_1 + x_2$

sub.to $x_1 \geq 0,\ x_2 \geq 0,\ x_1^2 + x_2^2 - 2x_2 \geq 0,$

$\left\| \begin{pmatrix} x_1 + 1 \\ x_2 \end{pmatrix} \right\| \leq 2$ (SOCP constraint)

$\Downarrow$ **Valid constraints and/or reformulation**

$$
\begin{array}{ll}
\text{max.} & -2x_1 + x_2 \\
\text{sub.to} & x_1 \geq 0,\ x_2 \geq 0,\ x_1^2 \geq 0,\ x_1 x_2 \geq 0,\ x_2^2 \geq 0,\ x_1^2 + x_2^2 - 2x_2 \geq 0, \\
& \left\| \begin{pmatrix} x_1 + 1 \\ x_2 \end{pmatrix} \right\| \leq 2,\ \left\| \begin{pmatrix} x_1^2 + x_1 \\ x_1 x_2 \end{pmatrix} \right\| \leq 2x_1,\ \left\| \begin{pmatrix} x_1 x_2 + x_2 \\ x_2^2 \end{pmatrix} \right\| \leq 2x_2.
\end{array}
$$

$\Downarrow$ **Linearization: Keep the linear terms, but replace each nonlinear term by a single independent variable**

$$
\begin{array}{ll}
\text{max.} & -2x_1 + x_2 \\
\text{sub.to} & x_1 \geq 0,\ x_2 \geq 0,\ X_{11} \geq 0,\ X_{12} \geq 0, X_{22} \geq 0, \\
& X_{11} + X_{22} - 2x_2 \geq 0, \\
& \left\| \begin{pmatrix} x_1 + 1 \\ x_2 \end{pmatrix} \right\| \leq 2, \left\| \begin{pmatrix} X_{11} + x_1 \\ X_{12} \end{pmatrix} \right\| \leq 2x_1, \left\| \begin{pmatrix} X_{12} + x_2 \\ X_{22} \end{pmatrix} \right\| \leq 2x_2.
\end{array}
$$

$\Downarrow$ **Valid constraints and/or reformulation**

$$\begin{array}{ll} \textbf{max.} & -2x_1 + x_2 \\ \textbf{sub.to} & x_1 \geq 0,\ x_2 \geq 0,\ x_1^2 + x_2^2 - 2x_2 \geq 0, \\ & \left\| \begin{pmatrix} x_1 + 1 \\ x_2 \end{pmatrix} \right\| \leq 2,\ \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix} \begin{pmatrix} 1 & x_1 & x_2 \end{pmatrix} \equiv \begin{pmatrix} 1 & x_1 & x_2 \\ x_1 & x_1^2 & x_1 x_2 \\ x_2 & x_1 x_2 & x_2^2 \end{pmatrix} \succeq O. \end{array}$$

$\Downarrow$ **Linearization**

$$\begin{array}{ll} \textbf{max.} & -2x_1 + x_2 \qquad \text{— SDP} \\ \textbf{sub.to} & x_1 \geq 0,\ x_2 \geq 0,\ X_{11} + X_{22} - 2x_2 \geq 0, \\ & \left\| \begin{pmatrix} x_1 + 1 \\ x_2 \end{pmatrix} \right\| \leq 2,\ \begin{pmatrix} 1 & x_1 & x_2 \\ x_1 & X_{11} & X_{12} \\ x_2 & X_{12} & X_{22} \end{pmatrix} \succeq O. \end{array}$$

Given a problem, there are various ways of adding valid constraints and reformulating the problem. They usually yield different convex relaxations.

Illustrative example again — 2

Original problem:

$$\begin{array}{ll} \text{max.} & -2x_1 + x_2 \\ \text{sub.to} & x_1 \geq 0,\ x_2 \geq 0,\ x_1^2 + x_2^2 - 2x_2 \geq 0, \\ & \left\| \begin{pmatrix} x_1 + 1 \\ x_2 \end{pmatrix} \right\| \leq 2 \ \text{(SOCP constraint)} \end{array},$$

we obtained two distinct convex relaxations.

$$\begin{array}{ll} \text{max.} & -2x_1 + x_2 \qquad \text{— SOCP} \\ \text{sub.to} & x_1 \geq 0,\ x_2 \geq 0,\ X_{11} \geq 0,\ X_{12} \geq 0, X_{22} \geq 0, \\ & X_{11} + X_{22} - 2x_2 \geq 0, \\ & \left\| \begin{pmatrix} x_1 + 1 \\ x_2 \end{pmatrix} \right\| \leq 2, \left\| \begin{pmatrix} X_{11} + x_1 \\ X_{12} \end{pmatrix} \right\| \leq 2x_1, \left\| \begin{pmatrix} X_{12} + x_2 \\ X_{22} \end{pmatrix} \right\| \leq 2x_2. \end{array}$$

