

# ON THE IMPORTANCE OF DUALITY IN OPTIMAL DESIGN

Aharon Ben-Tal

MINERVA Optimization Center  
Technion—Israel Institute of Technology

# WHAT IS “DUALITY THEORY”?

Kuhn (SIAM-AMS Proc., Vol. 9, 1976):

“A Duality Theory is made of the following elements:

- (a) A pair of optimization problems, one a minimization problem (problem ( $P$ )) and the other a maximization problem (program ( $D$ )) based on the same data.
- (b) Weak duality holds:  $\min(P) \geq \max(D)$ .
- (c) Necessary and sufficient conditions for optimality of a feasible pair is the equality of the corresponding objective function.”

## My addition

- (d) A computable (tractable) relation between the optimal solutions of ( $P$ ) and ( $D$ ).
- (e) Relation between: primal feasibility and dual boundedness and dual attainability.

# LAGRANGIAN DUALITY

$$(P) \quad \boxed{\inf_{x \in \mathbb{R}^n} \{f(x) \mid g_i(x) \leq 0, \quad i = 1, \dots, m\}}$$

$f, g_i$  convex

Lagrangian:  $L(x, y) = f(x) + \sum_{i=1}^n y_i g_i(x)$

Dual objective function:

$$g(y) = \inf_{x \in \mathbb{R}^n} L(x, y) \quad \text{a concave function}$$

$$\text{dom } g = \{y \mid g(y) > -\infty\}$$

$$(D) \quad \boxed{\sup_{y \in \mathbb{R}^m} \{g(y) \mid y \in \text{dom } g, \quad y \geq 0\}}$$

**Duality Theorem** If  $(P)$  satisfies

$$\exists \bar{x} : g_i(\bar{x}) < 0 \quad \forall i = 1, \dots, m \quad (\text{Slater condition})$$

then

$$\inf(P) = \max(D)$$

---

Extensions to  $\infty$ -dimensional spaces

# CONIC DUALITY

$K =$  closed convex pointed cone,  $\text{int } K \neq \emptyset$

$$(P) \quad \inf_{x \in \mathbb{R}^n} \{c^T x \mid Ax - b \in K\} \quad A = m \times n$$

( $P$ ) is *strictly feasible* if  $\exists \bar{x} : A\bar{x} - b \in \text{int } K$

The conic-dual of ( $P$ ) is

$$(D) \quad \sup_{y \in \mathbb{R}^m} \{b^T y \mid A^T y = c, y \in K_*\}$$

$K_*$  is the *dual cone* of  $K$

$$K_* = \{y \mid y^T x \geq 0, \quad \forall x \in K\}$$

( $D$ ) is strictly feasible if  $\exists \bar{y} \in \text{int } K_* : A^T \bar{y} = c$

## Examples of dual cones

$$K = \mathbb{R}_+^n, \quad K_* = K$$

$$K = L_n = \left\{ (x_1, \dots, x_{n-1}, x_n) \mid x_n \geq \sqrt{x_1^2 + \dots + x_{n-1}^2} \right\},$$
$$K_* = K$$

$$K = S_+^n = \{A \in \mathbb{R}^{n \times n} \mid A^T = A, A \succeq 0\}, \quad K_* = K$$

$$K = \{x \mid Ax \geq 0\} \quad K_* = \{A^T y \mid y \geq 0\}$$

# SDP DUALITY

$$(P) \quad \min_{x \in \mathbb{R}^n} \{ \langle c, x \rangle \mid \mathcal{A}x - B \succeq 0 \}$$

$$\mathcal{A}x = x_1 A_1 + \cdots + x_n A_n \quad \mathcal{A} : \mathbb{R}^n \rightarrow \mathbf{S}^m$$

$$A_i, B \in \mathbf{S}^m,$$

$$\text{Dual of } (P) : \max_{Y \in \mathbf{S}^m} \{ \langle B, Y \rangle \mid \mathcal{A}^* Y = c, Y \succeq 0 \}$$

$$\langle B, Y \rangle = \text{Trace}(BY)$$

$$\mathcal{A}^* : \mathbf{S}^m \rightarrow \mathbb{R}^n, \quad \mathcal{A}^* Y = (\text{Tr}(Y A_1), \dots, \text{Tr}(Y A_n))^T$$

