Constrained Minimization

- Algos & • Theory

The general inequality constrained convex minimization problem is

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & g_i(\mathbf{x}) \le \mathbf{0}, \quad i = 1, \ldots, m \\ & A\mathbf{x} = b \end{array} \tag{4.105}$$

where f as well as the g_i's are convex and twice continuously differentiable.

Constraints above give a convex set

[If $h_j(x) = 0$ then $h_j(x) \le 0$ & $-h_j(x) \le 0 \Rightarrow$ If both convex then h_j shld be affine]

$L_{\alpha_i}(g_i) : \{x \mid g_i(x) \le \alpha_i\}$ ... Sublevel set

If fn $g_i(x)$ is convex $L_{\alpha_i}(g_i)$ will be convex & so will be $\bigcap_i L_{\alpha_i}(g_i) \cap \{x \mid Ax = b\}$

Introducing the auxiliary variable $t \in \mathbb{R}$, we can rewrite (4.105) equivalently as

$\min$ t → A linear objective

subject to

Convex?

HOMEWORK: IDENTIFY NON-AFFINE $h_j(x) = 0$ that yields a convex domain.

In general:

$$\boxed{\begin{array}{lll} \min\limits_{x} & f(x) & \\ \text{s.t} & g_i(x) \leq 0 & i = 1 \ldots n \\ & h_j(x) = 0 & j = 1 \ldots m \end{array}}$$

For a while no convexity assumptions on g_i & f will be considered.

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ & x \in S_x \end{array}$$

(S_x is convex set)

$$\left.\begin{array}{ll} \min\limits_{z} & z \\ & (z, x) \in S_{(z,x)} \end{array}\right.$$

(S_z is convex set)

$z \in \mathbb{R}$ & $S_{(z,x)} = \{\mathbb{R} \times S_x\} \cap \{(z, x) \mid f(x) \leq z\}$

More generally, a convex program can be written as minimization of a linear function $c^T x$ ($x \in \mathbb{R}$ and $c=1$ above) over a convex feasible region F_c

$$\min \; c^T x$$
$$\text{subject to } \; x \in F_c$$

Recall definition of a conic program

$$\min \; c^T x$$
$$\text{s.t } \; Ax - b \in K$$

where K is a proper cone.

Claim:

Any convex program can be written as a conic program

Proof:

Given a convex optimisation problem

$$\min_{x \in \mathbb{R}^n} \quad c^T x$$

$$\text{subject to} \quad x \in F_c$$

Embed $\mathbb{R}^n$ into $\mathbb{R}^{n+1}$ as the hyperplane $H = \{1\} \times \mathbb{R}^n \subset \mathbb{R}^{n+1}$

and define a proper cone

$$K = cl\left(\left\{(t,x) \in \mathbb{R}^{n+1} : \frac{x}{t} \in F_c\right\}_{t>0}\right)$$

Let $d = \begin{pmatrix} 0 \\ c \end{pmatrix}$. We can write the above convex program as the following conic program:

$$\min_{(t,x) \in \mathbb{R}^{n+1}} \quad d^T x$$

$$\text{subject to} \quad x \in H \cap K$$

Needed a conic opt problem

Note that $x \in F_c$ iff $(1,x) \in K$

i.e. $x \in K \cap H$

Proof that K is a cone:

Let (t_1, x_1) & $(t_2, x_2) \in K$ and $\theta_1, \theta_2 \geq 0$

Consider $\theta_1(t_1, x_1) + \theta_2(t_2, x_2)$

$$\frac{\theta_1 x_1 + \theta_2 x_2}{\theta_1 t_1 + \theta_2 t_2} = \left(\frac{x_1}{t_1}\right)\underbrace{\left(\frac{\theta_1 t_1}{\theta_1 t_1 + \theta_2 t_2}\right)}_{\in [0,1]} + \left(\frac{x_2}{t_2}\right)\underbrace{\left(\frac{\theta_2 t_2}{\theta_1 t_1 + \theta_2 t_2}\right)}_{\in [0,1]}$$

Sum to 1

$=$ convex combination of $\left(\frac{x_1}{t_1}\right)$ and $\left(\frac{x_2}{t_2}\right)$

& therefore $\in C$

$$\Rightarrow \theta_1(t_1, x_1) + \theta_2(t_2, x_2) \in K.$$

Let us recall our discussion on linear programs (LP), dual of LP, conic programs & their duals

[Ref page 5 of http://www2.isye.gatech.edu/~nemirovs/ICMNemirovski.pdf]

LP

Affine objective

$$\begin{array}{ll} \min\limits_{\mathbf{x}\in\Re^n} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & -A\mathbf{x}+\mathbf{b}\le 0 \end{array}$$

Let: $\lambda \geq 0$ (i.e $\lambda \in \mathbb{R}^n_+$)

Then $\lambda^T(-Ax+b) \leq 0$

$$\Rightarrow c^Tx \geq c^Tx + \lambda^T(-Ax+b)$$

$$= \lambda^T b + (c - A^T\lambda)^T x$$

$$\geq \min_x \lambda^T b + (c - A^T\lambda)^T x$$

$$= \begin{cases} \lambda^T b & \text{if } A^T\lambda = c \\ -\infty & \text{if } A^T\lambda \neq c \end{cases}$$

independent of λ

independent of x

$$\Rightarrow \quad \boxed{\begin{array}{ll} \min\limits_{x\in\mathbb{R}^n} & c^Tx \\ \text{s.t} & Ax \geq b \end{array} \quad \geq \quad \begin{array}{ll} \max & b^T\lambda \\ & \lambda \geq 0 \\ \text{s.t} & A^T\lambda = c \end{array}}$$

Primal LP (lower bounded)

Dual LP (upper bounded)

Conic Program (CP)

$$\begin{array}{ll} \min\limits_{\mathbf{x}\in\Re^n} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & -A\mathbf{x}+\mathbf{b}\le_K 0 \end{array}$$

K is a regular/proper cone

Generalised cone program

$$\boxed{\begin{array}{ll} \min\limits_{x\in S} & \langle c, x\rangle_S \\ \text{subject to} & Ax - b \in K \end{array}}$$

We need an equivalent $\lambda \in K^*$ s.t

$$\langle \lambda, Ax-b \rangle \geq 0$$

This K^* s.t

$$K^* = \{\lambda \mid \langle \lambda, Ax-b\rangle \geq 0 \ \ \forall\, Ax-b \in K\}$$

is called the DUAL CONE of K (ie element that have +ve inner prod with each element of K) by dual) | by primal)

Called the weak duality theorem for Linear Program

$K_* = \{\lambda : \lambda^T \xi \geq 0 \, \forall \xi \in K\}$ is the cone dual to K

{defn on page 7 of http://www2.isye.gatech.edu/~nemirovs/ICMNemirovski.pdf}

With this, prove the following weak duality theorem for CONIC PROGRAM

$$\min_{x \in S} \langle c, x \rangle \quad \text{s.t. } Ax \geq_K b \quad \geq \quad \max_{\lambda \in K^*} \langle b, \lambda \rangle \quad \text{s.t. } A^T\lambda = c$$

Primal CP (lower bounded by dual)

Dual CP (upperbounded by primal)

Note: ① Both LP & CP dealt with affine objective

② CP dealt with the generalised conic inequalities

③ Later, in convex programs, we will deal with the more general convex functions in the objective

Notes:

① If $K = \mathbb{R}^n_+$, the CP is an LP

If $K = S^n_+$, the CP is an SDP

Set of all $n \times n$ symmetric positive semidefinite matrices

Semi-definite program

② Any generic convex program can be expressed as a cone program (CP) [H/W]

HOW ABOUT STRONG DUALITY FOR LPs?

[Page 21 of http://www2.isye.gatech.edu/~nemirovs/Lect_ModConvOpt.pdf]

Theorem 1.2.2 [Duality Theorem in Linear Programming] *Consider a linear programming program*

$$\min_x \left\{ c^T x \,\middle|\, Ax \geq b \right\} \tag{LP}$$

along with its dual

$$\max_y \left\{ b^T y \,\middle|\, A^T y = c, y \geq 0 \right\} \tag{LP*}$$

Then

1) *The duality is symmetric: the problem dual to dual is equivalent to the primal;*

2) *The value of the dual objective at every dual feasible solution is $\leq$ the value of the primal objective at every primal feasible solution*

Weak LP duality (already proved)

3) *The following 5 properties are equivalent to each other:*

$\exists x$ s.t. $Ax \geq b$

$-\infty < c^T x$, $Ax \geq b$

(i) *The primal is feasible and bounded below.*

$b^T y < \infty$, $A^T y = c$

(ii) *The dual is feasible and bounded above.*

$\exists \lambda$ s.t. $A^T \lambda = c$

(iii) *The primal is solvable.*

LP has a soln

(iv) *The dual is solvable.*

LP* has a soln

(v) *Both primal and dual are feasible.*

Whenever (i) $\equiv$ (ii) $\equiv$ (iii) $\equiv$ (iv) $\equiv$ (v) is the case, the optimal values of the primal and the dual problems are equal to each other. : Strong duality = (2) + (3)

Proof of 1 from page 21 of http://www2.isye.gatech.edu/~nemirovs/Lect_ModConvOpt.pdf

Proof. 1) is quite straightforward: writing the dual problem (LP*) in our standard form, we get

$$\min_y \left\{ -b^T y \,\middle|\, \begin{bmatrix} I_m \\ A^T \\ -A^T \end{bmatrix} y - \begin{pmatrix} 0 \\ -c \\ c \end{pmatrix} \geq 0 \right\},$$

where I_m is the m-dimensional unit matrix. Applying the duality transformation to the latter problem, we come to the problem

$$\max_{\xi,\eta,\zeta} \left\{ 0^T \xi + c^T \eta + (-c)^T \zeta \,\middle|\, \begin{array}{rcl} \xi & \geq & 0 \\ \eta & \geq & 0 \\ \zeta & \geq & 0 \\ \xi - A\eta + A\zeta & = & -b \end{array} \right\},$$

which is clearly equivalent to (LP) (set $x = \eta - \zeta$).

Similar Duality theorem for CP:

[page 7 of http://www2.isye.gatech.edu/~nemirovs/ICMNemirovski.pdf]

Note dual(dual(K)) = K if K is a closed

dual(K) is ALWAYS a closed cone

$$\min_x \left\{ c^T x : \underbrace{Ax - b \geq_K 0}_{\Leftrightarrow Ax - b \in K} \right\} \qquad \text{(CP)}$$

$$\max_\lambda \left\{ b^T \lambda : A^T \lambda = c, \lambda \geq_{K_*} 0 \right\}, \qquad \text{(D)}$$

Theorem 2.1. *Assuming A in* (CP) *is of full column rank, the following is true:*

(i) *The duality is symmetric:* (D) *is a conic problem, and the conic dual to* (D) *is (equivalent to)* (CP)*;*

(ii) [weak duality] Opt(D) $\leq$ Opt(CP)*;* → Already proved

(iii) [strong duality] *If one of the programs* (CP), (D) *is bounded and strictly feasible* (i.e., the corresponding affine plane intersects the interior of the associated cone), *then the other is solvable and* Opt(CP) = Opt(D). *If both* (CP), (D) *are strictly feasible, then both programs are solvable and* Opt(CP) = Opt(D)*;*

(iv) [optimality conditions] *Assume that both* (CP), (D) *are* $\boxed{\textit{strictly feasible.}}$ *Then a pair (x, λ) of feasible solutions to the problem is comprised of optimal solutions iff $c^T x = b^T \lambda$ ("zero duality gap"), same as iff $\lambda^T[Ax - b] = 0$ ("complementary slackness").*

H/W: Prove these in a manner similar to the duality theorem for LP

Note: Duality theorems for LP and CP are special cases of Lagrange duality that we will discuss later in the course

From dual of LP to dual of a general Optimisation problem

LP   Affine objective

$$\begin{array}{ll} \min\limits_{x \in \Re^n} & c^T x \\ \text{subject to} & -Ax + b \le 0 \end{array}$$

Let $\lambda \geq 0$ (i.e $\lambda \in \mathbb{R}^n_+$)

Then $\lambda^T(-Ax+b) \le 0$

$$\Rightarrow c^T x \geq c^T x + \lambda^T(-Ax+b)$$
$$= \lambda^T b + (c - A^T\lambda)^T x$$
$$\geq \min_x \lambda^T b + (c - A^T\lambda)^T x$$
$$= \begin{cases} \lambda^T b & \text{if } A^T\lambda = c \\ -\infty & \text{if } A^T\lambda \neq c \end{cases}$$

independent of λ independent of x

$\Rightarrow$

$$\boxed{\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} c^T x & \geq \max\limits_{\lambda \geq 0} b^T\lambda \\ \text{s.t } Ax \geq b & \text{s.t } A^T\lambda = c \end{array}}$$

Primal LP (lower bounded)   Dual LP (upper bounded)

$$\min_{x \in \mathbb{R}^n} f_0(x)$$
subject to $f_i(x) \le 0 \quad i = 1..m$
$$h_j(x) = 0 \quad j = 1..p$$

Let $\lambda \geq 0$ (ie $\lambda \in \mathbb{R}^p_+$) and x be feasible. Then

$$\Rightarrow f_0(x) \geq f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{j=1}^p \mu_j h_j(x)$$
$$\geq \underbrace{\min_x \underbrace{\left(f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{j=1}^p \mu_j h_j(x)\right)}_{L(x,\lambda,\mu)}}_{L^*(\lambda,\mu)}$$

over all x
Not just feasible x

$$\begin{array}{l} \min\limits_{x \in \mathbb{R}^n} f(x) \geq \max\limits_{\lambda,\mu} L^*(\lambda,\mu) \\ \text{s.t } f_i(x) \le 0 \\ \quad h_j(x) = 0 \end{array}$$

5. Duality

- Lagrange dual problem
- weak and strong duality
- geometric interpretation
- optimality conditions
- perturbation and sensitivity analysis
- examples
- generalized inequalities

