Slater's condition: proof

∃ a strictly feasible point ($Ax = b$, $g(x) < 0$), problem convex

- $p^* = \inf_x f(x)$ s.t. $Ax = b$, $g(x) \le 0$

  m ineq.

  n equalities

  e.g., $\inf x^2$ s.t. $e^{x+2} - 3 \le 0$

- $\tilde{A} = \{ (u, v, t) \mid \exists x.\ u = Ax - b \quad v \ge g(x) \quad t \ge f(x) \}$

  $\mathbb{R}^n$ $\mathbb{R}^m$ $\mathbb{R}$

  e.g., $\tilde{A} = \{ (v, t) \mid v \ge e^{x+2} - 3,\ t \ge x^2 \}$

  $\subset \mathbb{R}^n$

Picture of set $\tilde{A}$

$L(y,z) = \inf_x f(x) + y^T(Ax-b) + z^T g(x)$

$= \inf_{(u,v,t)\in\hat{A}} t + y^T u + z^T v$

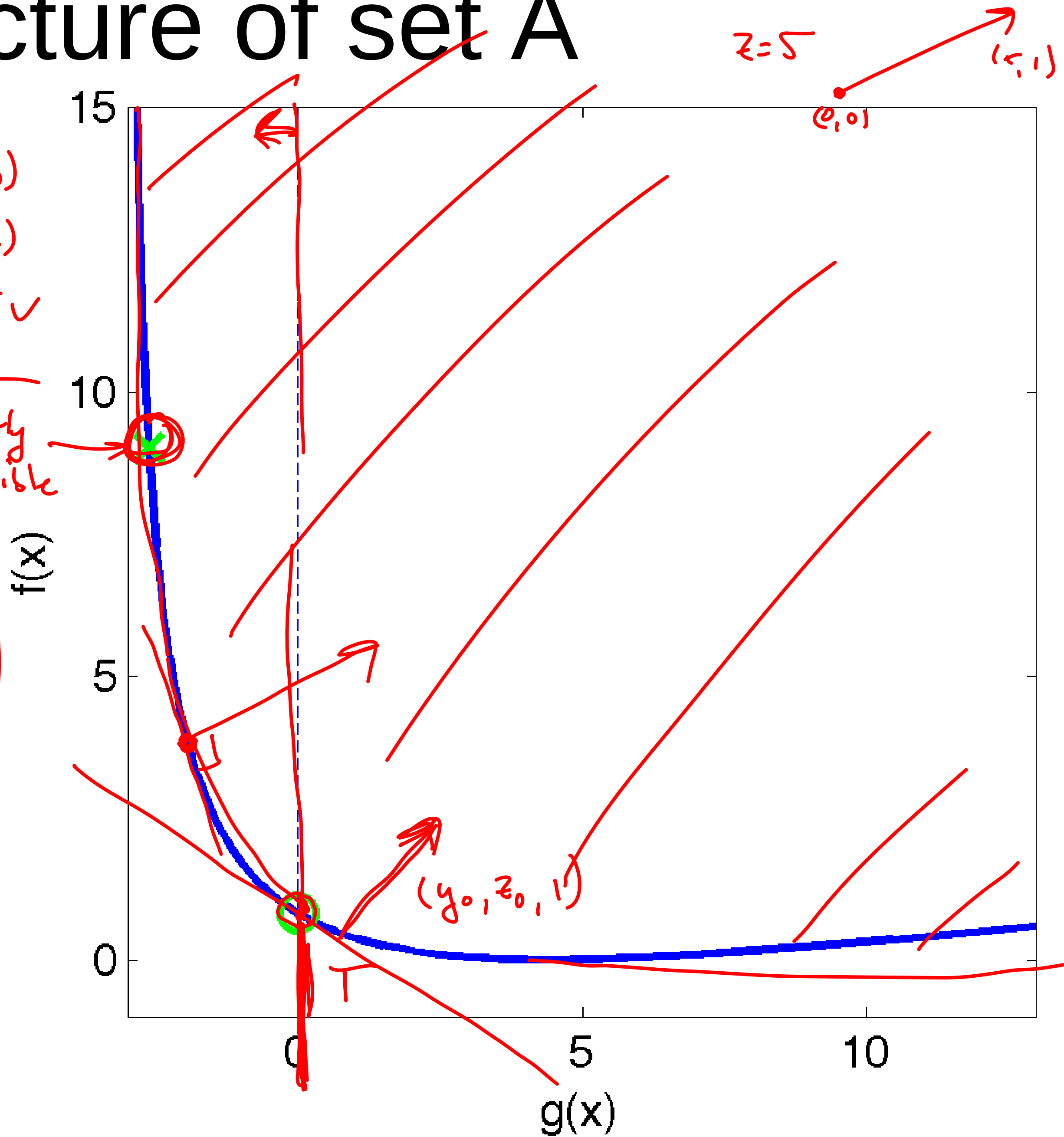
$= \inf_{(u,v,t)\in\tilde{A}} \begin{pmatrix} y \\ z \\ 1 \end{pmatrix} \cdot \begin{pmatrix} u \\ v \\ t \end{pmatrix}$