$\implies$

$$(D) \quad \max_{Y \in \mathbf{S}^m} \{ \text{Tr}(BY) \mid \text{Tr}(Y A_i) = c_i, i = 1, \dots, n, Y \succeq 0 \}$$

## CONIC DUALITY THEOREM

$$(P) \quad \inf\{c^T x \mid Ax - b \in K\}$$

$$(D) \quad \sup\{b^T y \mid A^T y = c, y \in K_*\}$$

1. Dual of  $(D) = (P)$
2. Weak duality:  $\inf(P) \geq \sup(D)$
3. Strong duality: If one of the  $(P)$   $(D)$  is strictly feasible and bounded, then its dual is solvable and  $\inf(P) = \sup(D)$ . If both are strictly feasible, then  $\min(P) = \max(D)$ .
4. Complementarity: If  $(P)$  or  $(D)$  is strictly feasible and bounded, then a feasible pair  $x, y$  is an optimal pair, if and only if

$$y^T (Ax - b) = 0$$

# **THE POWER OF DUALITY IN STRUCTURAL DESIGN**

- A typical Structural Design problem is

$$\begin{aligned} \text{Compl}_{\mathcal{F}}(t) &\equiv \sup_{f \in \mathcal{F}} \text{Compl}_f(t) \rightarrow \min \\ t_i &\succeq 0, \\ \underline{\rho}_i &\leq \text{Tr}(t_i) \leq \bar{\rho}_i, \quad i = 1, \dots, n, \\ \sum_{i=1}^n \text{Tr}(t_i) &\leq w \end{aligned}$$

where

- $\text{Compl}_f(t) = \sup_{v \in \mathcal{V}} [f^T v - \frac{1}{2} v^T A(t) v]$
- $A(t) = \sum_{i=1}^n \sum_{s=1}^S b_{is}^T t_i b_{is}$
- $\mathcal{V} = \{v \in \mathbb{R}^m \mid Rv \leq r\}$
- $t_i \in \mathbf{S}^d$ ,  $i = 1, \dots, n$ , are design variables
- $\mathcal{F} \in \mathbb{R}^m$  is the set of loading scenarios.

- From now on, we assume that

A)  $\mathcal{V}$  satisfies the Slater condition:

$$\exists \bar{v} : R\bar{v} < r$$

B) The ground structure does not admit rigid body motions:

$$\sum_{i,s} b_{is} b_{is}^T \succ 0$$

C)  $0 \leq \underline{\rho}_i < \bar{\rho}_i$  and

$$\sum_i \underline{\rho}_i < w$$

## Semidefinite reformulation of Structural Design problem

- The function

$$\text{Compl}_f(t) = \sup_{v: Rv \leq r} \left[ f^T v - \frac{1}{2} v^T \left[ \sum_{i,s} b_{is} t_i b_{is}^T \right] v \right]$$

of  $f \in \mathbb{R}^m$  and  $t = (t_1, \dots, t_n)$ ,  $t_i \in \mathbf{S}_+^d$  is

“**semidefinite representable**”:

$$f \in \mathbb{R}^m, t \in (\mathbf{S}_+^d)^n, \tau \geq \text{Compl}_f(t)$$

$\Leftrightarrow$

$\exists \mu \geq 0 :$

$$\left( \begin{array}{cc} 2\tau - 2r^T \mu & -f^T + \mu^T R \\ -f + R^T \mu & \underbrace{\sum_{i=1}^n \sum_{s=1}^S b_{is} t_i b_{is}^T}_{A(t)} \end{array} \right) \succeq 0;$$

$$t_i \succeq 0, i = 1, \dots, n.$$

- With “semidefinite representation” of  $\text{Compl}_f(t)$ , we can immediately pose the *multi-load* SD problem as a semidefinite program:

$$\begin{aligned}
& \tau \rightarrow \min \\
& \left( \begin{array}{cc} 2\tau - 2r^T \mu_\ell & -f_\ell^T + \mu_\ell^T R \\ -f_\ell + R^T \mu_\ell & \sum_{i=1}^n \sum_{s=1}^S b_{is} t_i b_{is}^T \end{array} \right) \succeq 0, \quad \ell = 1, \dots, k; \\
& t_i \succeq 0, \quad i = 1, \dots, n; \\
& \sum_{i=1}^n \text{Tr}(t_i) \leq w; \\
& \underline{\rho}_i \leq \text{Tr}(t_i) \leq \bar{\rho}_i, \quad i = 1, \dots, n; \\
& \mu_\ell \geq 0, \quad \ell = 1, \dots, k
\end{aligned}$$

with design variables  $t_i \in \mathbf{S}^d$ ,  $i = 1, \dots, n$ ,  
 $\mu_\ell \in \mathbb{R}^{\#\text{obst}}$ ,  $\tau \in \mathbb{R}$ , where  $\#\text{obst} = \dim r$ .

- The **robust** SD problem also can be posed as a semidefinite program *provided that there are no obstacles*:

$$\nu = \mathbb{R}^m.$$

- **Proposition**: Let

$$\mathcal{F} = \{f = Qu \mid u^T u \leq 1\} \quad [Q \in \mathbf{M}^{m,k}].$$

Then the function

$$\text{Compl}_{\mathcal{F}}(t) = \sup_v \sup_{u: u^T u \leq 1} \left[ u^T Q^T v - \frac{1}{2} v^T \left[ \sum_{i,s} b_{is} t_i b_{is}^T \right] v \right]$$

of  $t = (t_1, \dots, t_n) \in (\mathbf{S}_+^d)^n$ , is *semidefinite representable*:

$$t \in (\mathbf{S}_+^d)^n, \tau \geq \text{Compl}_{\mathcal{F}}(t)$$

$\Leftrightarrow$

$$\begin{pmatrix} 2\tau I_k & Q^T \\ Q & \sum_{i=1}^n \sum_{s=1}^S b_{is} t_i b_{is}^T \end{pmatrix} \succeq 0.$$

Consequently, in the case in question, the robust SD

problem is equivalent to the semidefinite program

$$\begin{array}{r}
 \tau \rightarrow \min \\
 \left( \begin{array}{cc} 2\tau I_k & Q^T \\ Q & \sum_{i=1}^n \sum_{s=1}^S b_{is} t_i b_{is}^T \end{array} \right) \succeq 0; \\
 t_i \succeq 0, \quad i = 1, \dots, n; \\
 \sum_{i=1}^n \text{Tr}(t_i) \leq w; \\
 \underline{\rho}_i \leq \text{Tr}(t_i) \leq \bar{\rho}_i, \quad i = 1, \dots, n;
 \end{array}$$

with design variables  $t_i \in \mathbf{S}^d$ ,  $\tau \in \mathbb{R}$ .

“Universal” semidefinite form of the SD problem

- Both multi-load and robust cases of the SD problem are covered by the following generic semidefinite program:

(Pr)

$$\begin{aligned}
 & \tau \rightarrow \min \\
 & \left( \begin{array}{cc} 2\tau I_p + \mathcal{D}_\ell z + D_\ell & [\mathcal{E}_\ell z + E_\ell]^T \\ [\mathcal{E}_\ell z + E_\ell] & \sum_{i=1}^n \sum_{s=1}^S b_{is} t_i b_{is}^T \end{array} \right) \succeq 0, \quad l = 1, \dots, K; \\
 & t_i \succeq 0, \quad i = 1, \dots, n; \\
 & \sum_{i=1}^n \text{Tr}(t_i) \leq w; \\
 & \underline{\rho}_i \leq \text{Tr}(t_i) \leq \bar{\rho}_i, \quad i = 1, \dots, n; \\
 & z \geq 0,
 \end{aligned}$$

where

- design variables are  $\{t_i \in \mathbf{S}^d\}_{i=1}^n$ ,  $z \in \mathbb{R}^N$ ,  $\tau \in \mathbb{R}$ ;
- data are  $m \times d$  matrices  $b_{is}$ , affine mappings