Lagrangian

standard form problem (not necessarily convex)

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

variable $x \in \mathbf{R}^n$, domain $\mathcal{D}$, optimal value $p^\star$

Lagrangian: $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$, with $\mathbf{dom}\, L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$,

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

- weighted sum of objective and constraint functions
- λ_i is Lagrange multiplier associated with $f_i(x) \leq 0$
- ν_i is Lagrange multiplier associated with $h_i(x) = 0$

Lagrange dual function

$L(x,\lambda,\mu) \le f_0(x)$

$L^*(\lambda,\mu) \le f_0(x)$

$L^*(\lambda,\mu) \le p^*$

Lagrange dual function: $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$,

$L^*(\lambda, \nu)$ or
$$
\begin{aligned}
g(\lambda,\nu) &= \inf_{x\in\mathcal{D}} L(x,\lambda,\nu) \\
&= \inf_{x\in\mathcal{D}} \left(f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right)
\end{aligned}
$$

g is concave, can be $-\infty$ for some λ, ν

lower bound property: if $\lambda \succeq 0$, then $g(\lambda,\nu) \le p^\star$

proof: if $\tilde{x}$ is feasible and $\lambda \succeq 0$, then

$$f_0(\tilde{x}) \ge L(\tilde{x},\lambda,\nu) \ge \inf_{x\in\mathcal{D}} L(x,\lambda,\nu) = g(\lambda,\nu)$$

minimizing over all feasible $\tilde{x}$ gives $p^\star \ge g(\lambda,\nu)$

Weak duality

(Refer derivation on page 9)

Prove: $L^*(\lambda,\nu)$ or $g(\lambda,\nu)$ is concave

$$
\begin{aligned}
& g(\theta\lambda_1 + (1-\theta)\lambda_2, \theta\nu_1 + (1-\theta)\nu_2) \\
&= \inf_{x\in D} \left[f_0(x)\cdot(\theta + 1 - \theta) + \sum_{i=1}^m (\theta\lambda_{1_i} + (1-\theta)\lambda_{2_i}) f_i(x) + \sum_{i=1}^p (\theta\nu_{1_i} + (1-\theta)\nu_{2_i}) h_i(x) \right]
\end{aligned}
$$

$$
= \inf_{x\in D} \left[\theta\left[f_0(x) + \sum_{i=1}^m \lambda_{1_i} f_i(x) + \sum_{i=1}^p \nu_{1_i} h_i(x)\right] + (1-\theta)\left[f_0(x) + \sum_{i=1}^m \lambda_{2_i} f_i(x) + \sum_{i=1}^p \nu_{2_i} h_i(x)\right] \right]
$$

$$\geq \theta \left\{ \inf_{x \in D} \left(f_0(x) + \sum_{i=1}^{m} \lambda_{1i} f_i(x) + \sum_{i=1}^{p} \nu_{1i} h_i(x) \right) \right\}$$

$$+ (1-\theta) \left\{ \inf_{x \in D} \left(f_0(x) + \sum_{i=1}^{m} \lambda_{2i} f_i(x) + \sum_{i=1}^{p} \nu_{2i} h_i(x) \right) \right\}$$

$$= \theta g(\lambda_1, \mu_1) + (1-\theta) g(\lambda_2, \mu_2) \Rightarrow g(\lambda, \mu) \equiv L^*(\lambda, \mu)$$

is CONCAVE

Least-norm solution of linear equations

$$\begin{array}{ll} \text{minimize} & x^T x \\ \text{subject to} & Ax = b \end{array}$$

substitute for x

dual function Lagrange fn

- Lagrangian is $L(x, \nu) = x^T x + \nu^T (Ax - b)$
- to minimize L over x, set gradient equal to zero:

$$\nabla_x L(x, \nu) = 2x + A^T \nu = 0 \quad \Longrightarrow \quad \underbrace{x = -(1/2) A^T}_{}\nu$$

Hint: in general: Find $\Rightarrow$ primal vars $= f$ (dual vars)

pt of min on $\nabla_x^2 L(x, \nu) = 2I \succ 0$

- plug in in L to obtain g:

$$g(\nu) = L((-1/2) A^T \nu, \nu) = -\frac{1}{4} \nu^T A A^T \nu - b^T \nu$$

Lagrange dual

a concave function of ν

lower bound property: $p^\star \geq -(1/4) \nu^T A A^T \nu - b^T \nu$ for all ν

in fact $p^* \geq \max_{\nu} g(\nu) = b^T (A A^T)^{-1} b$

[can show that lower bnd can be attained]

Soln to H/w:

- 06/11/2013. For the problem of least norm solution of linear equations (page no 13), show that A is an m x n matrix with m < n and if A has full row rank, strong duality holds, that is, there exists a point x satisfying the primal constraints such that the lower bound obtained using weak duality is actually attained. Hint: Refer to this. **Deadline:** 8th November.

Ans: • One way is to express soln to $Ax=b$ as sum of a particular solution to $Ax=b$ ($x_{particular}$) and a vector from null space of A ($x_{nullspace}$) [see pg 172 of
http://www.cse.iitb.ac.in/~CS709/notes/LinearAlgebra.pdf]

$$\mathbf{x}_{complete} = \mathbf{x}_{particular} + \mathbf{x}_{nullspace} \qquad (3.35)$$

• Another way is to find some value for x s.t. $Ax=b$ and $x^Tx = b^T(AA^T)^{-1}b$

since this will mean a feasible point exists at which primal objective = lower bound on the minimisation problem

which will mean that primal optimal solution = dual optimal solution

Verify that for $x = A^T(AA^T)^{-1}b$
$Ax=b$ & $x^Tx = b^T(AA^T)^{-1}b$

Standard form LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, \quad x \succeq 0 \end{array}$$

dual function

- Lagrangian is

$$\begin{array}{rcl} L(x, \lambda, \nu) & = & c^T x + \nu^T (Ax - b) - \lambda^T x \\ & = & -b^T \nu + (c + A^T \nu - \lambda)^T x \end{array}$$

- L is linear in x, hence

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = \begin{cases} -b^T \nu & A^T \nu - \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases}$$

  g is linear on affine domain $\{(\lambda, \nu) \mid A^T \nu - \lambda + c = 0\}$, hence concave

lower bound property: $p^\star \geq -b^T \nu$ if $A^T \nu + c \succeq 0$

Equality constrained norm minimization

$$\begin{array}{ll} \text{minimize} & \|x\| \\ \text{subject to} & Ax = b \end{array}$$

dual function

$$g(\nu) = \inf_x (\|x\| - \nu^T A x + b^T \nu) = \begin{cases} b^T \nu & \|A^T \nu\|_* \leq 1 \\ -\infty & \text{otherwise} \end{cases}$$

where $\|v\|_* = \sup_{\|u\| \leq 1} u^T v$ is dual norm of $\| \cdot \|$

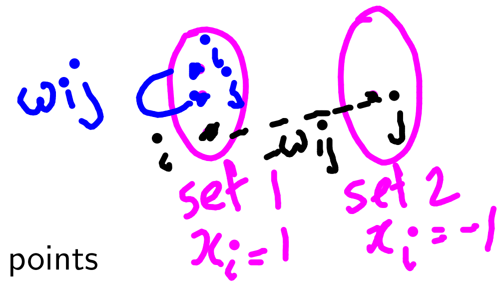
proof: follows from $\inf_x (\|x\| - y^T x) = 0$ if $\|y\|_* \leq 1$, $-\infty$ otherwise

- if $\|y\|_* \leq 1$, then $\|x\| - y^T x \geq 0$ for all x, with equality if $x = 0$
- if $\|y\|_* > 1$, choose $x = tu$ where $\|u\| \leq 1$, $u^T y = \|y\|_* > 1$:

$$\|x\| - y^T x = t(\|u\| - \|y\|_*) \to -\infty \quad \text{as } t \to \infty$$

lower bound property: $p^\star \geq b^T \nu$ if $\|A^T \nu\|_* \leq 1$

Two-way partitioning



$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \dots, n \end{array}$$

- a nonconvex problem; feasible set contains 2^n discrete points
- interpretation: partition $\{1, \dots, n\}$ in two sets; W_{ij} is cost of assigning i, j to the same set; $-W_{ij}$ is cost of assigning to different sets

dual function

$$\begin{aligned} g(\nu) = \inf_x (x^T W x + \sum_i \nu_i (x_i^2 - 1)) &= \inf_x x^T (W + \mathbf{diag}(\nu)) x - \mathbf{1}^T \nu \\ &= \begin{cases} -\mathbf{1}^T \nu & W + \mathbf{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

lower bound property: $p^\star \geq -\mathbf{1}^T \nu$ if $W + \mathbf{diag}(\nu) \succeq 0$

example: $\nu = -\lambda_{\min}(W)\mathbf{1}$ gives bound $p^\star \geq n\lambda_{\min}(W)$

Study the connection between w & λmin for diff choices of w

Lagrange dual and conjugate function

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Ax \preceq b, \quad Cx = d \end{array}$$

dual function

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in \mathbf{dom}\, f_0} \left(f_0(x) + (A^T \lambda + C^T \nu)^T x - b^T \lambda - d^T \nu \right) \\ &= -f_0^*(-A^T \lambda - C^T \nu) - b^T \lambda - d^T \nu \end{aligned}$$

- recall definition of conjugate $f^*(y) = \sup_{x \in \mathbf{dom}\, f} (y^T x - f(x))$
- simplifies derivation of dual if conjugate of f_0 is kown

example: entropy maximization

$$f_0(x) = \sum_{i=1}^n x_i \log x_i, \qquad f_0^*(y) = \sum_{i=1}^n e^{y_i - 1}$$

The dual problem

Lagrange dual problem

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

- finds best lower bound on $p^\star$, obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted $d^\star$
- λ, ν are dual feasible if $\lambda \succeq 0$, $(\lambda, \nu) \in \mathbf{dom}\, g$
- often simplified by making implicit constraint $(\lambda, \nu) \in \mathbf{dom}\, g$ explicit

example: standard form LP and its dual (page 5–5)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0 \end{array} \qquad\qquad \begin{array}{ll} \text{maximize} & -b^T \nu \\ \text{subject to} & A^T \nu + c \succeq 0 \end{array}$$

Weak and strong duality

weak duality: $d^\star \le p^\star$

- always holds (for convex and nonconvex problems)
- can be used to find nontrivial lower bounds for difficult problems

  for example, solving the SDP

$$\begin{array}{ll} \text{maximize} & -\mathbf{1}^T \nu \\ \text{subject to} & W + \mathbf{diag}(\nu) \succeq 0 \end{array}$$

  gives a lower bound for the two-way partitioning problem on page 5–7

strong duality: $d^\star = p^\star$

- does not hold in general
- (usually) holds for convex problems
- conditions that guarantee strong duality in convex problems are called **constraint qualifications**

For LP: Feasibility of primal & dual

For Conic Prog: Strict feasibility ie $\exists\ x \in \text{int}(K)$

Slater's constraint qualification

strong duality holds for a convex problem

$I \subseteq D$

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& Ax = b
\end{array}
$$

if it is strictly feasible, *i.e.*,

$$
\exists x \in \textbf{int}\, \mathcal{D}: \qquad f_i(x) < 0, \quad i = 1, \ldots, m, \qquad Ax = b
$$

- also guarantees that the dual optimum is attained (if $p^\star > -\infty$)
- can be sharpened: *e.g.*, can replace $\textbf{int}\, \mathcal{D}$ with $\textbf{relint}\, \mathcal{D}$ (interior relative to affine hull); linear inequalities do not need to hold with strict inequality, . . .
- there exist many other types of constraint qualifications

[Proof of strong duality under constraint qualification in section 5.3.2 pg 234 onwards of cvx book]

Inequality form LP

primal problem

$$
\begin{array}{ll}
\text{minimize} & c^T x \\
\text{subject to} & Ax \preceq b
\end{array}
$$

dual function

$$
g(\lambda) = \inf_x \left((c + A^T \lambda)^T x - b^T \lambda \right) = \begin{cases} -b^T \lambda & A^T \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases}
$$

dual problem

$$
\begin{array}{ll}
\text{maximize} & -b^T \lambda \\
\text{subject to} & A^T \lambda + c = 0, \quad \lambda \succeq 0
\end{array}
$$

- from Slater's condition: $p^\star = d^\star$ if $A\tilde{x} \prec b$ for some $\tilde{x}$
- in fact, $p^\star = d^\star$ except when primal and dual are infeasible

Quadratic program

primal problem (assume $P \in \mathbf{S}^n_{++}$)

$$\begin{array}{ll} \text{minimize} & x^T P x \\ \text{subject to} & \boxed{Ax \preceq b} \end{array}$$

convex polyhedron

dual function

$$g(\lambda) = \inf_x \left(x^T P x + \lambda^T (Ax - b)\right) = -\frac{1}{4}\lambda^T A P^{-1} A^T \lambda - b^T \lambda$$

dual problem

$$\begin{array}{ll} \text{maximize} & -(1/4)\lambda^T A P^{-1} A^T \lambda - b^T \lambda \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

- from Slater's condition: $p^\star = d^\star$ if $A\tilde{x} \prec b$ for some $\tilde{x}$
- in fact, $p^\star = d^\star$ always

A nonconvex problem with strong duality

$$\begin{array}{ll} \text{minimize} & x^T A x + 2b^T x \\ \text{subject to} & x^T x \leq 1 \end{array}$$

$A \not\succeq 0$, hence nonconvex

dual function: $g(\lambda) = \inf_x (x^T (A + \lambda I) x + 2b^T x - \lambda)$

- unbounded below if $A + \lambda I \not\succeq 0$ or if $A + \lambda I \succeq 0$ and $b \notin \mathcal{R}(A + \lambda I)$
- minimized by $x = -(A + \lambda I)^\dagger b$ otherwise: $g(\lambda) = -b^T (A + \lambda I)^\dagger b - \lambda$

dual problem and equivalent SDP:

$$\begin{array}{ll} \text{maximize} & -b^T (A + \lambda I)^\dagger b - \lambda \\ \text{subject to} & A + \lambda I \succeq 0 \\ & b \in \mathcal{R}(A + \lambda I) \end{array} \qquad \begin{array}{ll} \text{maximize} & -t - \lambda \\ \text{subject to} & \begin{bmatrix} A + \lambda I & b \\ b^T & t \end{bmatrix} \succeq 0 \end{array}$$

strong duality although primal problem is not convex (not easy to show)