strictly feasible

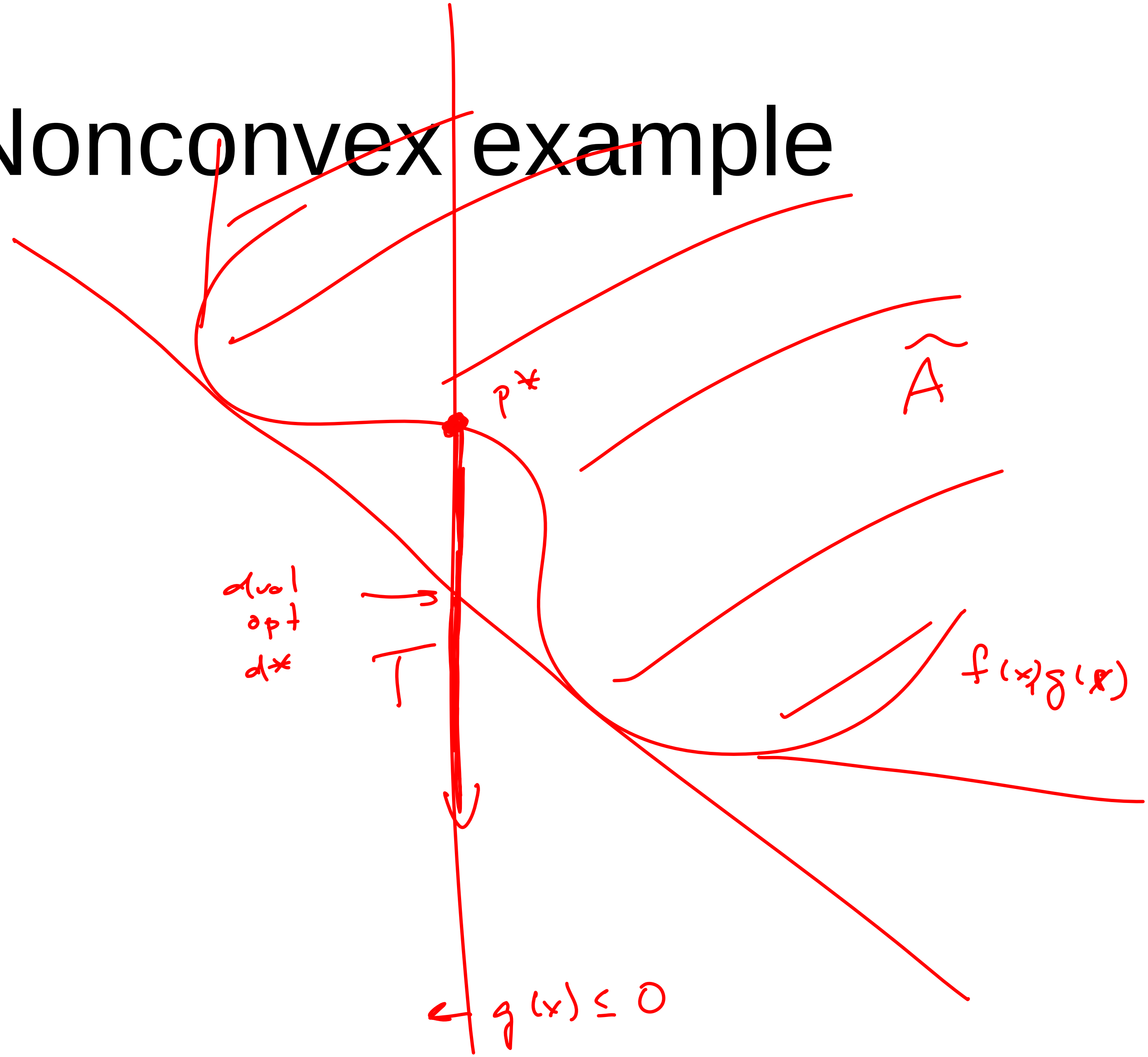
$T = \{(0,0,t) \mid t < p^*\}$

$L(y_0,z_0) = \inf_{(u,v,t)\in\tilde{A}} \underline{\begin{pmatrix} u \\ v \\ t \end{pmatrix} \cdot \begin{pmatrix} y_0 \\ z_0 \\ 1 \end{pmatrix}}$

$= p^*$



Nonconvex example



Interpretations

Suppose x^*, y^*, z^* optimal

$Ax = b + \epsilon$

$g(x) \le \mu$

$$L(x, y, z) = f(x) + y^T(Ax–b) + z^Tg(x)$$

- Prices or sensitivity analysis

$L(x^*, y^*, z^*) = f(x) + y^T(Ax - b - \epsilon) + z^T(g(x) - \mu)$

$\Delta L = -y^T\epsilon - z^T\mu$

- Certificate of optimality

Optimality conditions

$\sum_i z_i g_i(x)$

- $L(x, y, z) = f(x) + y^T(Ax - b) + z^T g(x)$
- Suppose strong duality, (x^*, y^*, z^*) optimal

(1) $Ax^* = b$ (2) $g(x^*) \leq 0$ (3) $z^* \geq 0$

$$x^* = \arg\min_x L(x, y^*, z^*) \iff \underbrace{0 \in \partial f(x^*) + A^T y^* + \sum_i z_i^* \partial g_i(x^*)}_{(4)}$$

$$f(x^*) = L(y^*, z^*) = \inf_x L(x, y^*, z^*)$$
$$\leq L(x^*, y^*, z^*) = f(x^*) + z^{*T} g(x^*) \leq f(x^*)$$

$$\Rightarrow z^{*T} g(x^*) = 0 \quad (5)$$

(1) - (5): KKT conditions

Karush Kuhn Tucker

KKT $\Longleftrightarrow$ primal, dual opt ; strong duality

Optimality conditions

- $L(x, y, z) = f(x) + \underline{y^T(Ax-b)} + \underline{z^T g(x)}$
- Suppose (x, y, z) satisfy KKT:

$$\underline{Ax = b} \qquad \underline{g(x) \le 0}$$

x is feasible

y, z feasible

$$\underline{z \ge 0} \qquad \underline{z^T g(x) = 0}$$

$L(x, y, z) = f(x)$

y, z opt

$L(y, z) = f(x)$

$$0 \in \partial f(x) + A^T y + \textstyle\sum_i z_i \partial g_i(x)$$

$\Longleftrightarrow$ $x = \arg\min_{x'} L(x', y, z)$

$f(x) = L(y, z)$ strong duality

y, z dual opt

x primal opt

Using KKT

- Can often use KKT to go from primal to dual optimum (or vice versa)
- E.g., in SVM:

  $\alpha_i > 0 \iff y_i(x_i^T w + b) = 1$

- Means $b = y_i - x_i^T w$ for any such i
  - typically, average a few in case of roundoff

Set duality

$t = 0$

$\Rightarrow x^T y \geq 0 \;\; \forall x \in C$

$\Rightarrow 1 \geq x^T(-y)$

- Let C be a set with $0 \in \text{conv}(C)$
- $C^* = \{\, y \mid x^T y \leq 1 \text{ for all } x \in C \,\}$
- Let $K = \{\, (x, s) \mid x \in sC \,\}$ $\quad s \geq 0 \,\}$
- $K^* = \{\, (y, t) \mid x^T y + st \geq 0 \text{ for all } \cancel{x}\,(x,s) \in K \,\}$

$t \geq 0 \;\; \forall (y,t) \in K^* \quad \Leftarrow \quad (0,0,\ldots,0,1) \in K \quad \Rightarrow \quad x^T y + t \geq 0 \;\; \forall x \in C$

suppose $t > 0$

$x^T \frac{y}{t} + 1 \geq 0 \qquad 1 \geq x^T\left(-\frac{y}{t}\right) \;\; \forall x \in C$

$\Leftrightarrow \left(-\frac{y}{t}\right) \in C^*$

What is set duality good for?

