5. Convexity

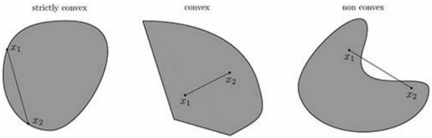
Convex Problems, Lagrangian Basics and Duality

Sun Woo Lim

Jan 23, 2021

Convex Set

Definition of a Convex Set



$K \subseteq \mathbb{R}^n$ is a **convex set** if $\forall x_1, x_2 \in K$, $\forall \lambda \in [0, 1]$, $\lambda x_1 + (1 - \lambda)x_2 \in K$
"The Convex Combination also is in the set".

✓ Which of the following is / are convex?

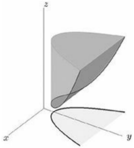
1. **Empty Set** ϕ
2. **A set of single point** $\{x_0\}$
3. $\{z \in R^n : ||z - z_0||_2 \leq \epsilon\}$ **for some** $\epsilon > 0$
4. $\{z \in R^n : ||z - z_0||_2 = \epsilon\}$ **for some** $\epsilon > 0$
5. $[-2, -1] \cup [1, 2]$

1,2,3 are convex and 4,5 are not.

Operations preserving Convexity of a set

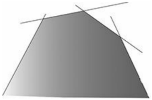
Operations preserving Convexity of a set

1. **Intersection of convex sets**
2. **Hyperplane** $\{x|a^T x - b = 0\}$ **and Half Spaces** $\{x|a^T x - b \leq 0\}$ **and** $\{x|a^T x - b > 0\}$
3. **Projection of a convex set onto a hyperplane**



4. **"Convex Hull of** A**"** $Co(A) := \{\sum_{i=1}^{m} \lambda_i x_i | x_i \in A, \lambda_i \geq 0, \forall i, \sum \lambda_i = 1\}$: Smallest convex set containing A
5. **"Conic Hull of** A**" :=**$\{\sum_{i=1}^{m} \lambda_i x_i | x_i \in A, \lambda_i \geq 0, \forall i\}$
6. **"Affine Hull of** A**" :=**$\{\sum_{i=1}^{m} \lambda_i x_i | x_i \in A, \sum \lambda_i = 1\}$

Q) Why is a polyhedron convex?



Operations preserving Convexity of a set : Proof

Claim 1) Intersection of convex sets is convex.
Suppose K_1 and K_2 are convex sets in $\mathbb{R}^n$.
Pick $x_1, x_2 \in K_1 \cap K_2, \lambda \in [0,1]$.
Need to show $\lambda x_1 + (1-\lambda)x_2 \in K_1 \cap K_2$.

Since $x_1, x_2 \in K_1$ and K_1 is convex, $\lambda x_1 + (1-\lambda)x_2 \in K_1$.
Since $x_1, x_2 \in K_2$ and K_2 is convex, $\lambda x_1 + (1-\lambda)x_2 \in K_2$.
Thus, $\lambda x_1 + (1-\lambda)x_2 \in K_1 \cap K_2$

Convex Functions

Convex / Concave Functions defined on the domain of a convex set

For $f : R^n \to R$ defined for $x \in dom(f)$ and assume f takes ∞ outside the domain. For f defined on convex domain, f is convex if $\lambda f(x_1) + (1-\lambda) f(x_2) \geq f(\lambda x_1 + (1-\lambda) x_2)$ and concave if $\lambda f(x_1) + (1-\lambda) f(x_2) \leq f(\lambda x_1 + (1-\lambda) x_2)$

✓ **Properties of Convex Functions**

1. **Pointwise Supremum of convex sets is a convex function**
2. **Nonnegative linear combination of Convex Functions is a convex function**
3. $f : \mathbb{R}^n \to \mathbb{R}$ **is a convex function** $\leftrightarrow$ $epi(f)$ **is a convex set : "Epigraph Characterization of a Convex Function"**
   $epi(f) := \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : x \in dom(f), t \in \mathbb{R}, f(x) \leq t\}$
4. Jensen's Inequality : For a convex function f and a random variable X, $E(f(X)) \geq f(E(X))$
5. Local minimum of convex function is a global minimum and strictly convex function has at most one global minimum

✓ **Iff conditions for differentiable convex functions**

For f which has an open domain and differentiable on $dom(f)$,

1. **First order (gradient) condition for convexity**
   f convex $\leftrightarrow f(y) \geq f(x) + \nabla f(x)^T (y-x), \forall x, y \in dom(f)$.
2. **Second order (Hessian) condition for convexity**
   f convex $\leftrightarrow \nabla^2 f(x) \succeq 0, \forall x \in dom(f)$. "Hessian is PSD".