Geometry of the dual [page 292 onwards, section 4.4.3 of http://www.cse.iitb.ac.in/~CS709/notes/BasicsOfConvexOptimization.pdf]

Let Primal be.

$$\begin{aligned} \min_{\mathbf{x}\in\mathcal{D}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \le 0,\ i = 1, 2, \ldots, m \end{aligned} \tag{4.80}$$

[$h_j(x) = 0$ can be expressed as $h_j(x) \le 0$ & $-h_j(x) \le 0$

The dual is

$$\begin{aligned} \max_{\lambda\in\Re^m} \quad & L^*(\lambda) \\ \text{subject to} \quad & \lambda \ge \mathbf{0} \end{aligned} \tag{4.81}$$

where: $L^*(\lambda) = \min_{\mathbf{x}\in\mathcal{D}} L(\mathbf{x}, \lambda) = \min_{\mathbf{x}\in\mathcal{D}} f(\mathbf{x}) + \lambda^T \mathbf{g}(\mathbf{x})$

Let

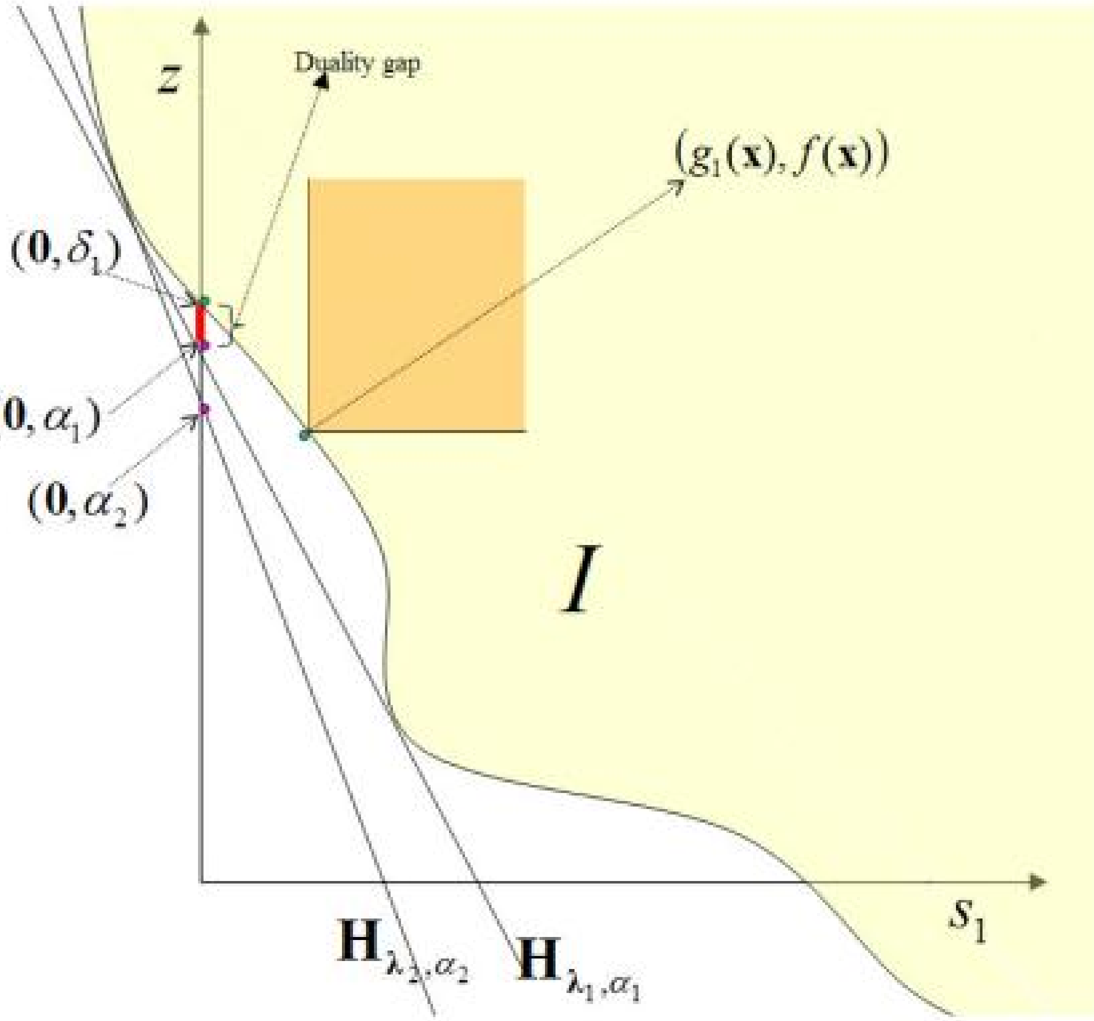
Deceptively similar to epigraph

$$\boxed{\mathcal{I} = \{(\mathbf{s}, z) \mid \mathbf{s} \in \Re^m,\ z \in \Re,\ \exists \mathbf{x} \in \mathcal{D} \textit{ with } g_i(\mathbf{x}) \le s_i\ \forall 1 \le i \le m,\ f(\mathbf{x}) \le z\}}$$

Note: $\mathcal{I} \subseteq \mathbb{R}^{m+1}$

Plot:- Example of the set $\mathcal{I}$ for a single constraint (*i.e.*, for $n = 1$).

Plot with s_1 on x-axis and z on y-axis



1. If $f(x)$ is convex and each of $g_i(x)$ are convex, then I will be convex

2. Feasible region of primal problem (4.80) corresponds to $\boxed{\text{subset of } I \text{ with } s_1 \leq 0}$

3. Solution to primal problem corresponds to point in I with $\boxed{\text{lowest value of } z \text{ such that } s_1 = 0 \ldots \text{ in the figure it is } (0, \delta_1)}$

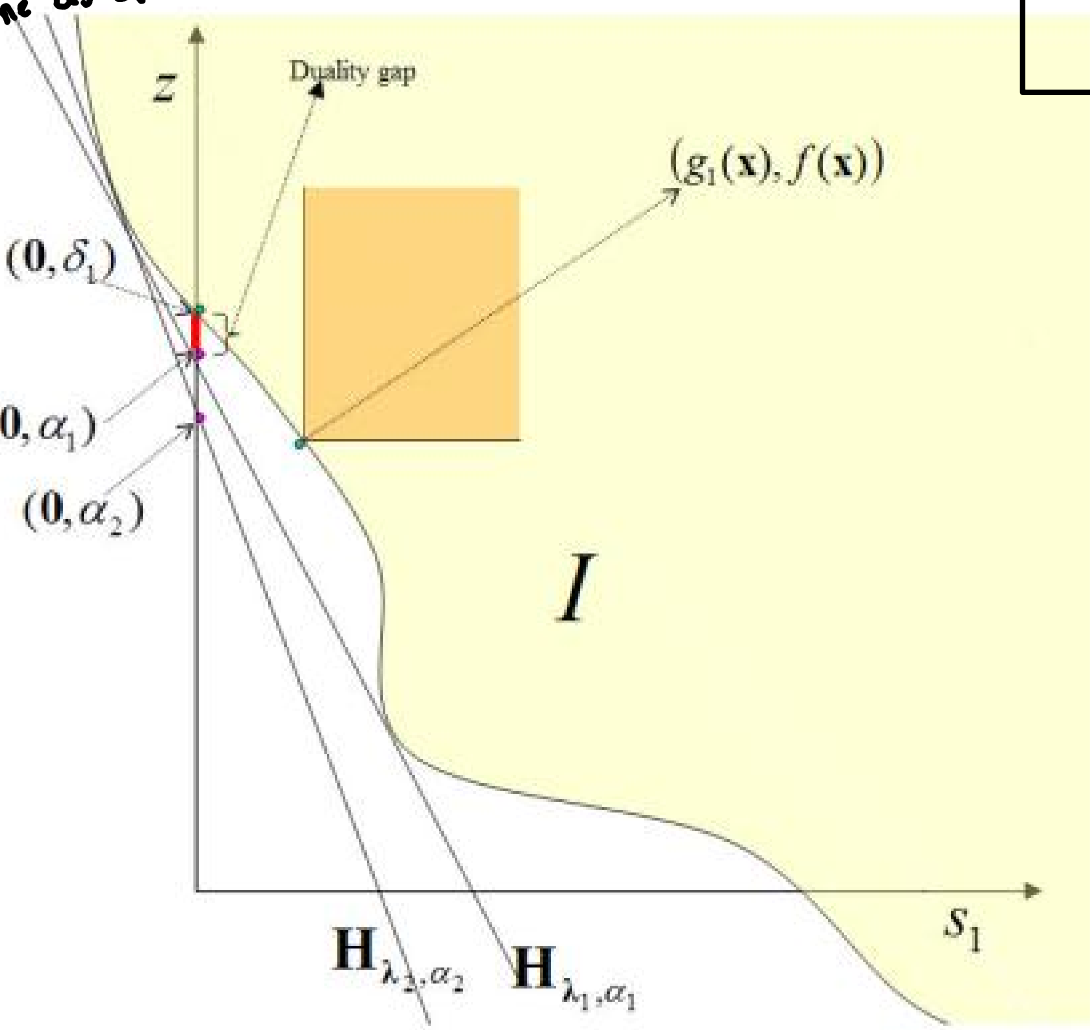
④

Let us define a hyerplane $\mathcal{H}_{\lambda,\alpha}$, parametrized by $\lambda \in \Re^m$ and $\alpha \in \Re$

$$\mathcal{H}_{\lambda,\alpha} = \{(\mathbf{s}, z) \,|\lambda^T.\mathbf{s} + z = \alpha\}$$

⑤ Of all $H_{\lambda,\alpha}$ that lie below I, consider the one which has as high a value of z-intercept as possible. This must be a $\boxed{\text{supporting hyperplane}}$

same as α



⑥ The problem in ⑤ can be specified as the following optimization problem:

$$\boxed{\begin{array}{ll} \max & \alpha \\ \text{subject to} & \mathcal{H}^+_{\lambda,\alpha} \supseteq \mathcal{I} \end{array}}$$

where $\mathcal{H}^+_{\lambda,\alpha}$ is the half space ABOVE $\mathcal{H}_{\lambda,\alpha}$

$$\mathcal{H}^+_{\lambda,\alpha} = \{(\mathbf{s}, z) \,|\, \lambda^T.\mathbf{s} + z \geq \alpha\}$$

By definitions of $\mathcal{I}$, $\mathcal{H}^+_{\lambda,\alpha}$ and the subset relation

$$\boxed{\begin{array}{ll} \max & \alpha \\ \text{subject to} & \lambda^T.\mathbf{s} + z \geq \alpha \;\forall (\mathbf{s}, z) \in \mathcal{I} \end{array}}$$

7 If $(s,z) \in \mathcal{I}$ then for all $s' \geq s$ we will have $\boxed{(s', z) \in \mathcal{I}}$

8 Therefore we cannot afford to have any component λ_i to be $\boxed{\text{negative}}$

9 Thus we can add the constraint $\boxed{\lambda \geq 0}$ to the above problem w/o changing soln.

.

$$
\begin{aligned}
\max \quad & \alpha \\
\text{subject to} \quad & \lambda^T.\mathbf{s} + z \geq \alpha \ \forall (\mathbf{s}, z) \in \mathcal{I} \\
& \lambda \geq \mathbf{0}
\end{aligned}
$$

10 Using the fact that every point on the form boundary $(\mathcal{I}) = \partial \mathcal{I}$ must be of the form

$$\boxed{(s', z') = (g_1(x')\ g_2(x') \ldots g_n(x'), f(x'))}$$

we get the following equivalent optimisation problem:

$$
\begin{aligned}
\max \quad & \alpha \\
\text{subject to} \quad & \lambda^T.\mathbf{g}(\mathbf{x}) + f(\mathbf{x}) \geq \alpha \ \forall \mathbf{x} \in \mathcal{D} \\
& \lambda \geq \mathbf{0}
\end{aligned}
$$

11 Recalling that $L(x, \lambda) = \lambda^T g(x) + f(x)$, we obtain

$$
\begin{array}{ll}
\max\limits_{(\alpha,\lambda)} & \alpha \\
\text{subject to} & L(\mathbf{x}, \lambda) \geq \alpha \ \forall \mathbf{x} \in \mathcal{D} \\
& \lambda \geq \mathbf{0}
\end{array}
$$

Since, $L^*(\lambda) = \min\limits_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \lambda)$, we can deal with (equivalently)

$$
\begin{array}{ll}
\max\limits_{(\alpha,\lambda)} & \alpha \\
\text{subject to} & L^*(\lambda) \geq \alpha \\
& \lambda \geq \mathbf{0}
\end{array}
$$