$$z \mapsto \mathcal{D}_\ell z + D_\ell; \mathbb{R}^N \rightarrow \mathbf{S}^d, \quad z \mapsto \mathcal{E}_\ell z + E_\ell : \mathbb{R}^N \rightarrow \mathbf{M}^{m,p},$$

$$\ell = 1, \dots, K, \text{ and reals } 0 \leq \underline{\rho}_i < \bar{\rho}_i, \quad i = 1, \dots, n,$$

$$w > 0.$$

(Pr)

$$\begin{aligned} & \tau \rightarrow \min \\ & \left( \begin{array}{cc} 2\tau I_p + \mathcal{D}_\ell z + D_\ell & [\mathcal{E}_\ell z + E_\ell]^T \\ [\mathcal{E}_\ell z + E_\ell] & \sum_{i=1}^n \sum_{s=1}^S b_{is} t_i b_{is}^T \end{array} \right) \succeq 0, \quad l = 1, \dots, K; \\ & t_i \succeq 0, \quad i = 1, \dots, n; \\ & \sum_{i=1}^n \text{Tr}(t_i) \leq w; \\ & \underline{\rho}_i \leq \text{Tr}(t_i) \leq \bar{\rho}_i, \quad i = 1, \dots, n; \\ & z \geq 0, \end{aligned}$$

- From the computational viewpoint, a disadvantage of (Pr) is its huge design dimension. E.g., when (Pr) comes from obstacle-free truss design, the dimension of the design vector is

$$\frac{M^2}{2} \pm O(M),$$

$M$  being the number of nodes. For a planar truss with  $15 \times 15$  nodal grid, this dimension is  $> 25,000$ . For a spatial truss with  $10 \times 10 \times 10$  nodal grid, the dimension is  $> 500,000$ .

- It turns out that the semidefinite dual to (Pr) admits *analytical* elimination of “most” of the variables, which in many important cases allows to reduce dramatically the dimension (and thus, the computational complexity) of the problem.

Passing from ( $Pr$ ) to its dual, we get the problem

$$\begin{aligned}
& \text{maximize} \quad -\phi \equiv -\sum_{\ell=1}^K \text{Tr}(D_\ell \alpha_\ell + 2E_\ell^T V_\ell) \\
& \quad \quad \quad -\sum_{i=1}^n \left[ \bar{\rho}_i \sigma_i^+ - \underline{\rho}_i \sigma_i^- \right] - w\gamma \\
& \left( \begin{array}{cc} \alpha_\ell & V_\ell^T \\ V_\ell & \beta_\ell \end{array} \right) \succeq 0, \quad \ell = 1, \dots, K \\
& \quad \quad \quad [\alpha_\ell \in \mathbf{S}^p, \beta_\ell \in \mathbf{S}^m, V_\ell \in \mathbf{M}^{m,p}], \\
& \quad \quad \quad \tau_i \succeq 0, \quad i = 1, \dots, n \quad [\tau_\ell \in \mathbf{S}^{D_\ell}], \\
& \quad \quad \quad \sigma_i^+, \sigma_i^- \geq 0, \quad i = 1, \dots, n \quad [\sigma_i^+, \sigma_i^- \in \mathbb{R}], \\
& \quad \quad \quad \gamma \geq 0 \quad [\gamma \in \mathbb{R}], \\
& \quad \quad \quad \eta \geq [\eta \in \mathbb{R}^N], \quad (D_{ini}) \\
& \quad \quad \quad 2\sum_{\ell=1}^K \text{Tr}(\alpha_\ell) = 1, \\
& \quad \quad \quad \sum_{\ell=1}^K [\mathcal{D}_\ell^* \alpha_\ell + 2\mathcal{E}_\ell^* V_\ell] + \eta = 0, \\
& \quad \quad \quad \sum_{\ell=1}^K \sum_{s=1}^S b_{is}^T \beta_\ell b_{is} + \tau_i \\
& \quad \quad \quad + [\sigma_i^- - \sigma_i^+ - \gamma] I_d = 0, \quad i = 1, \dots, n,
\end{aligned}$$

the design variables being  $\alpha_\ell, \beta_\ell, V_\ell, \ell = 1, \dots, K, \eta, \tau_i$ .