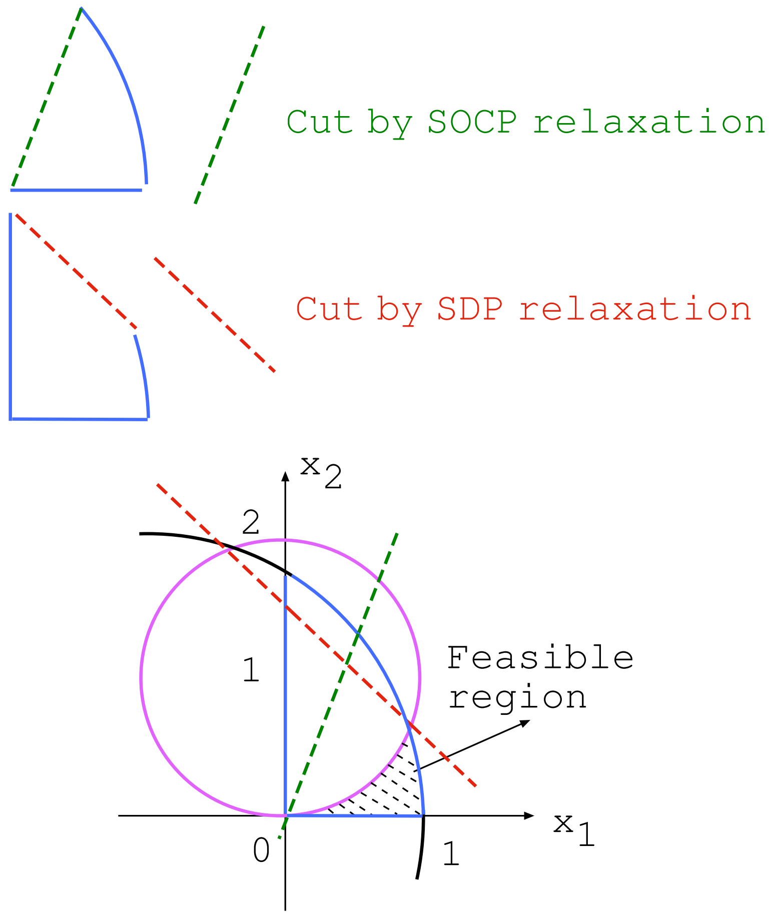
$$\begin{array}{ll} \text{max.} & -2x_1 + x_2 \qquad \text{— SDP} \\ \text{sub.to} & x_1 \geq 0,\ x_2 \geq 0, X_{11} + X_{22} - 2x_2 \geq 0, \\ & \left\| \begin{pmatrix} x_1 + 1 \\ x_2 \end{pmatrix} \right\| \leq 2, \begin{pmatrix} 1 & x_1 & x_2 \\ x_1 & X_{11} & X_{12} \\ x_2 & X_{12} & X_{22)} \end{pmatrix} \succeq O. \end{array}$$

Illustrative example again — 3

Original problem: max. $-2x_1 + x_2$

sub.to $x_1 \geq 0,\ x_2 \geq 0,\ x_1^2 + x_2^2 - 2x_2 \geq 0,$

$$\left\| \begin{pmatrix} x_1 + 1 \\ x_2 \end{pmatrix} \right\| \leq 2 \text{ (SOCP constraint)},$$



Some examples of valid constraints

- **Universally valid constraints.**

(a) SDP type:

$$u(x)^T u(x) = \begin{pmatrix} 1 & x_1 & x_2 & x_1^2 & x_1x_2 & x_2^2 \\ x_1 & x_1^2 & x_1x_2 & x_1^3 & x_1^2x_2 & x_1x_2^2 \\ x_2 & x_1x_2 & x_2^2 & x_1^2x_2 & x_1x_2^2 & x_2^3 \\ x_1^2 & x_1^3 & x_1^2x_2 & x_1^4 & x_1^3x_2 & x_1^2x_2^2 \\ x_1x_2 & x_1^2x_2 & x_1x_2^2 & x_1^3x_2 & x_1^2x_2^2 & x_1x_2^3 \\ x_2^2 & x_1x_2^2 & x_2^3 & x_1^2x_2^2 & x_1x_2^3 & x_2^4 \end{pmatrix} \succeq O,$$

where $u(x) = \begin{pmatrix} 1 & x_1 & x_2 & x_1^2 & x_1x_2 & x_2^2 \end{pmatrix}$