This problem can be restated as

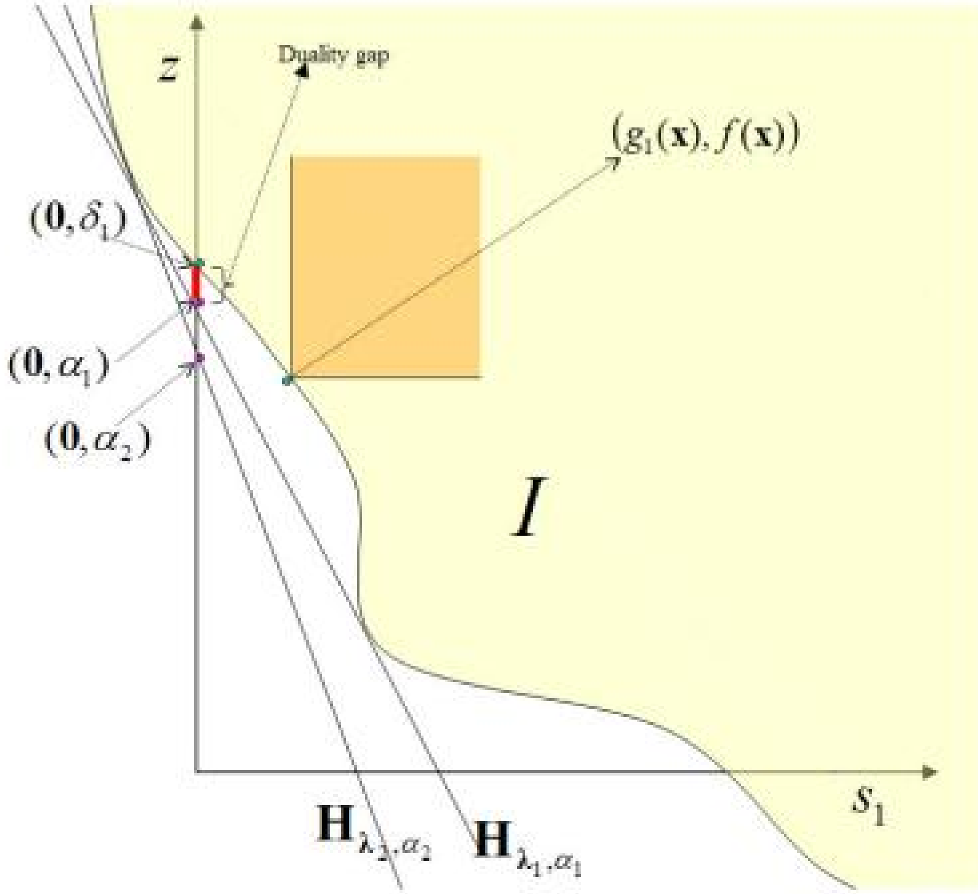
$$
\begin{array}{ll}
\max\limits_{\lambda} & L^*(\lambda) \\
\text{subject to} & \lambda \geq 0
\end{array}
$$

This is precisely the dual problem.

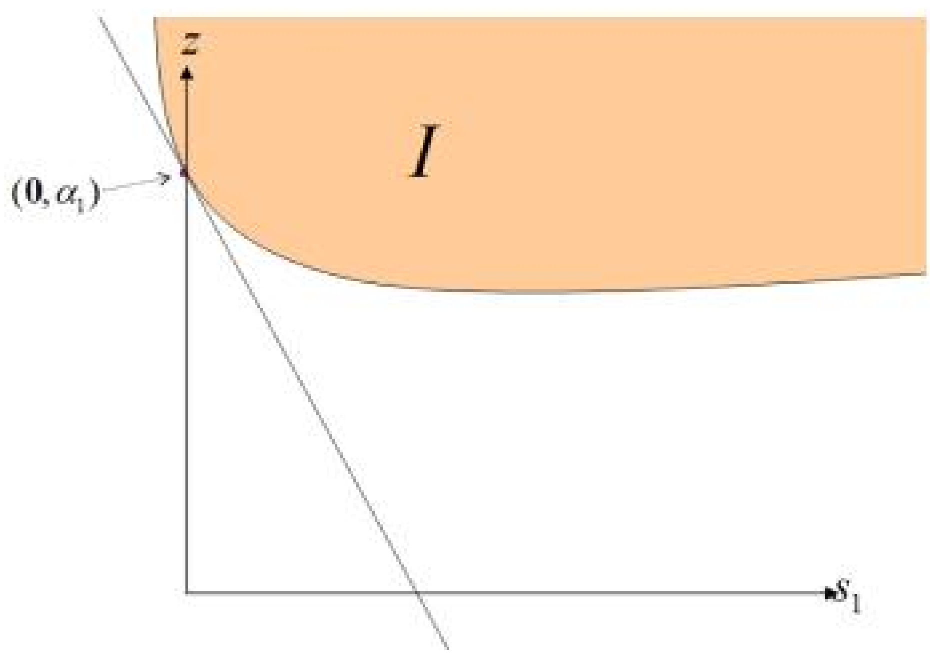
(12) What is effect of convexity of I on gap between primal & dual?

Non convex $I \Rightarrow$ There COULD BE a gap between $(0, \delta_1)$ and $(0, \alpha_1)$

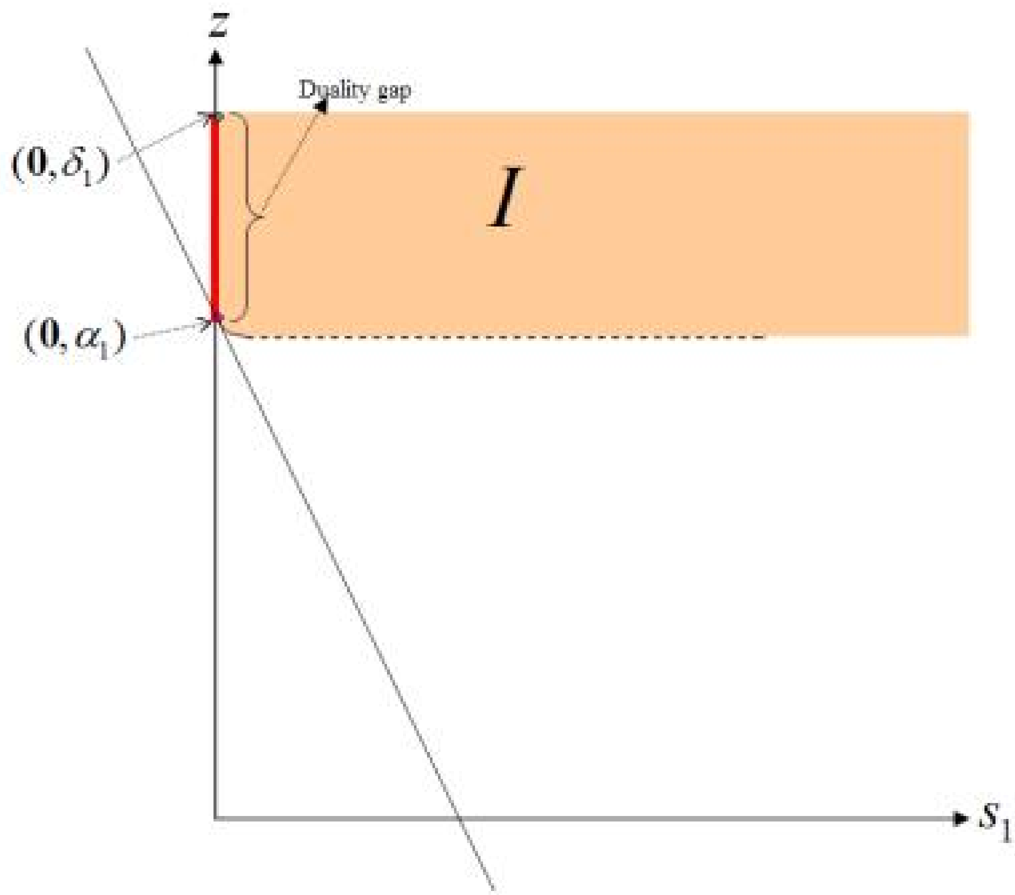
[we can NEVER prove that a gap WONT exist]



Well behaved convex $I \Rightarrow$ No duality gap



Not well-behaved convex I (as in the case of semi-definite programs) $\Rightarrow$ Gap might exist



Necessary conditions for constrained optimality : [Page 284 onwards of http://www.cse.iitb.ac.in/~CS709/notes/BasicsOfConvexOptimization.pdf]

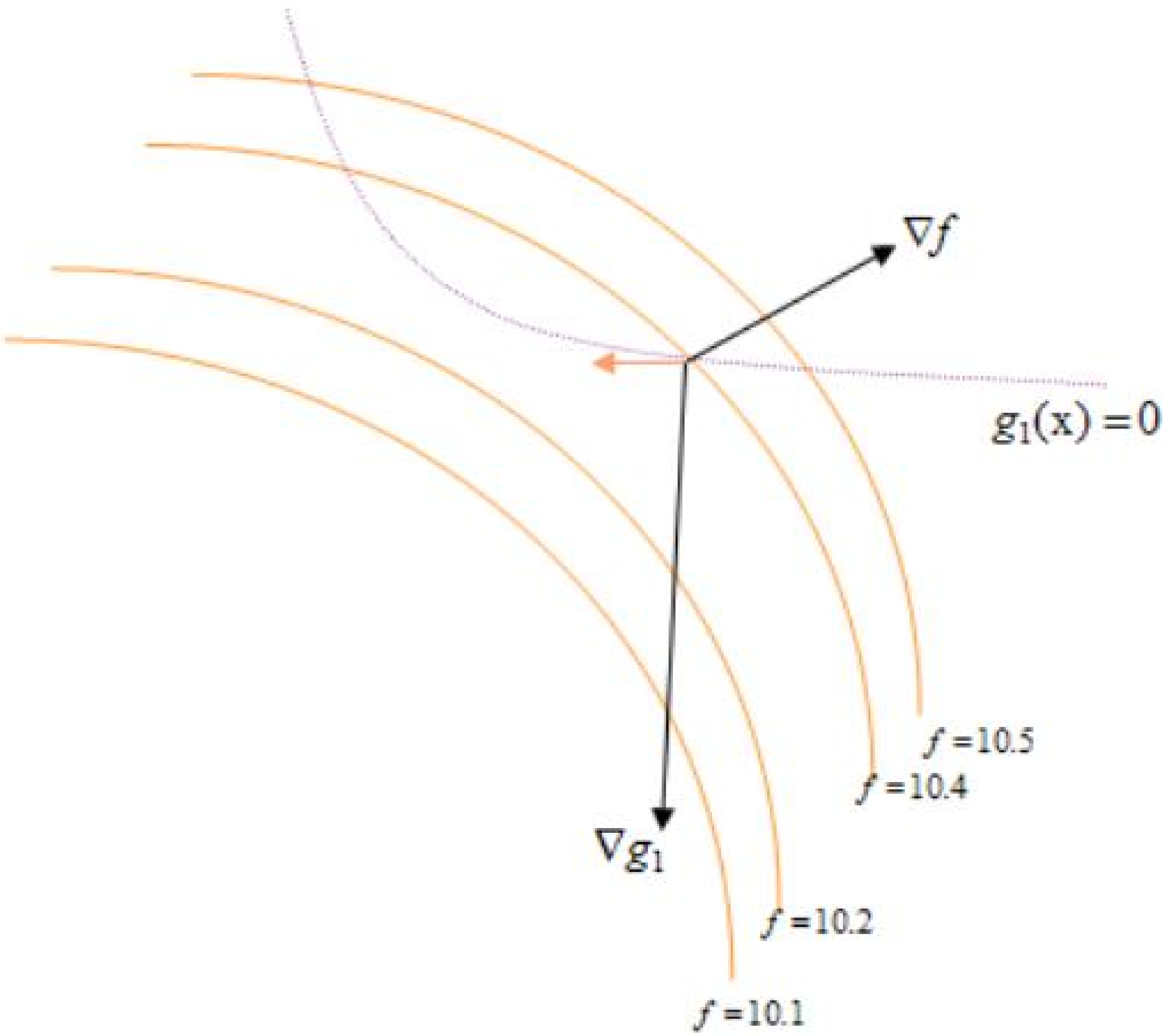


Figure 4.39: At any non-optimal and non-saddle point of the equality constrained problem, the gradient of the constraint will not be parallel to that of the function.

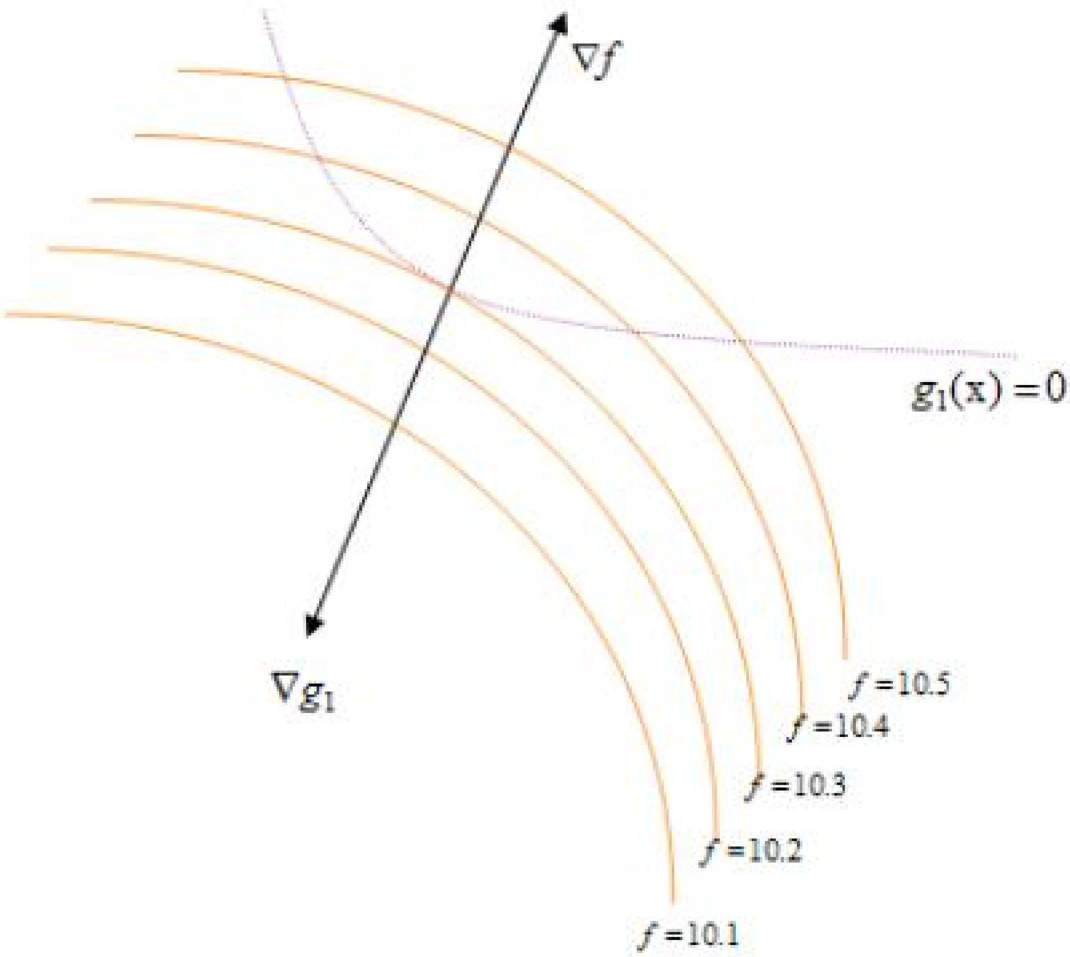


Figure 4.40: At the equality constrained optimum, the gradient of the constraint must be parallel to that of the function.