Convex Functions : Proof

Claim 2) Pointwise supremum of convex functions is convex

Assume that the domain of the supremum is nonempty.
Given f_α, $\alpha \in J$, each f_α is assumed to be convex, $_\alpha : \mathbb{R}^n \to \mathbb{R}$.
Define $g(x) = sup_{\alpha \in J} f_\alpha(x)$.
Define $dom(g) = \cap_{\alpha \in J} dom(f_\alpha)$.
$\forall x_1, x_2 \in dom(g)$, and $\forall \alpha \in [0,1]$, $f_\alpha(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f_\alpha(x_1) + (1-\lambda) f_\alpha(x_2) \leq \lambda g(x_1) + (1-\lambda) g(x_2)$.
Then, take supremum over $\alpha \in J$ on the LHS, leading to
$g(\lambda x_1 + (1-\lambda)x_2) \leq \lambda g(x_1) + (1-\lambda) g(x_2)$.

Convex Functions : Proof

Claim 3) f is convex $\leftrightarrow f(y) \geq f(x) + \nabla f(x)^T(y-x)$, for all $x, y \in dom(f)$: First Order Condition (FOC) for convexity

pf) skipped

Claim 4) f convex $\leftrightarrow \nabla^2 f(x) \geq 0, \forall x\ indom(f)$, assuming second order differentiability : Second Order Condition (SOC)

$\rightarrow$: Given $x, y \in dom(f)$, can write $f(x+\lambda(y-x)) = f(x) + \lambda \nabla f(x)^T(y-x) + \frac{\lambda^2}{2!}(y-x)^T \nabla^2 f(x)(y-x) + O(\lambda^3)$ by Taylor series expansion.
From this, conclude that letting $\lambda \rightarrow 0$, the Hessian must be PSD : $f(x+\lambda(y-x)) \geq f(x) + \lambda \nabla f(x)^T(y-x)$.

: $f(y) = f(x) + \nabla f(x)^T(y-x) + \frac{1}{2!}(y-x)^T \nabla^2 f(w)(y-x)$ a t some w on the line between x and y.
Thus, $f(y) \geq f(x) + \nabla f(x)^T(y-x)$, which is the FOC of convexity.

General Convex Optimization Problem

$min_x f_0(x)$: Objective Function
subject to (s.t.) $f_i(x) \leq 0, i = 1, 2, ...m,$: m inequality constraints
$h_i(x) = 0, i = 1, 2, ..., p$: p equality constraints

$\rightarrow$ This is a **Convex Optimization Problem** if $f_0, f_1, ..., f_m$ are convex functions and $h_1, ..., h_p$ are affine functions.

Terminologies

$\{x \in \{\cap_{i=0}^{m} dom(f_i) \cap \cap_{j=1}^{p} dom(h_i)\} | f_i(x) \leq 0, \forall i = 1, 2, ..., m, h_j(x) = 0, \forall j = 1, 2, ..., p\}$ is called a **Feasible Set**

An optimization problem having **an empty feasible set** is **infeasible**.

The **infimum** of the objective function over the **feasible set** is called the **primal optimal value**, denoted as $p*$.

If $\exists$ feasible x satisfying $f_0(x) = p*$, say x **attains** the optimum and x^* is called **primal optimal point**.

The set of feasible points at which the optimum is attained is called an **Optimal Set**

Constraints f_i or h_j is(are) **active** at feasible point x if $f_i(x) = 0$ or $h_j(x) = 0$ respectively. Else, they are **inactive** at x.

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a convex function with $dom(f)$. For $\alpha \in \mathbb{R}$, $S_\alpha := \{x \in \mathbb{R}^n | f(x) \leq \alpha\}$ is called a **Sublevel set** of f.

Atypical Optimization Problems

✓ An **Infeasible Problem** is an optimization problem having **an empty feasible set.**

ex) $min_x (x^2 + 3)$ s.t. $x \leq 0, -x + 2 \leq 0$

Here $p^* = \infty$, the problem is **infeasible**.