More generally, take a row vector consisting of a basis of the polynomials in $x_1, \ldots, x_n$ **with degree** ℓ **for** $u(x)$ **. [Lasserre'01].**

(b) SOCP (Second-Order Cone Programming) type:

$$\forall\ f_1, f_2 : \mathbb{R}^n \to \mathbb{R},\ \left\| \begin{pmatrix} f_1(x)^2 - f_2(x)^2 \\ 2f_1(x)f_2(x) \end{pmatrix} \right\| \le f_1(x)^2 + f_2(x)^2$$

- **Deriving valid constraints, "multiplication" of valid constraints:**

original constraints		new constraints
$\mathbb{R} \ni f(x) \geq 0,\ \mathbb{R} \ni g(x) \geq 0$	$\Rightarrow$	$f(x)g(x) \geq 0$ [Sherali et.al'92]
$f(x) \geq 0,\ G(x) \succeq O$	$\Rightarrow$	$f(x)G(x) \succeq 0$ [Lasserre'01]

$$F(x) \succeq O,\ G(x) \succeq O \Rightarrow F(x) \otimes G(x) \succeq 0 \text{ (Kronecker product)}$$

$$\left.\begin{array}{l} \|f(x)\| \leq f_0(x),\ f(x) \in \mathbb{R}^\ell \\ \|g(x)\| \leq g_0(x),\ g(x) \in \mathbb{R}^\ell \end{array}\right\} \Rightarrow \|f(x) \circ g(x)\| \leq f_0(x) g_0(x)$$

(SOCP constraints) (component-wise product)

5. Basic theory

POP: max. $f_0(x)$ sub.to $f(x) \in \mathcal{K}$, where

$\mathcal{K}$: a closed convex cone in $\mathbb{R}^m$, $f(x) \equiv (f_1(x), \ldots, f_m(x))$.

$\Downarrow$ **Linearization** — Keep the linear terms, but replace each
$\Downarrow$ nonlinear term by a single independent variable.

LOP: max. $F_0(x, y)$ sub.to $F(x, y) \in \mathcal{K}$,

where y denotes a new variable vector whose elements correspond to nonlinear terms appeared in the polynomials $f_j(x)$ $(j = 0, \ldots, m)$.

Lagrangian funct: $L(x, v) \equiv f_0(x) + \langle v, f(x) \rangle$ **for** $\forall x \in \mathbb{R}^n, v \in \mathcal{K}^*$

Under the Slater condition ($\exists x;\ f(x) \in$ **int** $\mathcal{K}$), **if** $\bar{\zeta}$ **is the opt. value of LOP then there exists** $\bar{v} \in \mathcal{K}^*$ **satisfying** $L(x, \bar{v}) = \bar{\zeta}$ **for** $\forall x \in \mathbb{R}^n$**.**

$$\begin{aligned} \textbf{Hence } \bar{\zeta} &= \max\{L(x, \bar{v}) : x \in \mathbb{R}^n\} \ \textbf{(a Lagrangian relaxation)} \\ &\geq \min_{v \in \mathcal{K}^*} \max\{L(x, v) : x \in \mathbb{R}^n\} \ \textbf{(Lagrangian dual relaxation)} \end{aligned}$$

Characterization of the projected feasible region of LOP onto $\mathbb{R}^n$:

$$\widehat{\mathcal{F}} \equiv \{x \in \mathbb{R}^n : F(x, y) \in \mathcal{K} \text{ for some } y\}$$

Define $\mathcal{L} \equiv \{v \in \mathcal{K}^* : \langle v, f(x) \rangle$ **is linear in** $x \in \mathbb{R}^n\}$ **and**

$$\widetilde{\mathcal{F}} \equiv \{x \in \mathbb{R}^n : \langle v, f(x) \rangle \geq 0 \text{ for every } v \in \mathcal{L}\}$$

"the set of linear consequences of $f(x) \in \mathcal{K}$" .

Then $\widehat{\mathcal{F}} \subseteq \widetilde{\mathcal{F}}$**, and (the closure of** $\widehat{\mathcal{F}}$**)** $= \widetilde{\mathcal{F}}$ **under** $\exists x;\ f(x) \in \text{ int } \mathcal{K}$**.**

6. Concluding remarks

The framework proposed in this talk for convex relaxation is quite general.

But we need to investigate various issues.

- **Effectiveness — How do we generate better bounds?**
- **Low cost — Resulting relaxed problems need to be solved cheaply**
- **How to combine this framework with other methods like the branch-and-bound method**
- **Parallel computation?**