The necessary condition for an optimum at $\mathbf{x}^*$ for the optimization problem in (4.75) with $m = 1$ can be stated as in (4.76), where the gradient is now $n+1$ dimensional with its last component being a partial derivative with respect to λ.

$$\nabla L(\mathbf{x}^*, \lambda^*) = \nabla f(\mathbf{x}^*) + \lambda^* \nabla g_1(\mathbf{x}^*) = 0 \tag{4.76}$$

Q: what about multiple equality constraints $g_1(x), g_2(x) \cdots g_m(x)$?

We will extend the necessary condition for optimality of a minimization problem with single constraint to minimization problems with multiple equality constraints (*i.e.*, $m > 1$. in (4.75)). Let $\mathcal{S}$ be the subspace spanned by $\nabla g_i(\mathbf{x})$ at any point $\mathbf{x}$ and let $\mathcal{S}_\perp$ be its orthogonal complement. Let $(\nabla f)_\perp$ be the component of ∇f in the subspace $\mathcal{S}_\perp$. At any solution $\mathbf{x}^*$, it must be true that the gradient of f has $(\nabla f)_\perp = 0$ (*i.e.*, no components that are perpendicular to all of the ∇g_i), because otherwise you could move $\mathbf{x}^*$ a little in that direction (or in the opposite direction) to increase (decrease) f without changing any of the g_i, *i.e.* without violating any constraints. Hence for multiple equality constraints, it must be true that at the solution $\mathbf{x}^*$, the space $\mathcal{S}$ contains the vector ∇f, *i.e.*, there are some constants λ_i such that $\nabla f(\mathbf{x}^*) = \lambda_i \nabla g_i(\mathbf{x}^*)$. We also need to impose that the solution is on the correct constraint surface (*i.e.*, $g_i = 0$, $\forall i$). In the same manner as in the case of $m = 1$, this can be encapsulated by introducing the Lagrangian $L(\mathbf{x}, \lambda) = f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x})$, whose gradient with respect to both $\mathbf{x}$, and λ vanishes at the solution.

This gives us the following necessary condition for optimality of (4.75):

$$\nabla L(\mathbf{x}^*, \lambda^*) = \nabla \left(f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) \right) = 0 \tag{4.77}$$

Irrespective of convexity of $f(x)$ or $g_i(x)$

Q: what about inequality constraints?

Now consider the general inequality constrained minimization problem

$$\begin{array}{lll} \min\limits_{\mathbf{x}\in\mathcal{D}} & f(\mathbf{x}) & \\ \text{subject to} & g_i(\mathbf{x}) \le 0 & i = 1, 2, \ldots, m \end{array}$$

See figure below for the case of $m=1$

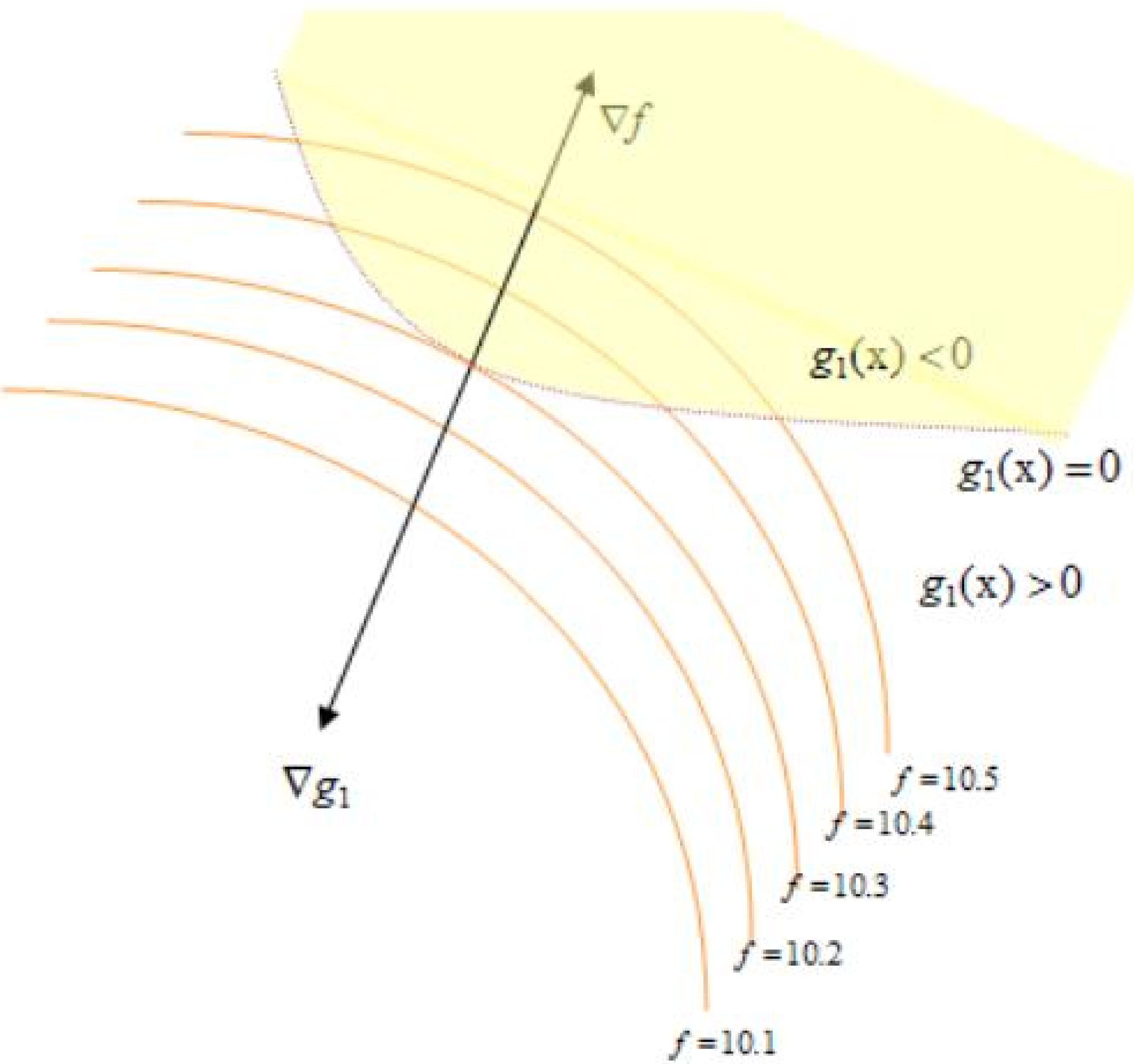


Figure 4.41: At the inequality constrained optimum, the gradient of the constraint must be parallel to that of the function.

Consider

$$L = f + \lambda g_1$$

If $g_1(x^*) < 0$
then $\nabla f(x^*) = 0$
and $\nabla L(x^*) = 0$
by setting $\lambda_1^* = 0$

If $g_1(x^*) = 0$
we have case of equality constrained minimization & therefore

$$\nabla L(x^*) = \nabla f(x^*) - \lambda_1^* \nabla g_1(x^*) = 0$$

In either case:

$$\nabla L(x^*) = 0 \quad \& \quad \lambda_1^* g_1(x^*) = 0$$

Q: What about multiple inequality constraints?

With multiple inequality constraints, for constraints that are active, as in the case of multiple equality constraints, ∇f must lie in the space spanned by the ∇g_i's, and if the Lagrangian is $L = f + \sum_{i=1}^{m} \lambda_i g_i$, then we must additionally have $\lambda_i \geq 0$, $\forall i$ (since otherwise f could be reduced by moving into the feasible region). As for an inactive constraint g_j ($g_j < 0$), removing g_j from L makes no difference and we can drop ∇g_j from $\nabla f = -\sum_{i=1}^{m} \lambda_i \nabla g_i$ or equivalently set $\lambda_j = 0$. Thus, the above KKT condition generalizes to $\lambda_i g_i(\mathbf{x}^*) = 0$, $\forall i$. The necessary condition for optimality of (4.78) is summarily given as

$$\nabla L(\mathbf{x}^*, \lambda^*) = \nabla \left(f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) \right) = 0$$

$$\forall i \quad \lambda_i g_i(\mathbf{x}) = 0 \tag{4.79}$$

Putting together the cases for equality and inequality constraints, we get necessary optimality conditions for any constrained optimization problem [summarized on the next page]

$$
\begin{array}{lll}
\min_{\mathbf{x}\in\mathcal{D}} & f(\mathbf{x}) & \\
\text{subject to} & g_i(\mathbf{x}) \leq 0, & i = 1,\ldots,m \\
& h_j(\mathbf{x}) = 0, & j = 1,\ldots,p \\
\text{variable } \mathbf{x} = (x_1,\ldots,x_n) & &
\end{array}
\tag{4.85}
$$

Suppose that the primal and dual optimal values for the above problem are attained and equal, that is, strong duality holds. Let $\widehat{\mathbf{x}}$ be a primal optimal and $(\widehat{\lambda}, \widehat{\mu})$ be a dual optimal point ($\widehat{\lambda} \in \Re^m, \widehat{\mu} \in \Re^p$). Thus,

$$
\begin{aligned}
f(\widehat{\mathbf{x}}) &= L^*(\widehat{\lambda}, \widehat{\mu}) \\
&= \min_{\mathbf{x}\in\mathcal{D}} f(\mathbf{x}) + \widehat{\lambda}^T \mathbf{g}(\mathbf{x}) + \widehat{\mu}^T \mathbf{h}(\mathbf{x}) \\
&\leq f(\widehat{\mathbf{x}}) + \widehat{\lambda}^T \mathbf{g}(\widehat{\mathbf{x}}) + \widehat{\mu}^T \mathbf{h}(\widehat{\mathbf{x}}) \\
&\leq f(\widehat{\mathbf{x}})
\end{aligned}
$$

The last inequality follows from the fact that $\widehat{\lambda} \geq \mathbf{0}$, $\mathbf{g}(\widehat{\mathbf{x}}) \leq \mathbf{0}$, and $\mathbf{h}(\widehat{\mathbf{x}}) = \mathbf{0}$. We can therefore conclude that the two inequalities in this chain must hold with equality. Some of the conclusions that we can draw from this chain of equalities are

Necessary conditions for optimality of constrained optimization problem are called KKT conditions

Karush-Kuhn-Tucker (KKT) conditions

the following four conditions are called KKT conditions (for a problem with differentiable f_i, h_i):

1. primal constraints: $f_i(x) \leq 0$, $i = 1, \ldots, m$, $h_i(x) = 0$, $i = 1, \ldots, p$
2. dual constraints: $\lambda \succeq 0$
3. complementary slackness: $\lambda_i f_i(x) = 0$, $i = 1, \ldots, m$
4. gradient of Lagrangian with respect to x vanishes:

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0$$

from page 5–17: if strong duality holds and x, λ, ν are optimal, then they must satisfy the KKT conditions

Complementary slackness

assume strong duality holds, $x^\star$ is primal optimal, $(\lambda^\star, \nu^\star)$ is dual optimal

$$\begin{aligned} f_0(x^\star) = g(\lambda^\star, \nu^\star) &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \right) \\ &\leq f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star f_i(x^\star) + \sum_{i=1}^p \nu_i^\star h_i(x^\star) \\ &\leq f_0(x^\star) \end{aligned}$$

hence, the two inequalities hold with equality

- $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$
- $\lambda_i^\star f_i(x^\star) = 0$ for $i = 1, \ldots, m$ (known as complementary slackness):

$$\lambda_i^\star > 0 \Longrightarrow f_i(x^\star) = 0, \qquad f_i(x^\star) < 0 \Longrightarrow \lambda_i^\star = 0$$

Necessary $\Big\downarrow$ conditions

Karush-Kuhn-Tucker (KKT) conditions

the following four conditions are called KKT conditions (for a problem with differentiable f_i, h_i):

1. primal constraints: $f_i(x) \leq 0$, $i = 1, \ldots, m$, $h_i(x) = 0$, $i = 1, \ldots, p$
2. dual constraints: $\lambda \succeq 0$
3. complementary slackness: $\lambda_i f_i(x) = 0$, $i = 1, \ldots, m$
4. gradient of Lagrangian with respect to x vanishes:

$$\nabla f_0(x) + \sum_{i=1}^m \lambda_i \nabla f_i(x) + \sum_{i=1}^p \nu_i \nabla h_i(x) = 0$$

from page 5–17: if strong duality holds and x, λ, ν are optimal, then they must satisfy the KKT conditions

(Sufficiency of) KKT conditions for convex problem

if $\tilde{x}$, $\tilde{\lambda}$, $\tilde{\nu}$ satisfy KKT for a convex problem, then they are optimal:

- from complementary slackness: $f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$
- from 4th condition (and convexity): $g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$

hence, $f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu})$ $\Rightarrow$

if **Slater's condition** is satisfied:

x is optimal if and only if there exist λ, ν that satisfy KKT conditions

- recall that Slater implies strong duality, and dual optimum is attained
- generalizes optimality condition $\nabla f_0(x) = 0$ for unconstrained problem

example: water-filling (assume $\alpha_i > 0$)