✓ Even a convex problem have $p^* = -\infty$.

ex) $min_x - log(x)$ s.t. $1 - x \leq 0$

The problem is **feasible** while the **optimal set** is empty.

✓ There is a problem having a **finite optimal value** where the **the optimal set is empty**.

ex) $inf_x (\frac{1}{x} + 2)$ s.t. $1 - x \leq 0$

Here $p^* = 2$, the problem is **feasible**.

The **the optimal set** is empty.

The optimal is **not attained**.

Indicator Functions

An **Indicator function** in this field is totally different from the indicator function in probability theory (1 if in the set, 0 if not in the set).

$$I_S(x) = \begin{cases} 0 & x \in S \\ \infty & x \notin S \end{cases}$$

✓ Two mostly used **indicator functions** are :

$$I_{\mathbb{R}\_}(x) = \begin{cases} 0 & x \leq 0 \\ \infty & x > 0 \end{cases} = sup_{z \geq 0} zx$$

$$I_{\{0\}}(x) = \begin{cases} 0 & x = 0 \\ \infty & x \neq 0 \end{cases} = sup_{z \in \mathbb{R}} zx$$

Apply two **indicator functions** to :

✓ $I_{\mathbb{R}\_}(f_i(x)) = sup_{\lambda_i \geq 0}\{\lambda_i f_i(x)\}$

✓ $I_{\{0\}}(h_j(x)) = sup_{\nu_j \in \mathbb{R}}\{\nu_j h_j(x)\}$

Later used in **the Lagrangian**.

Dual Norm

Dual Norm of an arbitrary Norm $||x||* := sup_{\{z\in\mathbb{R}^n:||z||\leq 1\}} z^T x$

✓ Note that every norm on $\mathbb{R}^n$ is a **convex function**.
pf) For all norm on $\mathbb{R}^n$, $\forall x_1, x_2 \in \mathbb{R}^n, \forall\lambda \in [0,1], ||\lambda x_1 + (1-\lambda)x_2|| \leq ||\lambda x_1|| + ||(1-\lambda)x_2|| = \lambda||x_1|| + (1-\lambda)||x_2||$

✓ Note that a **dual norm** is a **norm** so, is a **convex function**.
pf)

1. If $||x||^* = 0$, it means that $z^T x = 0$, $\forall z : ||z|| \leq 1$. $\therefore x = 0$. If $x \neq 0 \leftrightarrow ||x|| \neq 0$, since $z = \frac{x}{||x||}$, $||x||^* \geq \frac{||x||_2^2}{||x||} > 0$.
2. $||\alpha x||^* = max_{||z||\leq 1}|\alpha| z^T x = |\alpha|||x||^*$
3. Similarly check $||x_1 + x_2||^* \leq ||x_1||^* + ||x_2||^*$

✓ For $p = 1, 2, 3, ..., \infty$, dual of l_p norm is l_q norm for p, q satisfying $\frac{1}{p} + \frac{1}{q} = 1 \because$ using **Holder's Inequality.**

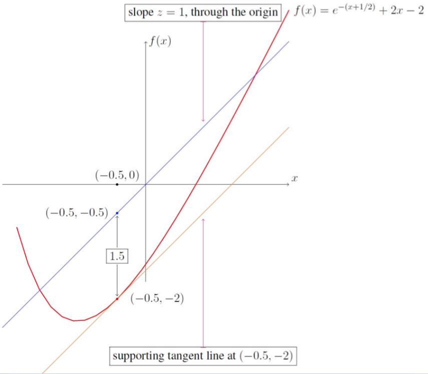
- dual of l_1 norm is l_∞ norm and dual of l_∞ norm is l_1 norm $||z||_1 = max_{u:||u||_\infty\leq 1} u^T z = ||z||_\infty^*$.
- dual of l_2 norm is l_2 norm : can prove this by **Cauchy-Schwarz Ineqaulity.** $||z||_2 = max_{u:||u||_2\leq 1} u^T z = ||z||_2^*$: **self dual**

Convex Conjugate of a Function

Let $f : \mathbb{R}^n \to \mathbb{R}$ **(need not be cvx ftn)** having a nonempty domain **(need not be cvx set)**.