However,  $\eta, \tau_i$  can be immediately eliminated.

Also, it can be shown that at optimality  $\beta_\ell = V_\ell \alpha_\ell^{-1} V_\ell^T$ .

With this elimination, the dual problem becomes

$$\begin{aligned}
& \text{minimize } \phi \equiv \sum_{\ell=1}^K \text{Tr}(D_\ell \alpha_\ell + 2E_\ell^T V_\ell) \\
& \quad + \sum_{i=1}^n \left[ \bar{\rho}_i \sigma_i^+ - \underline{\rho}_i \sigma_i^- \right] + w\gamma \\
& \left( \begin{array}{cc} A(\alpha) & B_i^T(V) \\ B_i(V) & (\gamma + \sigma_i^+ - \sigma_i^-) I_d \end{array} \right) \succeq 0, \quad l = 1, \dots, N, \quad (D1) \\
& \quad \sigma_i^+ - \sigma_i^- \geq 0, \quad i = 1, \dots, n; \\
& \quad \gamma \geq 0; \\
& \quad 2 \sum_{\ell=1}^K \text{Tr}(\alpha_\ell) = 1; \\
& \quad \sum_{\ell=1}^K [\mathcal{D}_\ell^* \alpha_\ell + 2\mathcal{E}_\ell^* V_\ell] \leq 0,
\end{aligned}$$

the design variables of the problem being

$$\begin{aligned}
\alpha &= \{ \alpha_\ell \in \mathbf{S}^p \}_{\ell=1}^K, \\
V &= \{ V_i \in \mathbf{M}^{m,p} \}_{i=1}^K, \\
\sigma &= \{ \sigma_i^\pm \in \mathbb{R} \}_{i=1}^n, \quad \gamma \in \mathbb{R}.
\end{aligned}$$

- When the dual problem is solved by path-following interior point methods, it is easy to recover a (nearly) optimal design from nearly optimal central path solutions to  $(D1)$ .
- Computational advantages of  $(D1)$  as compared to  $(Pr)$  are especially significant in the case of truss design.

$$\begin{array}{c}
 \tau \rightarrow \min \\
 \left( \begin{array}{cc} 2\tau & f_\ell^T \\ f_\ell & \sum_{i=1}^n t_i b_i b_i^T \end{array} \right) \succeq 0, \ell = 1, \dots, k \\
 t_i \geq 0, i = 1, \dots, n; \\
 \sum_{i=1}^n t_i \leq w
 \end{array} \quad (Pr)$$

$$\begin{array}{c}
 -2 \sum_{\ell=1}^k f_\ell^T v_\ell + w\gamma \rightarrow \min \\
 \left( \begin{array}{cccc} \alpha_1 & & & b_1^T v_1 \\ & \alpha_2 & & b_i^T v_2 \\ & & \ddots & \vdots \\ & & & \alpha_k & b_i^T v_k \\ v_i^T v_1 & b_i^T v_2 & \cdots & b_i^T v_k & \gamma \end{array} \right) \succeq 0, i = 1, \dots, n; \\
 \sum_{\ell=1}^k \alpha_\ell = 1
 \end{array} \quad (D1)$$

- $(Pr)$  has  $n + 1 = \frac{M^2}{2} \pm O(M)$  design variables and  $k$  “large” LMI’s (of size  $(m + 1) \times (m + 1)$ ).  
 $m = 2M \pm O(\sqrt{M})$  for planar and  
 $m = 3M \pm O(M^{2/3})$  for spatial trusses.
- $(D1)$  has  $k(m + 1) = O(1)Mk \ll n = O(M^2)$  design variables and  $n$  “small” (of sizes  $(k + 1) \times (k + 1)$ ) LMI’s.
- E.g., for planar truss with  $15 \times 15$  nodal grid and 3 loads:

Setting	Design dimension	Effort of analyzing LMI’s at a point, a.o.
$(Pr)$	25,096	37,309,230
$(D1)$	1,264	267,680

- In the case of Shape design, the advantages of the dual setting also can be quite significant.