- Related to norm duality
- Useful for helping visualize cones

Duality of norms

$\max_{x \,|\, \|x\|_1 \leq 1} y^T x = \max_i |y_i|$

- Dual norm definition

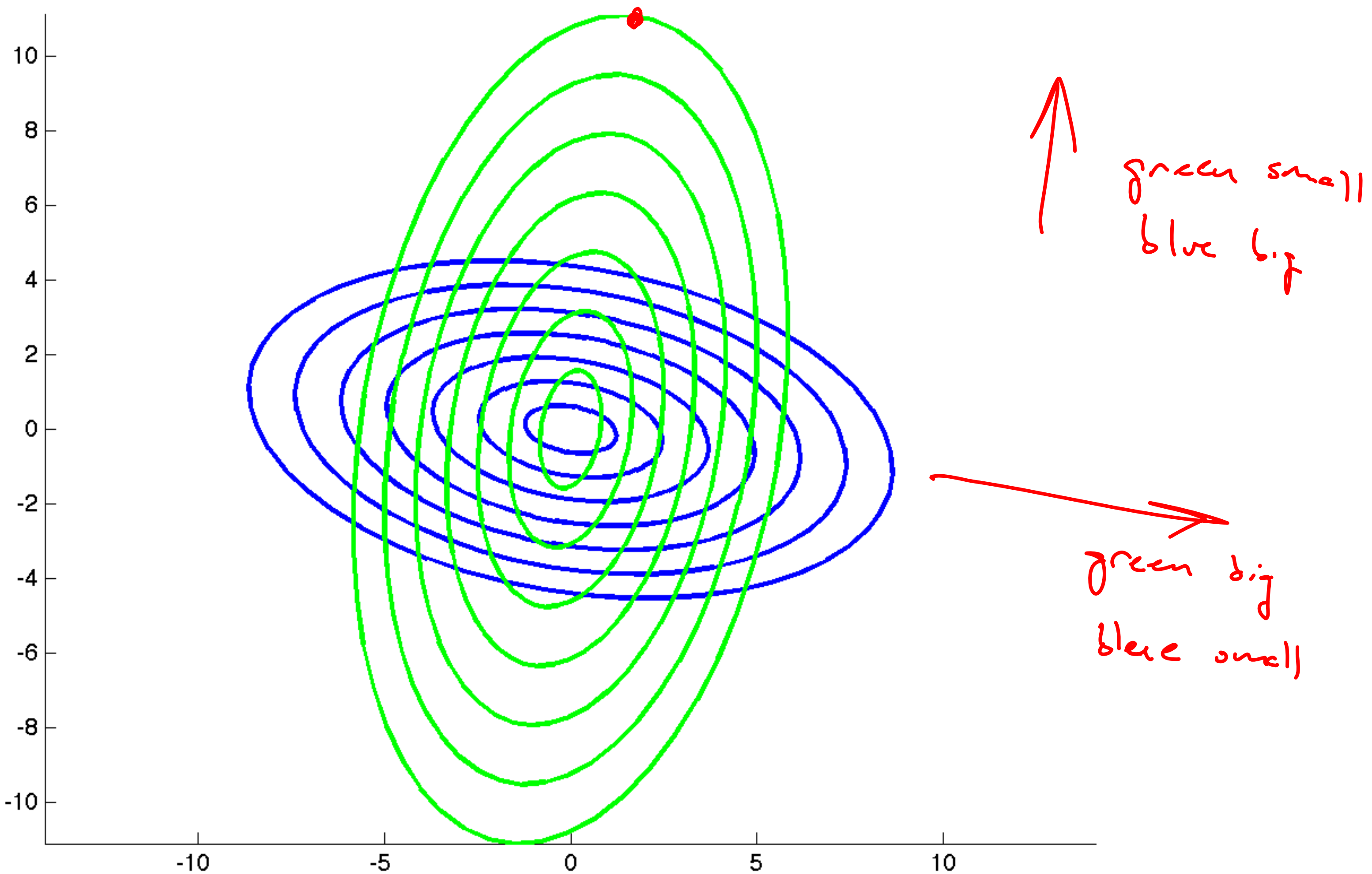
  $\|y\|_* = \max_{x \,|\, \|x\| \leq 1} y^T x$