$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^n \log(x_i + \alpha_i) \\ \text{subject to} & x \succeq 0, \quad \mathbf{1}^T x = 1 \end{array}$$

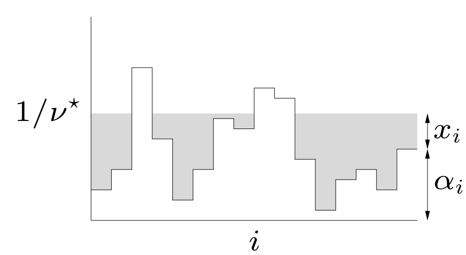
x is optimal iff $x \succeq 0$, $\mathbf{1}^T x = 1$, and there exist $\lambda \in \mathbf{R}^n$, $\nu \in \mathbf{R}$ such that

$$\lambda \succeq 0, \qquad \lambda_i x_i = 0, \qquad \frac{1}{x_i + \alpha_i} + \lambda_i = \nu$$

- if $\nu < 1/\alpha_i$: $\lambda_i = 0$ and $x_i = 1/\nu - \alpha_i$
- if $\nu \geq 1/\alpha_i$: $\lambda_i = \nu - 1/\alpha_i$ and $x_i = 0$
- determine ν from $\mathbf{1}^T x = \sum_{i=1}^n \max\{0, 1/\nu - \alpha_i\} = 1$

interpretation

- n patches; level of patch i is at height α_i
- flood area with unit amount of water
- resulting level is $1/\nu^\star$



Perturbation and sensitivity analysis

(unperturbed) optimization problem and its dual

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array} \qquad \begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

perturbed problem and its dual

$$\begin{array}{ll} \text{min.} & f_0(x) \\ \text{s.t.} & f_i(x) \leq u_i, \quad i = 1, \ldots, m \\ & h_i(x) = v_i, \quad i = 1, \ldots, p \end{array} \qquad \begin{array}{ll} \text{max.} & g(\lambda, \nu) - u^T\lambda - v^T\nu \\ \text{s.t.} & \lambda \succeq 0 \end{array}$$

- x is primal variable; u, v are parameters
- $p^\star(u, v)$ is optimal value as a function of u, v
- we are interested in information about $p^\star(u, v)$ that we can obtain from the solution of the unperturbed problem and its dual

global sensitivity result

assume strong duality holds for unperturbed problem, and that $\lambda^\star$, $\nu^\star$ are dual optimal for unperturbed problem

apply weak duality to perturbed problem:

$$\begin{array}{rcl} p^\star(u, v) & \geq & g(\lambda^\star, \nu^\star) - u^T\lambda^\star - v^T\nu^\star \\ & = & p^\star(0, 0) - u^T\lambda^\star - v^T\nu^\star \end{array}$$

sensitivity interpretation

- if $\lambda_i^\star$ large: $p^\star$ increases greatly if we tighten constraint i $(u_i < 0)$
- if $\lambda_i^\star$ small: $p^\star$ does not decrease much if we loosen constraint i $(u_i > 0)$
- if $\nu_i^\star$ large and positive: $p^\star$ increases greatly if we take $v_i < 0$; if $\nu_i^\star$ large and negative: $p^\star$ increases greatly if we take $v_i > 0$
- if $\nu_i^\star$ small and positive: $p^\star$ does not decrease much if we take $v_i > 0$; if $\nu_i^\star$ small and negative: $p^\star$ does not decrease much if we take $v_i < 0$

local sensitivity: if (in addition) $p^\star(u,v)$ is differentiable at $(0,0)$, then

$$\lambda_i^\star = -\frac{\partial p^\star(0,0)}{\partial u_i}, \qquad \nu_i^\star = -\frac{\partial p^\star(0,0)}{\partial v_i}$$

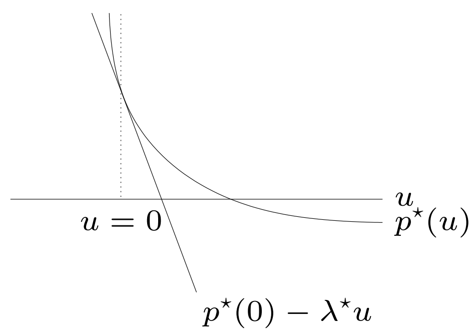
proof (for $\lambda_i^\star$): from global sensitivity result,

$$\frac{\partial p^\star(0,0)}{\partial u_i} = \lim_{t \searrow 0} \frac{p^\star(te_i,0) - p^\star(0,0)}{t} \geq -\lambda_i^\star$$

$$\frac{\partial p^\star(0,0)}{\partial u_i} = \lim_{t \nearrow 0} \frac{p^\star(te_i,0) - p^\star(0,0)}{t} \leq -\lambda_i^\star$$

hence, equality

$p^\star(u)$ for a problem with one (inequality) constraint:



Duality and problem reformulations

- equivalent formulations of a problem can lead to very different duals
- reformulating the primal problem can be useful when the dual is difficult to derive, or uninteresting

common reformulations

- introduce new variables and equality constraints
- make explicit constraints implicit or vice-versa
- transform objective or constraint functions

  $e.g.$, replace $f_0(x)$ by $\phi(f_0(x))$ with ϕ convex, increasing

Introducing new variables and equality constraints

$$\text{minimize} \quad f_0(Ax+b)$$

- dual function is constant: $g = \inf_x L(x) = \inf_x f_0(Ax+b) = p^\star$
- we have strong duality, but dual is quite useless

reformulated problem and its dual

$$\begin{array}{ll} \text{minimize} & f_0(y) \\ \text{subject to} & Ax+b-y=0 \end{array} \qquad \begin{array}{ll} \text{maximize} & b^T\nu - f_0^*(\nu) \\ \text{subject to} & A^T\nu = 0 \end{array}$$

dual function follows from

$$\begin{aligned} g(\nu) &= \inf_{x,y}(f_0(y) - \nu^T y + \nu^T Ax + b^T\nu) \\ &= \begin{cases} -f_0^*(\nu) + b^T\nu & A^T\nu = 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

norm approximation problem: minimize $\|Ax-b\|$

$$\begin{array}{ll} \text{minimize} & \|y\| \\ \text{subject to} & y = Ax-b \end{array}$$

can look up conjugate of $\|\cdot\|$, or derive dual directly

$$\begin{aligned} g(\nu) &= \inf_{x,y}(\|y\| + \nu^T y - \nu^T Ax + b^T\nu) \\ &= \begin{cases} b^T\nu + \inf_y(\|y\| + \nu^T y) & A^T\nu = 0 \\ -\infty & \text{otherwise} \end{cases} \\ &= \begin{cases} b^T\nu & A^T\nu = 0, \quad \|\nu\|_* \le 1 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

(see page 5–4)

dual of norm approximation problem

$$\begin{array}{ll} \text{maximize} & b^T\nu \\ \text{subject to} & A^T\nu = 0, \quad \|\nu\|_* \le 1 \end{array}$$

Implicit constraints

LP with box constraints: primal and dual problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & -\mathbf{1} \preceq x \preceq \mathbf{1} \end{array} \qquad \begin{array}{ll} \text{maximize} & -b^T\nu - \mathbf{1}^T\lambda_1 - \mathbf{1}^T\lambda_2 \\ \text{subject to} & c + A^T\nu + \lambda_1 - \lambda_2 = 0 \\ & \lambda_1 \succeq 0, \quad \lambda_2 \succeq 0 \end{array}$$

reformulation with box constraints made implicit

$$\begin{array}{ll} \text{minimize} & f_0(x) = \begin{cases} c^T x & -\mathbf{1} \preceq x \preceq \mathbf{1} \\ \infty & \text{otherwise} \end{cases} \\ \text{subject to} & Ax = b \end{array}$$

dual function

$$\begin{aligned} g(\nu) &= \inf_{-\mathbf{1}\preceq x\preceq \mathbf{1}} (c^T x + \nu^T(Ax - b)) \\ &= -b^T\nu - \|A^T\nu + c\|_1 \end{aligned}$$

dual problem: maximize $-b^T\nu - \|A^T\nu + c\|_1$

Problems with generalized inequalities

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \preceq_{K_i} 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

$\preceq_{K_i}$ is generalized inequality on $\mathbf{R}^{k_i}$

definitions are parallel to scalar case:

- Lagrange multiplier for $f_i(x) \preceq_{K_i} 0$ is vector $\lambda_i \in \mathbf{R}^{k_i}$
- Lagrangian $L : \mathbf{R}^n \times \mathbf{R}^{k_1} \times \cdots \times \mathbf{R}^{k_m} \times \mathbf{R}^p \to \mathbf{R}$, is defined as

$$L(x, \lambda_1, \cdots, \lambda_m, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i^T f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

- dual function $g : \mathbf{R}^{k_1} \times \cdots \times \mathbf{R}^{k_m} \times \mathbf{R}^p \to \mathbf{R}$, is defined as

$$g(\lambda_1, \ldots, \lambda_m, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda_1, \cdots, \lambda_m, \nu)$$

lower bound property: if $\lambda_i \succeq_{K_i^*} 0$, then $g(\lambda_1, \ldots, \lambda_m, \nu) \leq p^\star$

proof: if $\tilde{x}$ is feasible and $\lambda \succeq_{K_i^*} 0$, then

$$\begin{aligned} f_0(\tilde{x}) &\geq f_0(\tilde{x}) + \sum_{i=1}^{m} \lambda_i^T f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) \\ &\geq \inf_{x \in \mathcal{D}} L(x, \lambda_1, \ldots, \lambda_m, \nu) \\ &= g(\lambda_1, \ldots, \lambda_m, \nu) \end{aligned}$$

minimizing over all feasible $\tilde{x}$ gives $p^\star \geq g(\lambda_1, \ldots, \lambda_m, \nu)$

dual problem

$$\begin{array}{ll} \text{maximize} & g(\lambda_1, \ldots, \lambda_m, \nu) \\ \text{subject to} & \lambda_i \succeq_{K_i^*} 0, \quad i = 1, \ldots, m \end{array}$$

- weak duality: $p^\star \geq d^\star$ always
- strong duality: $p^\star = d^\star$ for convex problem with constraint qualification (for example, Slater's: primal problem is strictly feasible)

Semidefinite program

primal SDP $(F_i, G \in \mathbf{S}^k)$

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 F_1 + \cdots + x_n F_n \preceq G \end{array}$$

- Lagrange multiplier is matrix $Z \in \mathbf{S}^k$
- Lagrangian $L(x, Z) = c^T x + \mathbf{tr}\left(Z(x_1 F_1 + \cdots + x_n F_n - G)\right)$
- dual function

$$g(Z) = \inf_x L(x, Z) = \begin{cases} -\mathbf{tr}(GZ) & \mathbf{tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n \\ -\infty & \text{otherwise} \end{cases}$$

dual SDP

$$\begin{array}{ll} \text{maximize} & -\mathbf{tr}(GZ) \\ \text{subject to} & Z \succeq 0, \quad \mathbf{tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n \end{array}$$

$p^\star = d^\star$ if primal SDP is strictly feasible ($\exists x$ with $x_1 F_1 + \cdots + x_n F_n \prec G$)

12. Interior-point methods

- inequality constrained minimization
- logarithmic barrier function and central path
- barrier method
- feasibility and phase I methods
- complexity analysis via self-concordance
- generalized inequalities

Inequality constrained minimization

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array} \tag{1}$$

- f_i convex, twice continuously differentiable
- $A \in \mathbf{R}^{p \times n}$ with $\mathbf{rank}\, A = p$
- we assume $p^\star$ is finite and attained
- we assume problem is strictly feasible: there exists $\tilde{x}$ with

$$\tilde{x} \in \mathbf{dom}\, f_0, \qquad f_i(\tilde{x}) < 0, \quad i = 1, \ldots, m, \qquad A\tilde{x} = b$$

hence, strong duality holds and dual optimum is attained

Examples

- LP, QP, QCQP, GP
- entropy maximization with linear inequality constraints

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n x_i \log x_i \\ \text{subject to} & Fx \preceq g \\ & Ax = b \end{array}$$

  with $\mathbf{dom}\, f_0 = \mathbf{R}^n_{++}$

- differentiability may require reformulating the problem, $e.g.$, piecewise-linear minimization or ℓ_∞-norm approximation via LP
- SDPs and SOCPs are better handled as problems with generalized inequalities (see later)

Logarithmic barrier

reformulation of (1) via indicator function:

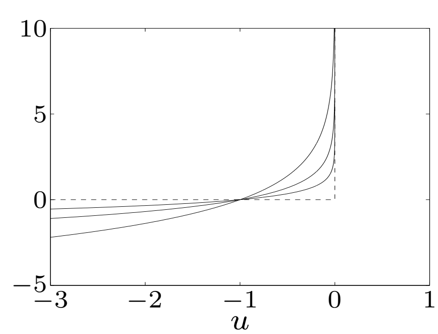
$$\begin{array}{ll} \text{minimize} & f_0(x) + \sum_{i=1}^m I_-(f_i(x)) \\ \text{subject to} & Ax = b \end{array}$$

where $I_-(u) = 0$ if $u \leq 0$, $I_-(u) = \infty$ otherwise (indicator function of $\mathbf{R}_-$)

approximation via logarithmic barrier

$$\begin{array}{ll} \text{minimize} & f_0(x) - (1/t) \sum_{i=1}^m \log(-f_i(x)) \\ \text{subject to} & Ax = b \end{array}$$

- an equality constrained problem
- for $t > 0$, $-(1/t)\log(-u)$ is a smooth approximation of I_-
- approximation improves as $t \to \infty$



logarithmic barrier function

$$\phi(x) = -\sum_{i=1}^{m} \log(-f_i(x)), \quad \mathbf{dom}\,\phi = \{x \mid f_1(x) < 0, \ldots, f_m(x) < 0\}$$