$f^*(z) := sup_{x \in \mathbb{R}^n}(z^T x - f(x))$: **Convex Conjugate (Fenchel Conjugate)**

✓ Property 1) f^* is always **convex ftn** and **lower semicontinuous : epigraph is a closed set**.
✓ Property 2) If f is convex and **lower semicontinuous** them $f^{**} = f$



Convex Conjugate of a Function

Examples

1. $f(x) = e^x \to f^*(z) = zlnz - z$
2. $f(x) = a^T x + b \to$

$$f^*(z) = \begin{cases} -b & z = a \\ \infty & z \neq a \end{cases}$$

3. $f(x) = ||x|| \to$ conjugate of a norm is "the indicator function of unit dual norm ball"

$$f^*(z) := sup_{x \in \mathbb{R}^m}(z^T x - ||x||) = sup(0, sup_{x \in \mathbb{R}^n - \{0\}}(z^T x - ||x||)$$

$$= \text{sup}(0, \text{sup}_{L>0} L(sup_{||x||=1}(z^T x - 1)) = sup(0, sup_{L>0} L(||z||^* - 1)) = \begin{cases} 0 & ||z||_* \leq 1 \\ \infty & ||z||_* > 1 \end{cases}$$

4. $I^*_{\mathbb{R}_-}(z) = I_{\mathbb{R}_+}(z)$. Also, $I_{\mathbb{R}_-}(x) = sup_{z \geq 0} zx$: either by direct calculation or applying dual of dual
5. $I^*_{\{0\}}(z) = 0$. Also, $I_{\{0\}}(x) = sup_{z \in \mathbb{R}} zx$: either by direct calculation or applying dual of dual
6. $I^*_B(z) = sup_{x \in \mathbb{R}^n}(z^T x - I_B(x)) = sup_{x \in B} z^T x = I_{B^\perp}$. Why? consider 1) z is orthogonal to x 2) otherwise.

✓ Used for transforming **primal** problem to its **dual**. Very useful in deriving the **Lagrange Dual Function**.

The Lagrangian

✓ Find a **Convex optimization** primal probelm to another **dual** problem, wanting that the **Dual** is easier to solve! Actually, can dualize non-convex problems to make a convex dual, but not dealt here.

✓ Formulate a function called **The Lagrangian** that integrates all the **constraints** into an **unconstrained problem**.

$L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ having $dom(L) = D \times \mathbb{R}^m \times \mathbb{R}^p$, $L(x, \lambda, \nu) := f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_j h_j(x)$
1) L **is convex in** x**.**
2) L **is affine in** λ, ν**.**

Key Idea behind the Lagrangian : Pay infinite price for disobeying the constrints.

$p* = inf_x[f_0(x) + \sum_{i=1}^m I_{\mathbb{R}_-}(f_i(x)) + \sum_{j=1}^p I_{\{0\}}(h_j(x))]$:**Pay infinite price for disobeying the constrints**

Then, use indicator functions : $I_{\mathbb{R}_-}(f_i(x)) = sup_{\lambda_i \geq 0}\{\lambda_i f_i(x)\}$ and $I_{\{0\}}(h_j(x)) = sup_{\nu_j \in \mathbb{R}}\{\nu_j h_j(x)\}$

$\rightarrow p* = inf_x[f_0(x) + \sum_{i=1}^m sup_{\lambda_i \geq 0}\{\lambda_i f_i(x)\} + \sum_{j=1}^p sup_{\nu_j \in \mathbb{R}}\{\nu_j h_j(x)\}$

$\rightarrow p* = inf_x sup_{\lambda \geq 0, \nu}[f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{j=1}^p \nu_j h_j(x)]$

$\rightarrow p* = inf_x sup_{\lambda \geq 0, \nu} L(x, \lambda, \nu)$

Lagrangian Duality

$p* = min_x max_{\lambda \geq 0, \nu}[f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x)]$

The Lagrange Dual Function

$g : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}, g(\lambda, \nu) := inf_{x \in D} L(x, \lambda, \nu)$

g is a **Concave Extended Real valued Function** possibly taking $-\infty$ as a function value or a function that is ∞ everywhere.

Why? Note that L is **affine : concave and convex** in λ, ν and **infimum over concave functions is concave**.

So what? In most cases, g has a global max! (examples we deal with has).