- Motivation: Holder's inequality

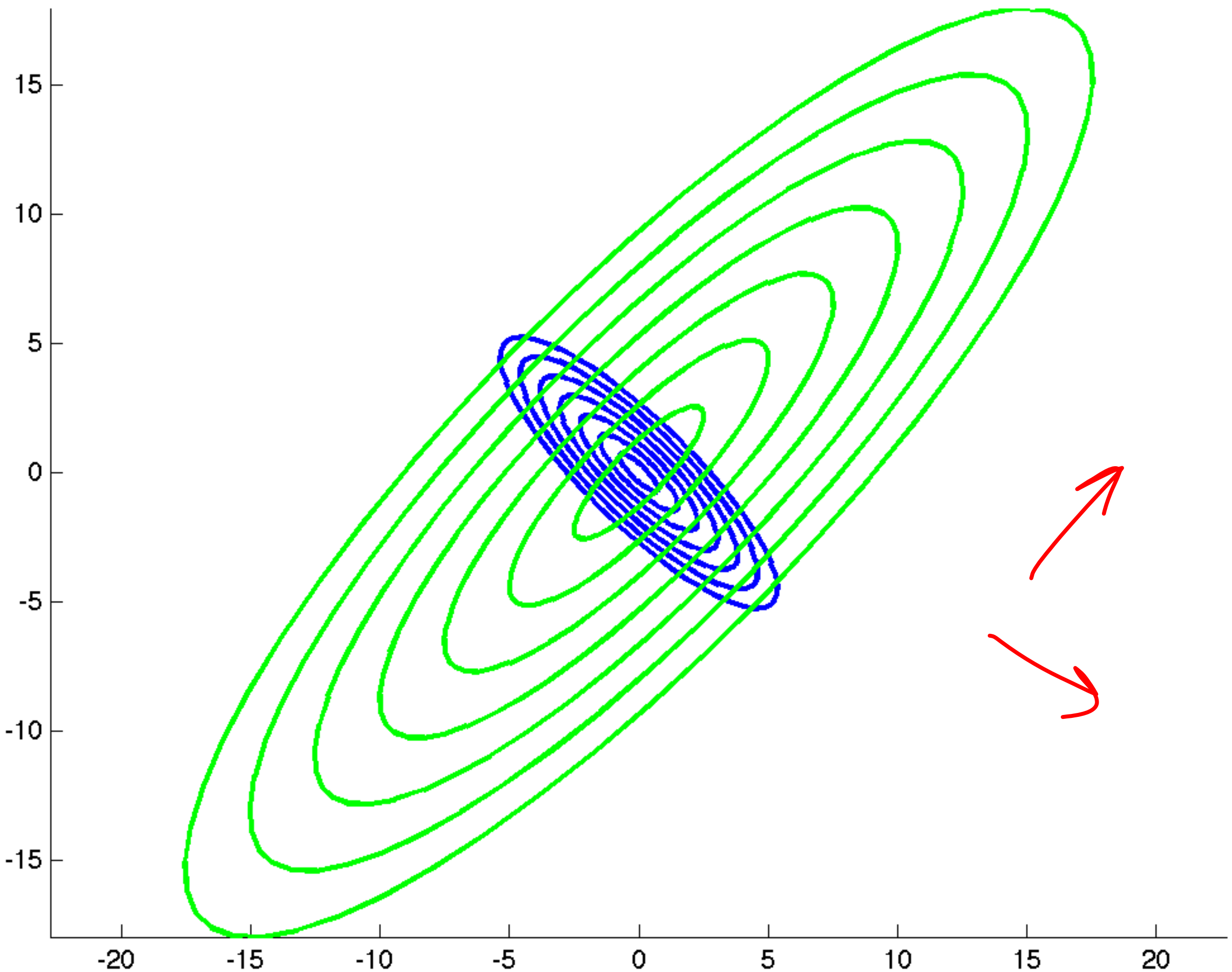
  $x^T y \leq \|x\| \, \|y\|_*$

  $x^T y = \|x\| \underbrace{\frac{x}{\|x\|}}_{\text{norm } 1} \cdot y \leq \|x\| \max_{x \,|\, \|x\| \leq 1} x \cdot y = \|x\| \, \|y\|_*$