- convex (follows from composition rules) [page 48 of http://www.cse.iitb.ac.in/~CS709/notes/eNotes/basicsOfUnivariateOptAndItsGeneralisation-withHighlights.pdf]
- twice continuously differentiable, with derivatives

$$\begin{aligned}
\nabla\phi(x) &= \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla f_i(x) \\
\nabla^2\phi(x) &= \sum_{i=1}^{m} \frac{1}{f_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T + \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla^2 f_i(x)
\end{aligned}$$

Central path

- for $t > 0$, define $x^\star(t)$ as the solution of

$$\begin{array}{ll}
\text{minimize} & t f_0(x) + \phi(x) \\
\text{subject to} & Ax = b
\end{array}$$

$\rightarrow$ gives $x = x_{\text{particular}} + x_{\text{null space}}$

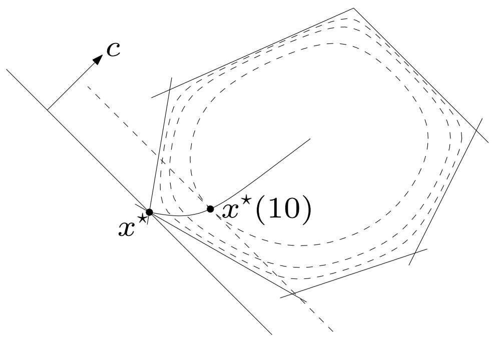
(for now, assume $x^\star(t)$ exists and is unique for each $t > 0$)

- central path is $\{x^\star(t) \mid t > 0\}$

example: central path for an LP

$$\begin{array}{ll}
\text{minimize} & c^T x \\
\text{subject to} & a_i^T x \le b_i, \quad i = 1, \ldots, 6
\end{array}$$

hyperplane $c^T x = c^T x^\star(t)$ is tangent to level curve of ϕ through $x^\star(t)$



Dual points on central path

$x = x^\star(t)$ if there exists a w such that

$$t\nabla f_0(x) + \sum_{i=1}^m \frac{1}{-f_i(x)}\nabla f_i(x) + A^T w = 0, \qquad Ax = b$$

- therefore, $x^\star(t)$ minimizes the Lagrangian

$$L(x, \lambda^\star(t), \nu^\star(t)) = f_0(x) + \sum_{i=1}^m \lambda_i^\star(t) f_i(x) + \nu^\star(t)^T(Ax - b)$$

  where we define $\lambda_i^\star(t) = 1/(-tf_i(x^\star(t))$ and $\nu^\star(t) = w/t$

- this confirms the intuitive idea that $f_0(x^\star(t)) \to p^\star$ if $t \to \infty$:

$$\begin{aligned} p^\star &\geq g(\lambda^\star(t), \nu^\star(t)) \\ &= L(x^\star(t), \lambda^\star(t), \nu^\star(t)) \\ &= f_0(x^\star(t)) - m/t \end{aligned}$$

Duality gap

Interpretation via KKT conditions

$x = x^\star(t)$, $\lambda = \lambda^\star(t)$, $\nu = \nu^\star(t)$ satisfy

1. primal constraints: $f_i(x) \leq 0$, $i = 1, \ldots, m$, $Ax = b$
2. dual constraints: $\lambda \succeq 0$
3. approximate complementary slackness: $-\lambda_i f_i(x) = 1/t$, $i = 1, \ldots, m$
4. gradient of Lagrangian with respect to x vanishes:

$$\nabla f_0(x) + \sum_{i=1}^m \lambda_i \nabla f_i(x) + A^T \nu = 0$$

difference with KKT is that condition 3 replaces $\lambda_i f_i(x) = 0$

Barrier method

---

given strictly feasible x, $t := t^{(0)} > 0$, $\mu > 1$, tolerance $\epsilon > 0$.

repeat

1. *Centering step.* Compute $x^\star(t)$ by minimizing $tf_0 + \phi$, subject to $Ax = b$.
2. *Update.* $x := x^\star(t)$.
3. *Stopping criterion.* **quit** if $m/t < \epsilon$.
4. *Increase* t. $t := \mu t$.

Duality gap check for convergence

- terminates with $f_0(x) - p^\star \leq \epsilon$ (stopping criterion follows from $f_0(x^\star(t)) - p^\star \leq m/t$)
- centering usually done using Newton's method, starting at current x
- choice of μ involves a trade-off: large μ means fewer outer iterations, more inner (Newton) iterations; typical values: $\mu = 10$–20
- several heuristics for choice of $t^{(0)}$

Convergence analysis

number of outer (centering) iterations: exactly

$$\left\lceil \frac{\log(m/(\epsilon t^{(0)}))}{\log \mu} \right\rceil$$

plus the initial centering step (to compute $x^\star(t^{(0)})$)

centering problem

$$\text{minimize} \quad tf_0(x) + \phi(x)$$

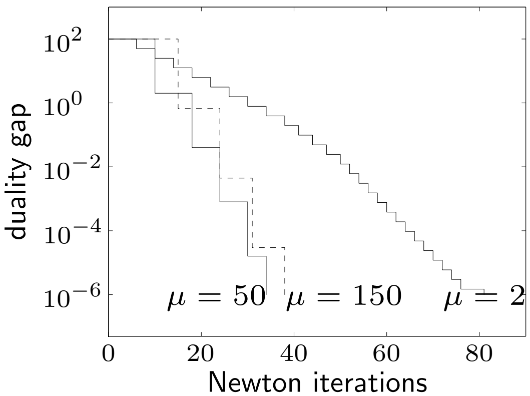
can be solved using any unconstrained optimisation algo

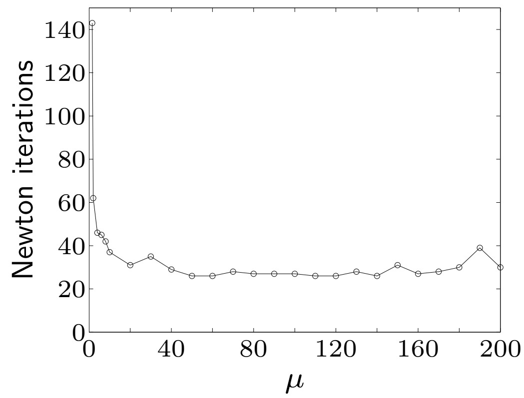
see convergence analysis of Newton's method

- $tf_0 + \phi$ must have closed sublevel sets for $t \geq t^{(0)}$
- classical analysis requires strong convexity, Lipschitz condition
- analysis via self-concordance requires self-concordance of $tf_0 + \phi$

Examples

inequality form LP ($m = 100$ inequalities, $n = 50$ variables)

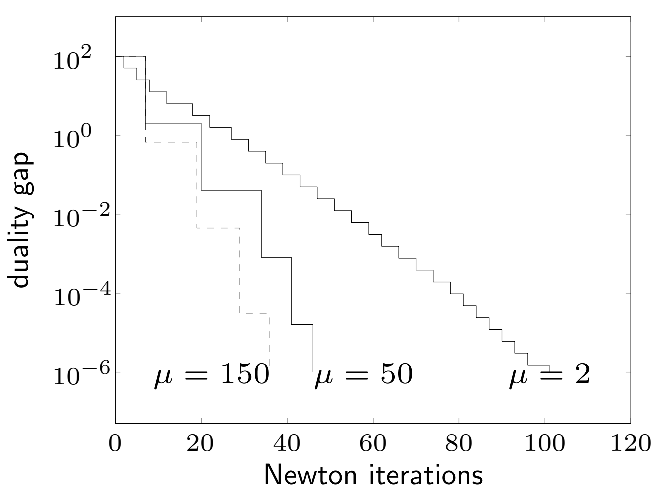




- starts with x on central path ($t^{(0)} = 1$, duality gap 100)
- terminates when $t = 10^8$ (gap 10^{-6})
- centering uses Newton's method with backtracking
- total number of Newton iterations not very sensitive for $\mu \geq 10$

geometric program ($m = 100$ inequalities and $n = 50$ variables)

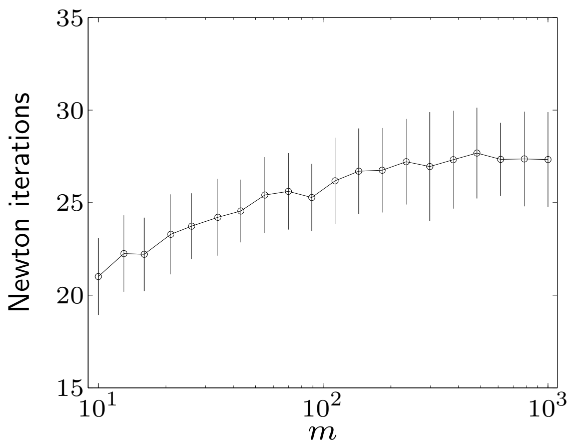
$$\begin{array}{ll} \text{minimize} & \log\left(\sum_{k=1}^{5} \exp(a_{0k}^T x + b_{0k})\right) \\ \text{subject to} & \log\left(\sum_{k=1}^{5} \exp(a_{ik}^T x + b_{ik})\right) \leq 0, \quad i = 1, \ldots, m \end{array}$$



family of standard LPs ($A \in \mathbf{R}^{m \times 2m}$)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, \quad x \succeq 0 \end{array}$$

$m = 10, \ldots, 1000$; for each m, solve 100 randomly generated instances



number of iterations grows very slowly as m ranges over a $100:1$ ratio

Feasibility and phase I methods

feasibility problem: find x such that

$$f_i(x) \leq 0, \quad i = 1, \ldots, m, \qquad Ax = b \tag{2}$$

phase I: computes strictly feasible starting point for barrier method

basic phase I method

$$\begin{array}{ll} \text{minimize (over } x, s) & s \\ \text{subject to} & f_i(x) \leq s, \quad i = 1, \ldots, m \\ & Ax = b \end{array} \tag{3}$$

Posing feasibility problem as an opt prob

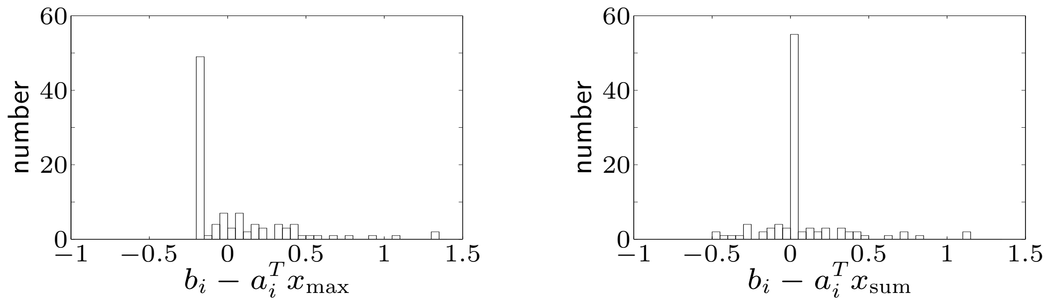
- if x, s feasible, with $s < 0$, then x is strictly feasible for (2)
- if optimal value $\bar{p}^\star$ of (3) is positive, then problem (2) is infeasible
- if $\bar{p}^\star = 0$ and attained, then problem (2) is feasible (but not strictly);
  if $\bar{p}^\star = 0$ and not attained, then problem (2) is infeasible

sum of infeasibilities phase I method

$$
\begin{array}{ll}
\text{minimize} & \mathbf{1}^T s \\
\text{subject to} & s \succeq 0, \quad f_i(x) \leq s_i, \quad i = 1, \ldots, m \\
& Ax = b
\end{array}
$$

for infeasible problems, produces a solution that satisfies many more inequalities than basic phase I method

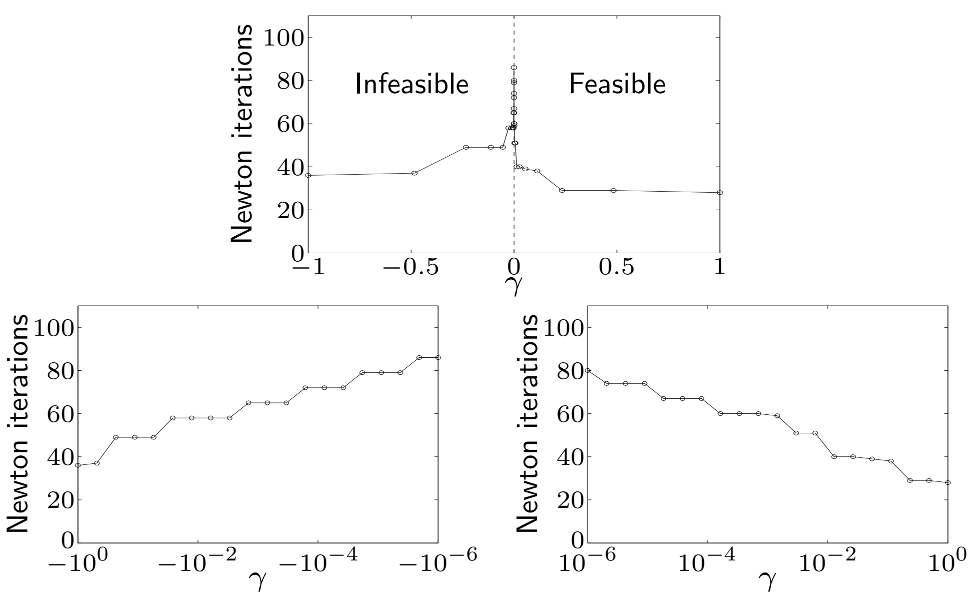
example (infeasible set of 100 linear inequalities in 50 variables)



left: basic phase I solution; satisfies 39 inequalities
right: sum of infeasibilities phase I solution; satisfies 79 solutions

example: family of linear inequalities $Ax \preceq b + \gamma \Delta b$

- data chosen to be strictly feasible for $\gamma > 0$, infeasible for $\gamma \leq 0$
- use basic phase I, terminate when $s < 0$ or dual objective is positive



number of iterations roughly proportional to $\log(1/|\gamma|)$

Complexity analysis via self-concordance

(Like in case of unconstrained optimization using Newton method)

same assumptions as on page 12–2, plus:

- sublevel sets (of f_0, on the feasible set) are bounded
- $tf_0 + \phi$ is self-concordant with closed sublevel sets

second condition

- holds for LP, QP, QCQP
- may require reformulating the problem, *e.g.*,

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n x_i \log x_i \\ \text{subject to} & Fx \preceq g \end{array} \quad \longrightarrow \quad \begin{array}{ll} \text{minimize} & \sum_{i=1}^n x_i \log x_i \\ \text{subject to} & Fx \preceq g, \quad x \succeq 0 \end{array}$$

- needed for complexity analysis; barrier method works even when self-concordance assumption does not apply

Newton iterations per centering step: from self-concordance theory

$$\#\text{Newton iterations} \leq \frac{\mu t f_0(x) + \phi(x) - \mu t f_0(x^+) - \phi(x^+)}{\gamma} + c$$

- bound on effort of computing $x^+ = x^\star(\mu t)$ starting at $x = x^\star(t)$
- γ, c are constants (depend only on Newton algorithm parameters)
- from duality (with $\lambda = \lambda^\star(t)$, $\nu = \nu^\star(t)$):

$$\begin{array}{l} \mu t f_0(x) + \phi(x) - \mu t f_0(x^+) - \phi(x^+) \\ \quad = \quad \mu t f_0(x) - \mu t f_0(x^+) + \displaystyle\sum_{i=1}^m \log(-\mu t \lambda_i f_i(x^+)) - m \log \mu \\ \quad \leq \quad \mu t f_0(x) - \mu t f_0(x^+) - \mu t \displaystyle\sum_{i=1}^m \lambda_i f_i(x^+) - m - m \log \mu \\ \quad \leq \quad \mu t f_0(x) - \mu t g(\lambda, \nu) - m - m \log \mu \\ \quad = \quad m(\mu - 1 - \log \mu) \end{array}$$

total number of Newton iterations (excluding first centering step)

$$\#\text{Newton iterations} \leq N = \left\lceil \frac{\log(m/(t^{(0)}\epsilon))}{\log \mu} \right\rceil \left(\frac{m(\mu - 1 - \log \mu)}{\gamma} + c \right)$$

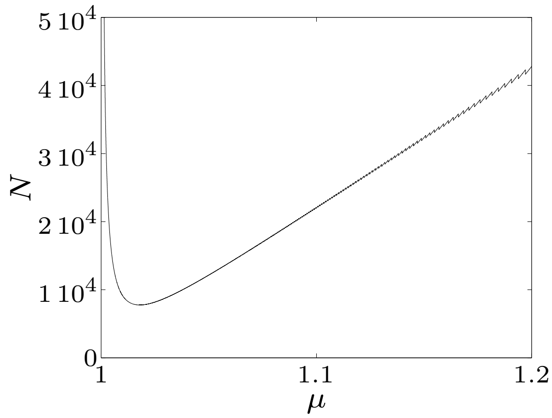


figure shows N for typical values of γ, c,

$$m = 100, \qquad \frac{m}{t^{(0)}\epsilon} = 10^5$$

- confirms trade-off in choice of μ
- in practice, #iterations is in the tens; not very sensitive for $\mu \geq 10$

polynomial-time complexity of barrier method

- for $\mu = 1 + 1/\sqrt{m}$:

$$N = O\left(\sqrt{m}\log\left(\frac{m/t^{(0)}}{\epsilon}\right)\right)$$

- number of Newton iterations for fixed gap reduction is $O(\sqrt{m})$
- multiply with cost of one Newton iteration (a polynomial function of problem dimensions), to get bound on number of flops

this choice of μ optimizes worst-case complexity; in practice we choose μ fixed ($\mu = 10, \ldots, 20$)

Generalized inequalities

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \preceq_{K_i} 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

- f_0 convex, $f_i : \mathbf{R}^n \to \mathbf{R}^{k_i}$, $i = 1, \ldots, m$, convex with respect to proper cones $K_i \in \mathbf{R}^{k_i}$
- f_i twice continuously differentiable
- $A \in \mathbf{R}^{p \times n}$ with $\mathbf{rank}\, A = p$
- we assume $p^\star$ is finite and attained
- we assume problem is strictly feasible; hence strong duality holds and dual optimum is attained

examples of greatest interest: SOCP, SDP

Generalized logarithm for proper cone

$\psi : \mathbf{R}^q \to \mathbf{R}$ is generalized logarithm for proper cone $K \subseteq \mathbf{R}^q$ if:

- $\mathbf{dom}\, \psi = \mathbf{int}\, K$ and $\nabla^2 \psi(y) \prec 0$ for $y \succ_K 0$
- $\psi(sy) = \psi(y) + \theta \log s$ for $y \succ_K 0$, $s > 0$ (θ is the degree of ψ)

examples

- nonnegative orthant $K = \mathbf{R}^n_+$: $\psi(y) = \sum_{i=1}^n \log y_i$, with degree $\theta = n$
- positive semidefinite cone $K = \mathbf{S}^n_+$:

$$\psi(Y) = \log \det Y \qquad (\theta = n)$$

For convexity analysis, ref page 42 of http://www.cse.iitb.ac.in/~CS709/notes/eNotes/basicsOfUnivariateOptAndItsGeneralisation-withHighlights.pdf

- second-order cone $K = \{y \in \mathbf{R}^{n+1} \mid (y_1^2 + \cdots + y_n^2)^{1/2} \leq y_{n+1}\}$:

$$\psi(y) = \log(y_{n+1}^2 - y_1^2 - \cdots - y_n^2) \qquad (\theta = 2)$$

properties (without proof): for $y \succ_K 0$,

$$\nabla\psi(y) \succeq_{K^*} 0, \qquad y^T\nabla\psi(y) = \theta$$

- nonnegative orthant $\mathbf{R}^n_+$: $\psi(y) = \sum_{i=1}^n \log y_i$

$$\nabla\psi(y) = (1/y_1, \ldots, 1/y_n), \qquad y^T\nabla\psi(y) = n$$

- positive semidefinite cone $\mathbf{S}^n_+$: $\psi(Y) = \log\det Y$

$$\nabla\psi(Y) = Y^{-1}, \qquad \mathbf{tr}(Y\nabla\psi(Y)) = n$$

- second-order cone $K = \{y \in \mathbf{R}^{n+1} \mid (y_1^2 + \cdots + y_n^2)^{1/2} \leq y_{n+1}\}$:

$$\psi(y) = \frac{2}{y_{n+1}^2 - y_1^2 - \cdots - y_n^2}\begin{bmatrix} -y_1 \\ \vdots \\ -y_n \\ y_{n+1} \end{bmatrix}, \qquad y^T\nabla\psi(y) = 2$$

Logarithmic barrier and central path

logarithmic barrier for $f_1(x) \preceq_{K_1} 0, \ldots, f_m(x) \preceq_{K_m} 0$:

$$\phi(x) = -\sum_{i=1}^m \psi_i(-f_i(x)), \qquad \mathbf{dom}\,\phi = \{x \mid f_i(x) \prec_{K_i} 0,\ i = 1, \ldots, m\}$$

- ψ_i is generalized logarithm for K_i, with degree θ_i
- ϕ is convex, twice continuously differentiable

central path: $\{x^\star(t) \mid t > 0\}$ where $x^\star(t)$ solves

$$\begin{array}{ll} \text{minimize} & tf_0(x) + \phi(x) \\ \text{subject to} & Ax = b \end{array}$$

Dual points on central path

$x = x^\star(t)$ if there exists $w \in \mathbf{R}^p$,

$$t\nabla f_0(x) + \sum_{i=1}^m Df_i(x)^T \nabla\psi_i(-f_i(x)) + A^T w = 0$$

($Df_i(x) \in \mathbf{R}^{k_i \times n}$ is derivative matrix of f_i)

- therefore, $x^\star(t)$ minimizes Lagrangian $L(x, \lambda^\star(t), \nu^\star(t))$, where

$$\lambda_i^\star(t) = \frac{1}{t}\nabla\psi_i(-f_i(x^\star(t))), \qquad \nu^\star(t) = \frac{w}{t}$$

- from properties of ψ_i: $\lambda_i^\star(t) \succ_{K_i^*} 0$, with duality gap

$$f_0(x^\star(t)) - g(\lambda^\star(t), \nu^\star(t)) = (1/t)\sum_{i=1}^m \theta_i$$

example: semidefinite programming (with $F_i \in \mathbf{S}^p$)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & F(x) = \sum_{i=1}^n x_i F_i + G \preceq 0 \end{array}$$

- logarithmic barrier: $\phi(x) = \log\det(-F(x)^{-1})$
- central path: $x^\star(t)$ minimizes $tc^T x - \log\det(-F(x))$; hence

$$tc_i - \mathbf{tr}(F_i F(x^\star(t))^{-1}) = 0, \quad i = 1, \ldots, n$$

- dual point on central path: $Z^\star(t) = -(1/t)F(x^\star(t))^{-1}$ is feasible for

$$\begin{array}{ll} \text{maximize} & \mathbf{tr}(GZ) \\ \text{subject to} & \mathbf{tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n \\ & Z \succeq 0 \end{array}$$

- duality gap on central path: $c^T x^\star(t) - \mathbf{tr}(GZ^\star(t)) = p/t$

Barrier method

given strictly feasible x, $t := t^{(0)} > 0$, $\mu > 1$, tolerance $\epsilon > 0$.

repeat

1. *Centering step.* Compute $x^\star(t)$ by minimizing $tf_0 + \phi$, subject to $Ax = b$.
2. *Update.* $x := x^\star(t)$.
3. *Stopping criterion.* **quit** if $(\sum_i \theta_i)/t < \epsilon$.
4. *Increase* t. $t := \mu t$.

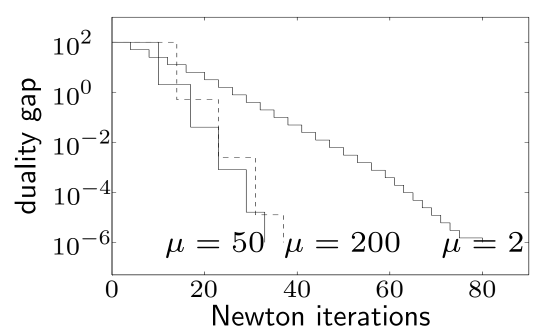
- only difference is duality gap m/t on central path is replaced by $\sum_i \theta_i/t$
- number of outer iterations:

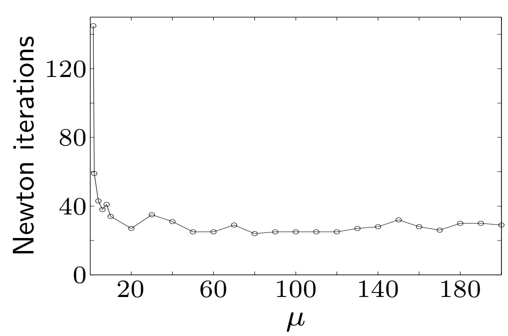
$$\left\lceil \frac{\log((\sum_i \theta_i)/(\epsilon t^{(0)}))}{\log \mu} \right\rceil$$

- complexity analysis via self-concordance applies to SDP, SOCP

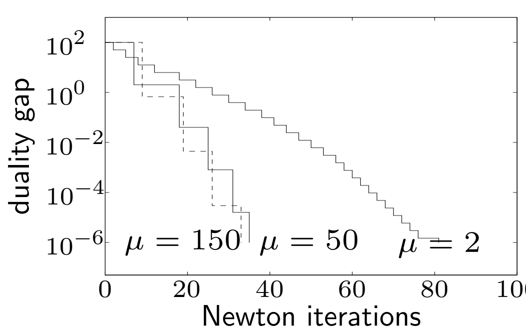
Examples

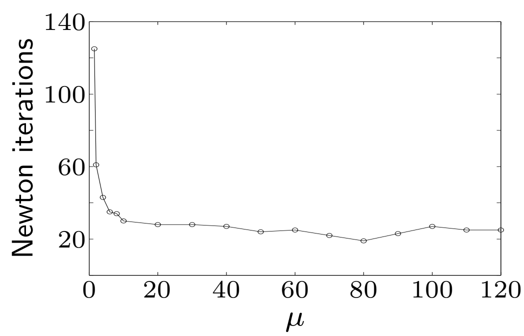
second-order cone program (50 variables, 50 SOC constraints in $\mathbf{R}^6$)





semidefinite program (100 variables, LMI constraint in $\mathbf{S}^{100}$)

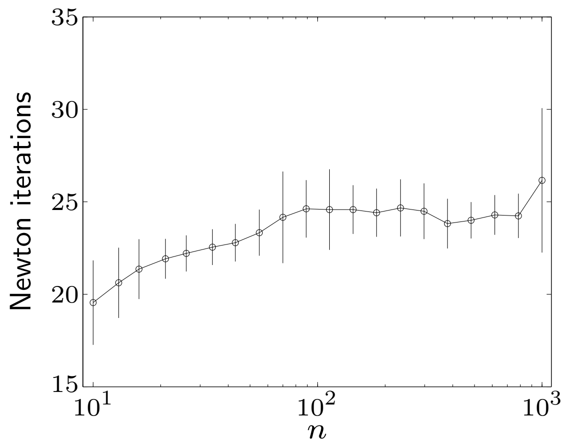




family of SDPs ($A \in \mathbf{S}^n$, $x \in \mathbf{R}^n$)

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T x \\ \text{subject to} & A + \mathbf{diag}(x) \succeq 0 \end{array}$$

$n = 10, \ldots, 1000$, for each n solve 100 randomly generated instances



Primal-dual interior-point methods

more efficient than barrier method when high accuracy is needed

- update primal and dual variables at each iteration; no distinction between inner and outer iterations
- often exhibit superlinear asymptotic convergence
- search directions can be interpreted as Newton directions for modified KKT conditions
- can start at infeasible points
- cost per iteration same as barrier method