Dual Optimization Problem

If g is not an everywhere ∞ function, the **Dual Problem**

$d^* := sup_{\lambda \succeq 0, \nu} g(\lambda, \nu) = max_{\lambda \succeq 0, \nu} min_x L(x, \lambda, \nu)$
$\leftrightarrow$ a convex problem : $-d^* := inf_{\lambda \succeq 0, \nu}\{-g(\lambda, \nu)\}$ a.k.a $inf_{\lambda, \nu}\{-g(\lambda, \nu)\}$ s.t. $\lambda \succeq 0$

✓ Always, $d^* \leq p^*$: **weak duality**

✓ Under "good" conditions, $d^* = p^*$: **strong duality**

If $d^* = p^*$: **strong duality** and let x^* be a **primal optimal point** and let (λ^*, ν^*) be a **dual optimal point**.

Then, $f_0(x^*) = g(\lambda^*, \nu^*) = inf_{x \in D} L(x, \lambda^*, \nu^*)$

Example of Lagrangian, Lagrangian Dual, Duality

$p^* = min_x[(x-1)^2+2]$ s.t. $x+2 \leq 0$ (In this case, solving the primal would be better though.)
Surely, $(x^*, p^*) = (-2, 11)$

Indicator function $I_{\mathbb{R}_-}(x+2) = sup_{\lambda \geq 0}\lambda(x+2)$.
$\rightarrow L : dom(f) = \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R} : L(x,\lambda) = (x-1)^2 + 2 + \lambda(x+2)$.
$p^* = min_x max_{\lambda \geq 0}[(x-1)^2 + 2 + \lambda(x+2)]$
$d^* = max_{\lambda \geq 0} min_x[(x-1)^2 + 2 + \lambda(x+2)]$
$g(\lambda) := min_x L(x,\lambda) = -\frac{\lambda^2}{4} + 3\lambda + 2$ and $d^* = sup_{\lambda \geq 0}(-\frac{\lambda^2}{4} + 3\lambda + 2)$.

$\therefore (\lambda^*, d^*) = (6, 11)$. **From Calculation**, you know **that Strong Duality Holds**.

✓ But how to easily check **if Strong Duality holds** without direct calculation like this : **Slater's Condition**.
✓ How to easily find the solution using the **KKT conditions**.

Weak Duality : $p^* \geq d^*$

Minimax Inequality (More general proof than just Weak Duality)

Claim) $\forall$ Sets A, B, $\forall$ function $f : A \times B \to \mathbb{R}$,

$min_{a \in A} max_{b \in B} f(a, b) \geq max_{b \in B} min_{a \in A} f(a, b)$

Proof of Minimax Inequality : from Wikipedia, Boyd and Vandenburghe (2004)

Define $g(z) \triangleq \inf_{w \in W} f(z, w)$.

$$\forall w, \forall z, g(z) \leq f(z, w)$$

$$\implies \forall w, \sup_z g(z) \leq \sup_z f(z, w)$$

$$\implies \sup_z g(z) \leq \inf_w \sup_z f(z, w)$$

$$\implies \sup_z \inf_w f(z, w) \leq \inf_w \sup_z f(z, w)$$

Example of Minimax Inequality : Game Theory

✓ A **Zero - Sum Game** setting where I pick a row, opponent picks a column. Assume rational agents.
✓ Entry is the payoff that I get from the opponent. Good if I go 1st? or 2nd?

7	-8	-7	-8	3	5
9	-5	10	-2	-10	5
-8	1	10	9	7	-2
9	10	0	6	9	3
3	10	6	10	4	-7

- Important assumption : rational agents!
- If I go first :
  1. I know that the opponent will choose the min element of the selected row.
  2. Thus, I gauge several rowmin values = -8,-10,-8,0,-7.
  3. Thus, I have to choose the 4th row : this is max-min : maximize the minimum.
- If I go second, the opponent knows that I will choose the max element of the selected column ... : min-max

✓ Let i be the row index and j be the column index of the **reward matrix**.
"Going 1st" $d^* = max_i min_j M_{ij} = 0$ vs "Going 2nd" $p^* = min_j max_i M_{ij} = 5$

1. Basic Version of the Slater's Condition

✓ A **Convex Problem** is given.
✓ If $\exists x \in relint(D)$ s.t. $f_i(x) < 0$, for $i = 1, 2, ..., m$ and $h_j(x) = 0$, for $j = 1, 2, ..., p$,

then $d^* = p^*$.