Dual norm examples

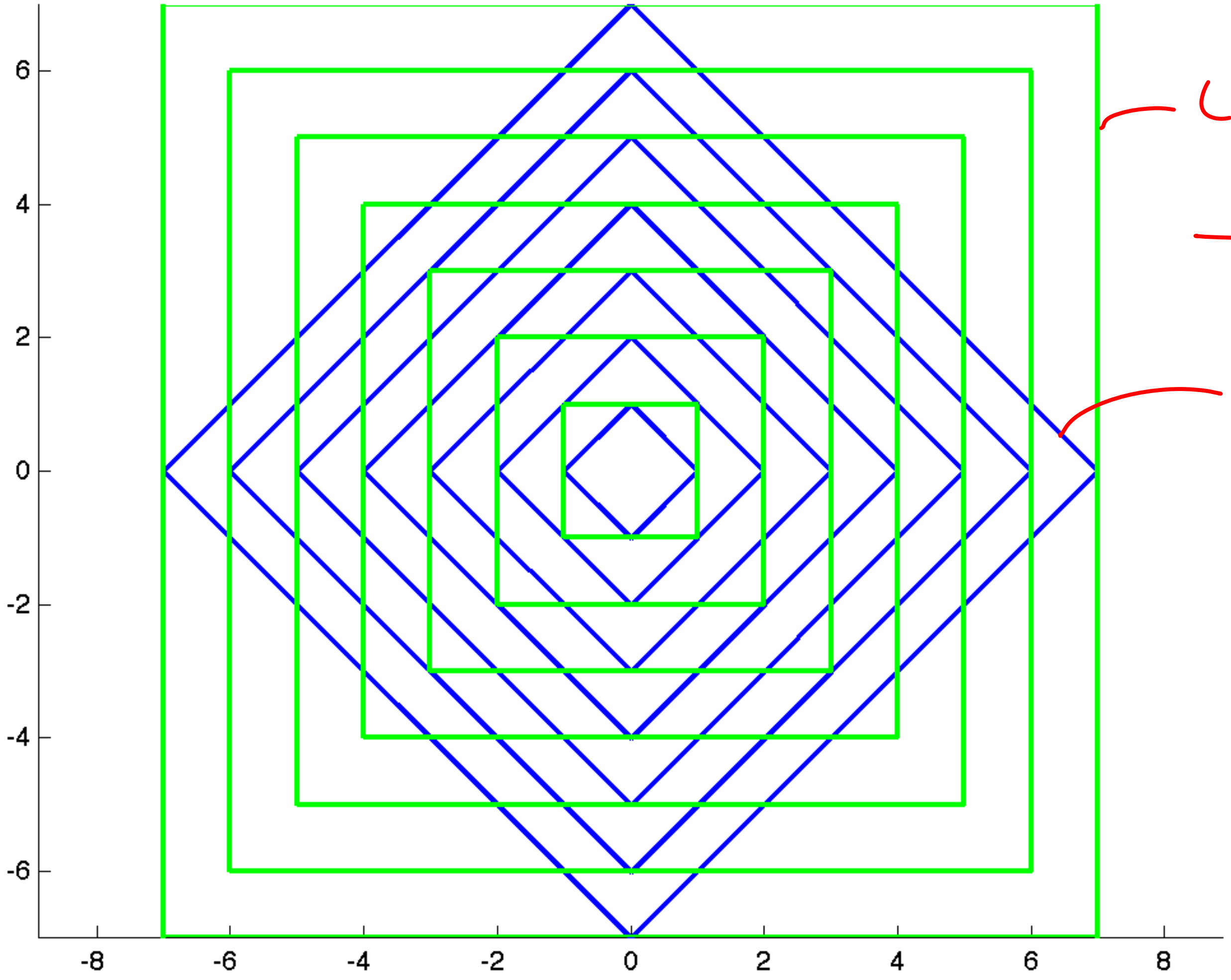


Dual norm examples



Dual norm examples

L_∞ small
L_1 big

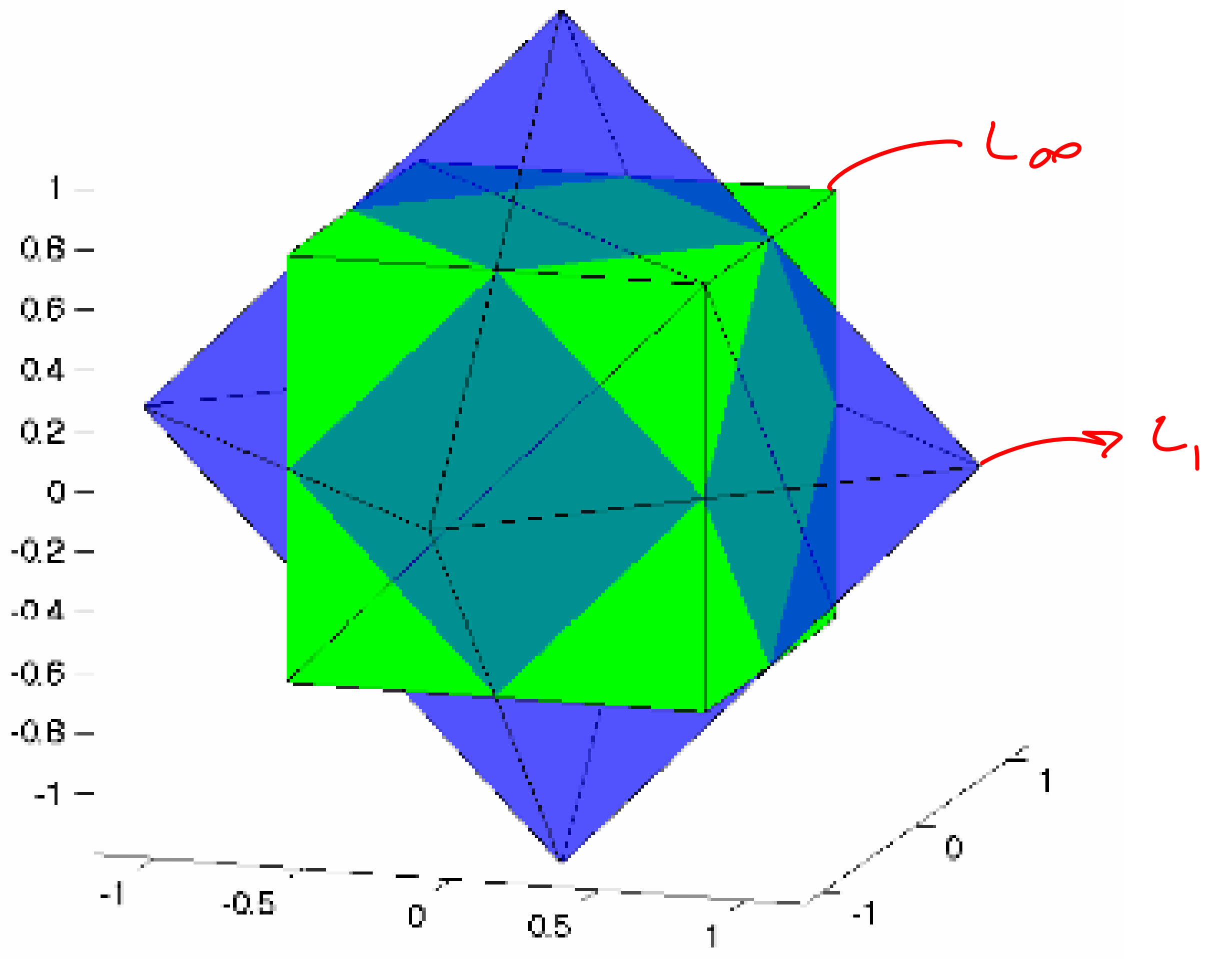


L_∞

L_1 sm
L_∞ big

L_1

Cuboctahedron



$||y||_*$ is a norm

- $||y||_* \geq 0$:
- $||ky||_* = |k| \, ||y||_*$:
- $||y||_* = 0$ iff $y = 0$:
- $||y_1+y_2||_* \leq ||y_1||_* + ||y_2||_*$

Dual-norm balls

- $\{ y \mid \|y\|_* \leq 1 \} =$ $\{ y \mid y^T x \leq 1 \ \forall x.\ \|x\| \leq 1 \}$

  $= \{ x \mid \|x\| \leq 1 \}^*$

- Duality of norms: $\|x\|_{**} = \|x\|$

Visualizing cones

rotate $u \rightarrow \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$

- Suppose we have some weird cone in high dimensions (say, $K = S_+$)
- Often easy to get a vector $\underline{u}$ in $\underline{K} \cap \underline{K^*}$
  - e.g., $\underline{I} \in \underline{S_+}$, $\underline{I} \in \underline{S_+}^* = \underline{S_+}$
- Plot $K \cap \{ u^T x = 1 \}$ and $K^* \cap \{ u^T x = 1 \}$ instead of K, K^*
  - saves a dimension

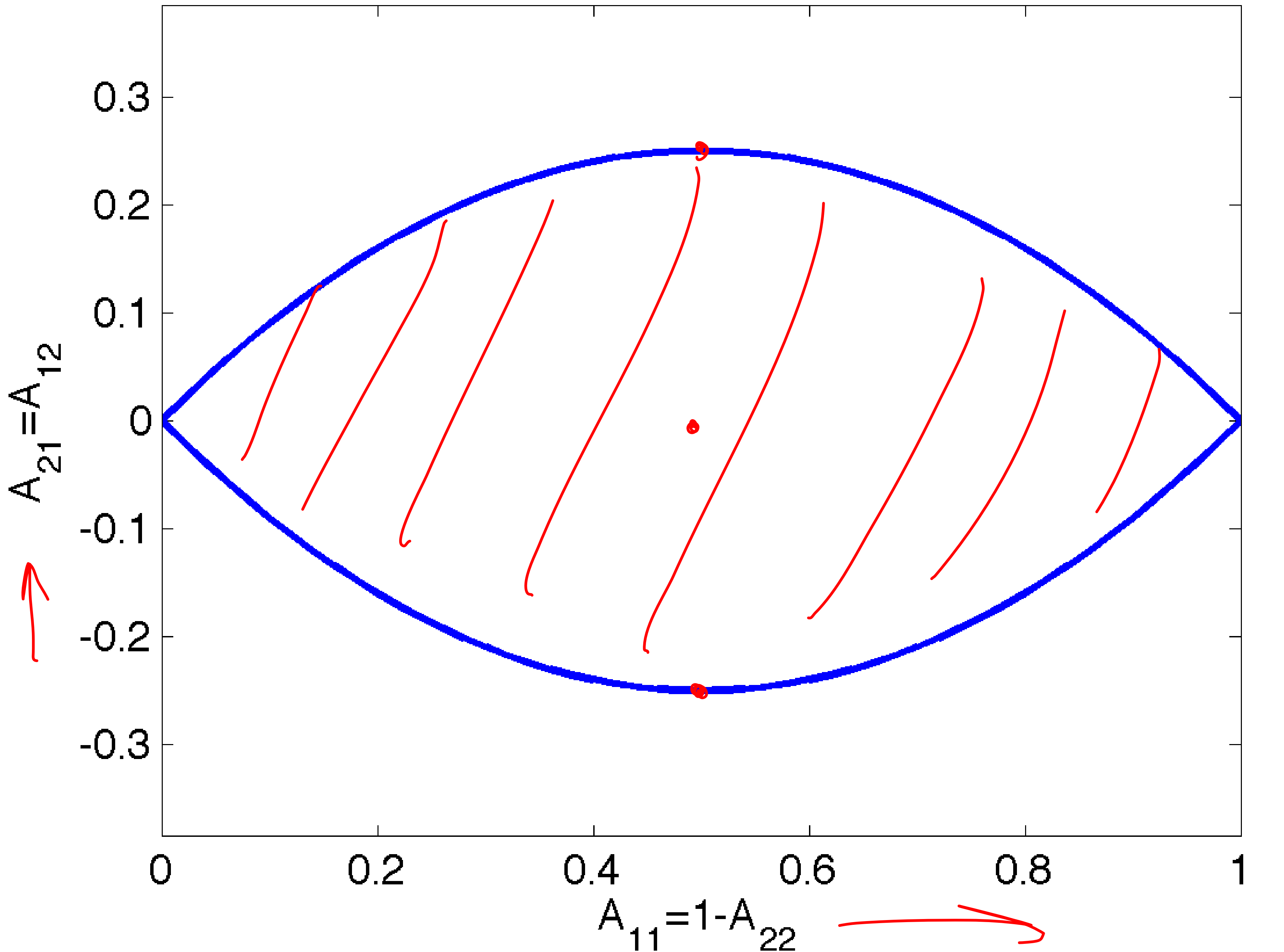
Visualizing S_+

$$X = \begin{pmatrix} a & c \\ c & b \end{pmatrix}$$

- Say, 2 x 2 symmetric matrices
- Add constraint $\underline{tr(X^T I) = 1}$

$tr(X) = 1 \quad a + b = 1$

- Result: a 2D set

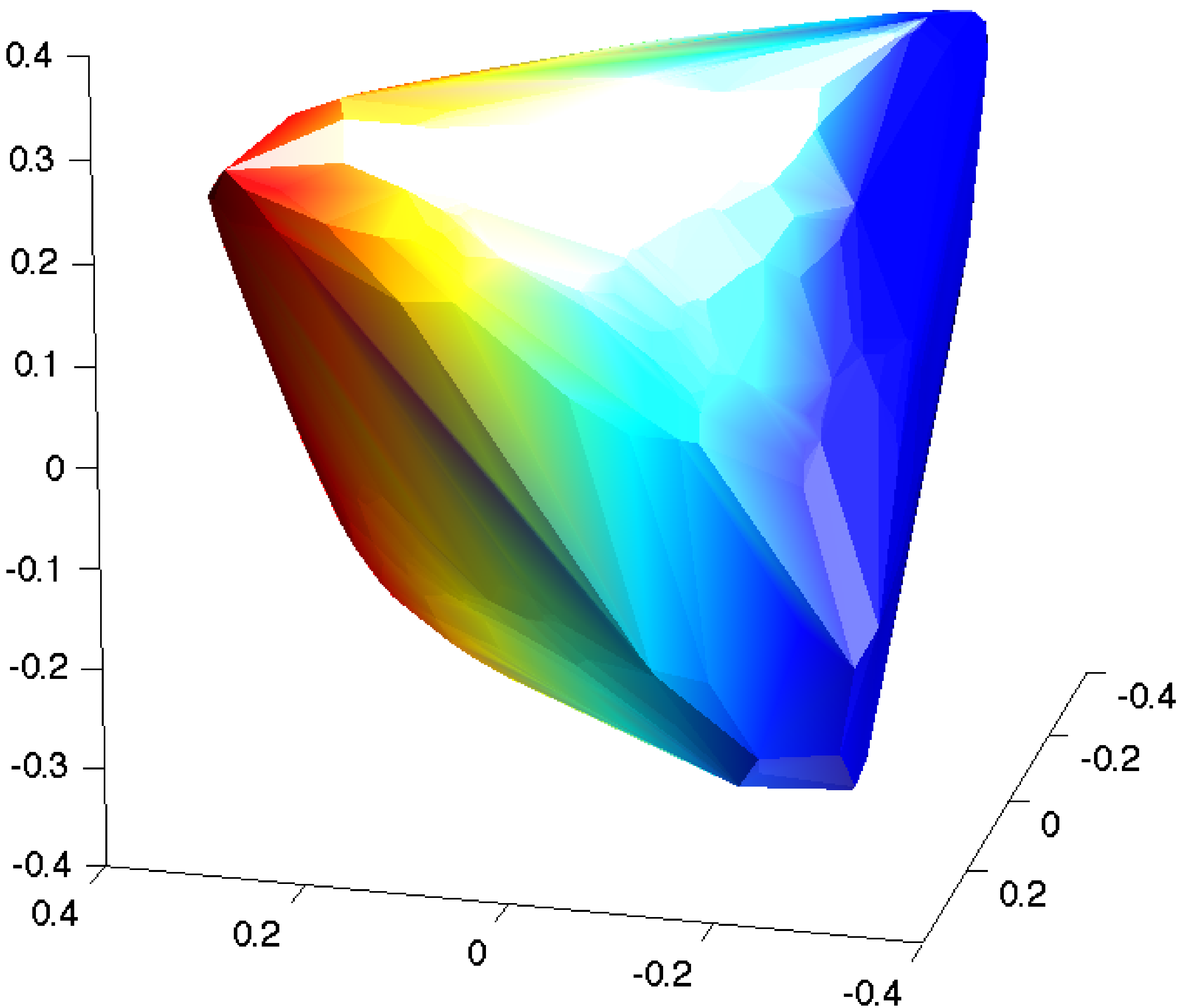


0.3
0.2
0.1
0
-0.1
-0.2
-0.3
$A_{21}=A_{12}$
0
0.2
0.4
0.6
0.8
1
$A_{11}=1-A_{22}$

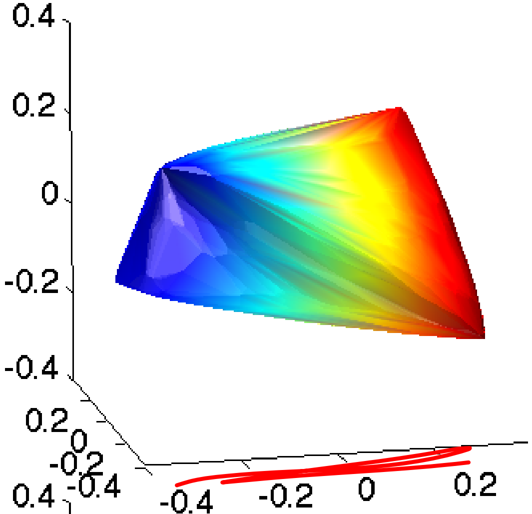
What about 3 x 3?