Moreover, if $p^* > -\infty$, $\exists$ dual optimal point (λ^*, ν^*).

2. Stronger Version of the Slater's Condition

✓ A **Convex Problem** is given.
✓ If $\exists x \in relint(D)$ s.t. $f_i(x) \leq 0$, for f_i : affine, $f_i(x) < 0$, for f_i not affine, and $h_j(x) = 0$, for $j = 1, 2, ..., p$,

then $d^* = p^*$.

Moreover, if $p^* > -\infty$, $\exists$ dual optimal point (λ^*, ν^*).

Relative Interior

✓$relint(S) := \{x \in S | \exists \epsilon > 0 : B_\epsilon(x) \cap aff(S)\}$: Interior of S as a subset of its affine hull.
OK to just regard this just as the interior of the domain at this level.
✓ What is the $relint(\{(1.5, 2), (3, 1)\})$?
Proof of the Slater's Condition using the supporting hyperplane theorem dealt next week.

Optimization Problem Solving Skills : Slack Variables, Relaxation, Conjugacy and Dual Norms

Example 1 : Slack Variables

$min_x(max_{i=1,2,\dots,n}x_i - min_{j=1,2,\dots,n}x_j)$ s.t. $Ax = b$ for $A \in \mathbb{R}^{m,n}$ and $b \in \mathbb{R}^m$.

Introduce slack variables t and u which represent the maximum and minmum respectively.

$min_{x,t,u}(t-u)$ s.t. $Ax = b$, $t \geq x_i, i = 1,2,...,n$, $u \leq x_i, i = 1,2,...,n$.

Example 2 : Slack Variables, Relaxation and Dual Norms

$p^* = min_{x\in\mathbb{R}^n}||Ax - y||_1 + \mu||x||_2$, $A \in \mathbb{R}^{m,n}$ and $y \in \mathbb{R}^m, \mu > 0$.

Use a slack variable $z \in \mathbb{R}^m$ that is elementwise bigger than or equal to elements of the absolute value of $Ax - y$.

Use a slack variable $t \in \mathbb{R}$ that is bigger than or equal to $||x||_2$.

Relaxation in the **Feasible Region**. Why??

$min_{x,z,t}z^T1 + \mu t$ s.t. $|(Ax)_i - y_i| \leq z_i$, $i = 1,2,...,m$ and $||x||_2 \leq t$.

Hint : use the dual norm : $||z||_2 = max_{u:||u||_2\leq 1}u^Tz$ and $||z||_1 = max_{u:||u||_\infty\leq 1}u^Tz$

$||Ax - y||_1 + \mu||x||_2 = max_{u:||u||_\infty\leq 1}\{u^T(Ax - y) + \mu \cdot max_{v:||v||_2\leq 1}v^Tx\}$.

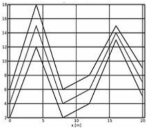
$p^* = min_x max_{u,v:||u||_\infty\leq 1,||v||_2\leq 1}\{u^T(Ax - y) + \mu v^Tx\}$.

$d^* = max_{u,v:||u||_\infty\leq 1,||v||_2\leq 1} min_x\{u^T(Ax - y) + \mu v^Tx\}$

Example 3 : Slack Variables and Relaxation

Find the path that minimizes the total length of the path (Form this problem as an SOCP).

Exercise 9.2 (A slalom problem) A two-dimensional skier must slalom down a slope, by going through n parallel gates of known position (x_i, y_i), and of width c_i, $i = 1, \dots, n$. The initial position (x_0, y_0) is given, as well as the final one, (x_{n+1}, y_{n+1}). Here, the x-axis represents the direction down the slope, from left to right, see Figure 9.24.



Problem from Optimization Models (Calafiore El Ghaoui).

Example 4 : Slack Variables and Relaxation

Warehouse Location Problem : minimize the maximum distance (Form this problem as an SOCP).

$min_x max||x - y_i||_2, i = 1, 2, ..., m$. Here, y_i are vectors of the cities, which are fixed.

Using slack variable relaxation, $min_{x,t} t$ s.t. $||x - y_i||_2 \leq t, i = 1, 2, ..., m$.