$$\begin{matrix} a & d & f \\ d & b & e \\ f & e & c \end{matrix}$$

- 6 parameters in raw form
- Still 5 after tr(X)=1 $a+b+c=1$
- Try setting entire diagonal to 1/3
  - plot off-diagonal elements d, e, f

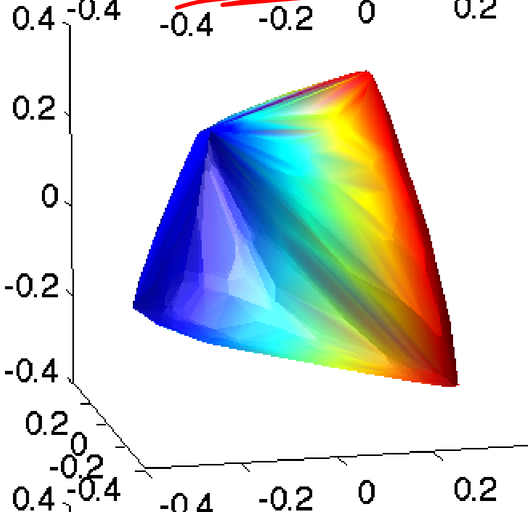

0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
0.4
0.2
0
-0.2
-0.4
-0.4
-0.2
0
0.2

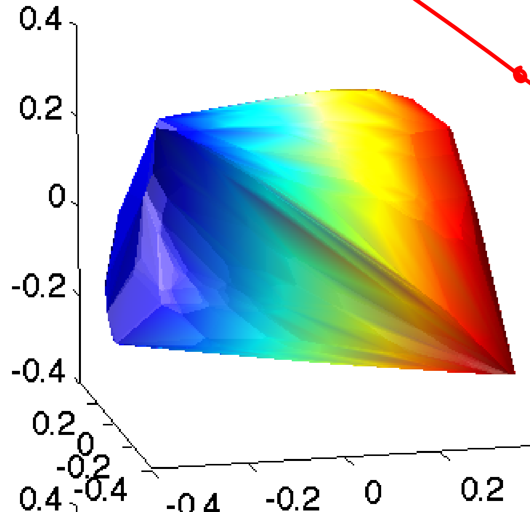
Visualizing 3*3 symmetric semidefinite matrices



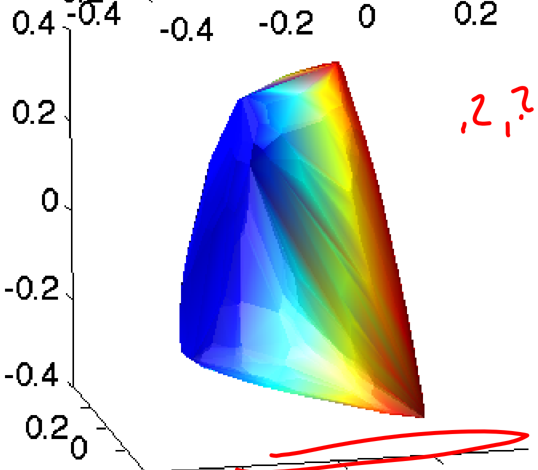
.6, .2, .2

.4, .4, .2 on diagonal

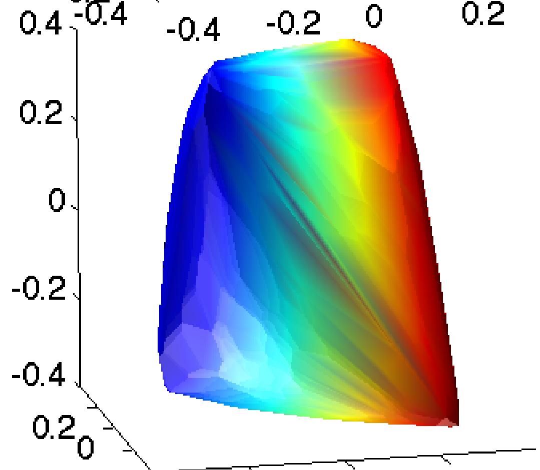


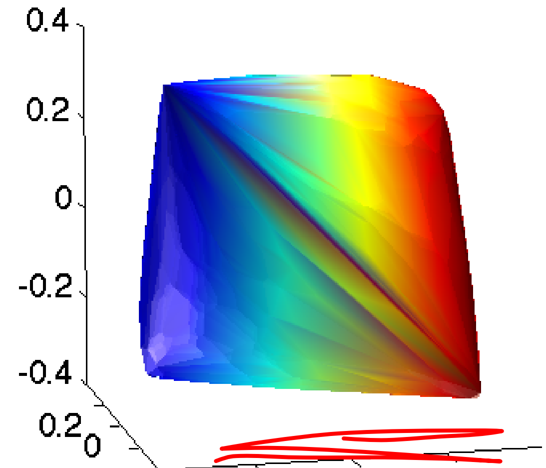


.2, .6, .2



.2, .2, .6





Multi-criterion optimization

- Ordinary feasible region

  $g_i(x) \le 0 \quad \forall i$

- Indecisive optimizer: wants all of

  $\min f_1(x) \quad \min f_2(x) \quad \min f_3(x) \quad \ldots$

Buying the perfect car

	\$K	0-60	MPG
Porshe	150	4 s	15
BMW	50	6 s	20
Ford	20	8 s	20
Prius	25	9 s	40
Hummer	50	10 s	10

$f_1(x) = \text{price} \Rightarrow \text{Ford}$

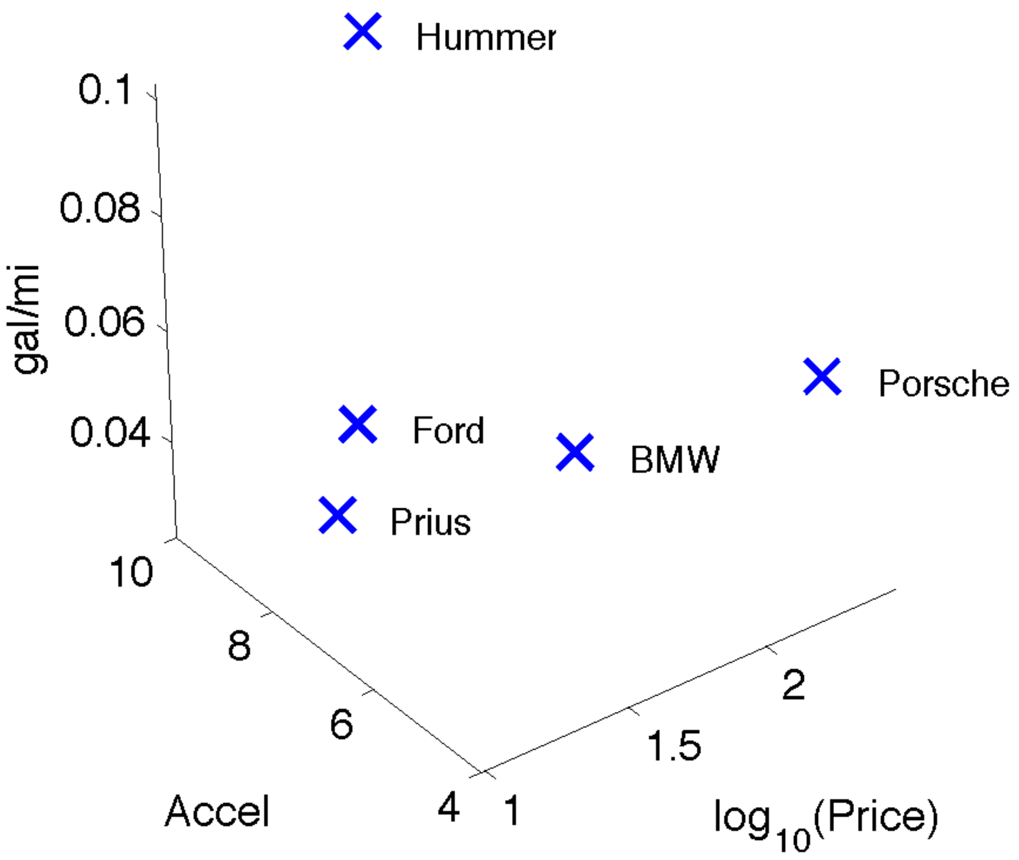
$f_2(x) = \text{accel} \Rightarrow \text{Porsche}$

$f_3(x) = -\text{MPG} \Rightarrow \text{Prius}$

$f_2(x) + \frac{f_1(x)}{20} + \frac{f_2(x)}{20} \Rightarrow \text{BMW}$

$\max\ h_1(f_1(x)) + h_2(f_2(x)) + h_3(f_3(x))$

Pareto optimality



x* Pareto optimal =

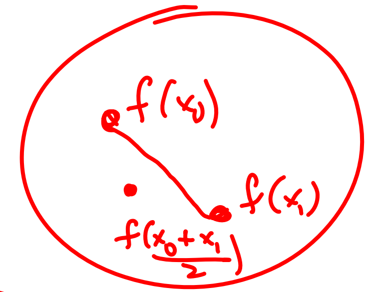
x^* could be optimal for monotone u_1, u_2, u_3

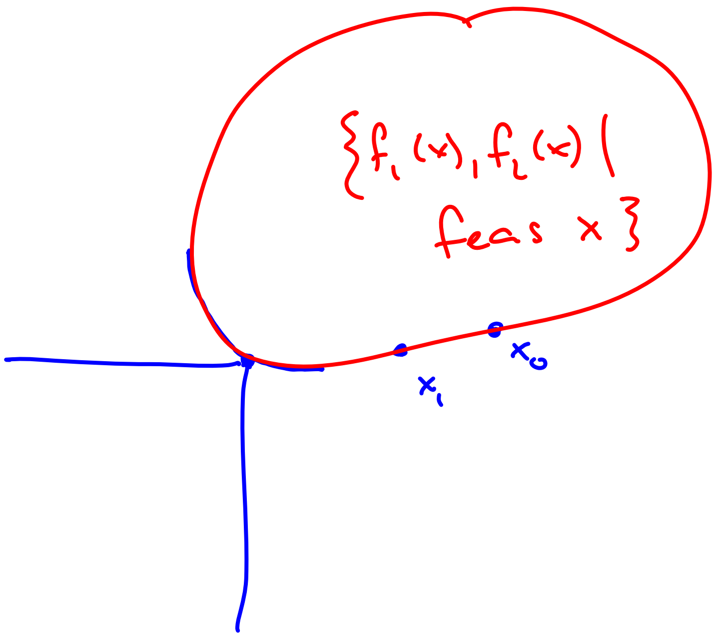
(or lim x optimal for monotone u_1 u_2 u_3)

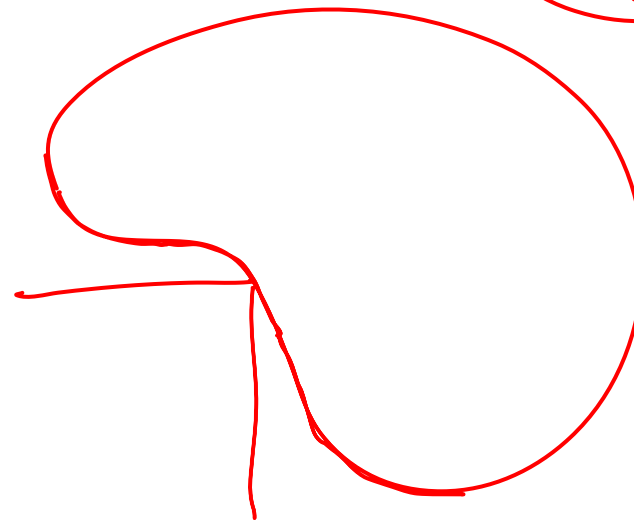
$\Longleftrightarrow$ x^* not strictly dominated

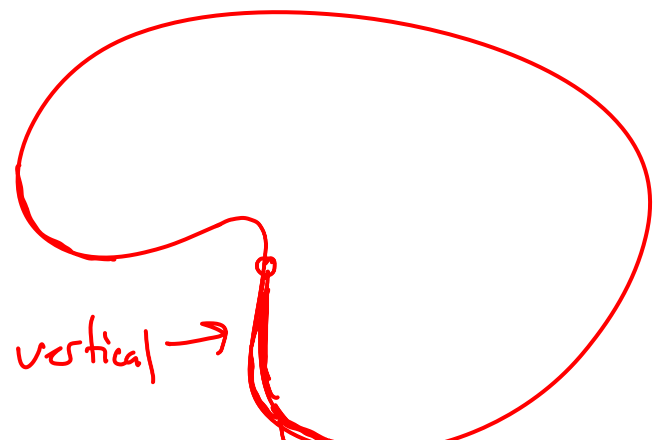
$$\{(u,v) \mid u \geq f_1(x),\ v \geq f_2(x)\}$$

Pareto examples









Scalarization

- To find Pareto optima of convex problem:

$$\min \sum_i w_i f_i(x) \qquad w_i > 0 \qquad \sum_i w_i = 1$$

See Boyd