Consider, e.g., the obstacle-free planar Shape problem with rectangular cells and with simple bounds.

- The primal problem is:

$$\begin{aligned}
 & \tau \rightarrow \min \\
 & \left( \begin{array}{cc} 2\tau & -f_\ell^T \\ -f_\ell & \sum_{i=1}^n \sum_{s=1}^4 b_{is} t_i b_{is}^T \end{array} \right) \succeq 0, \ell = 1, \dots, k; \\
 & t_i \succeq 0, i = 1, \dots, n; \\
 & \sum_{i=1}^n \text{Tr}(t_i) \leq w \\
 & [\tau \in \mathbb{R}, t_i \in \mathbf{S}^3]
 \end{aligned}$$

- The dual problem is

$$-2 \sum_{\ell=1}^k f_{\ell}^T v_{\ell} + w\gamma \rightarrow \min$$

$$\left( \begin{array}{c|c|c|c} \alpha_1 & & & v_1^T b_{i1} \\ & \ddots & & \dots \\ & & \alpha_1 & v_1^T b_{iS} \\ \hline & \ddots & & \dots \\ \hline & & \alpha_k & v_k^T b_{i1} \\ & & \ddots & \dots \\ & & & \alpha_k & v_k^T b_{iS} \\ \hline b_{i1}^T v_1 \cdots b_{iS}^T v_1 & \cdots & b_{i1}^T v_k \cdots b_{iS}^T v_k & \gamma I_3 \end{array} \right) \succeq 0, \quad i = 1, \dots, n;$$

$$2 \sum_{\ell=1}^k \alpha_{\ell} = 1.$$

$$[\alpha_{\ell}, \gamma \in \mathbb{R}, v_{\ell} \in \mathbb{R}^m]$$

- E.g., for planar shape with  $14 \times 14$  cells and 3 loads:

Setting	Design dimension	Effort of analyzing LMI's at a point, a.o.
(Pr)	1,177	37,309,230
(D1)	1,264	71,608

## From dual back to primal

- If  $(D1)$  were the usual semidefinite dual of  $(Pr)$ , the problem dual to  $(D1)$  was  $(Pr)$ .

In fact,  $(D1)$  is not the semidefinite dual to  $(Pr)$  — it is obtained from this dual by eliminating part of the variables. It turns out that the semidefinite dual to  $(D1)$  is a nontrivial (and instructive) equivalent reformulation of  $(Pr)$ , namely, the problem

$$\tau \rightarrow \min$$

$$\left( \begin{array}{c|ccc} 2\tau I_p + \mathcal{D}_\ell z + D_\ell & [q_{11}^l]^T \cdots [q_{1S}^l]^T & \cdots & [q_{n1}^l]^T \cdots [q_{nS}^l]^T \\ \hline q_{11}^l & t_1 & & \\ \cdots & \ddots & & \\ q_{1S}^l & & t_1 & \\ \hline \cdots & & \ddots & \\ \hline q_{n1}^l & & & t_n \cdots \\ \cdots & & & \ddots \\ q_{nS}^l & & & t_n \end{array} \right) \succeq 0,$$

$$l = 1, \dots, K;$$

$$\underline{\rho}_i \leq \text{Tr}(t_i) \leq \bar{\rho}_i, \quad i = 1, \dots, n;$$

$$\sum_{i=1}^n \text{Tr}(t_i) \leq w;$$

$$\sum_{i=1}^n \sum_{s=1}^S b_{is} q_{is}^l = \mathcal{E}_\ell z + E_\ell, \quad l = 1, \dots, k;$$

$$z \geq 0, \quad (Pr^+)$$

the design variables being symmetric  $d \times d$  matrices  $t_i$ ,  $i = 1, \dots, n$ ,  $d \times p$  matrices  $q_{is}^l$ ,  $l = 1, \dots, K$ ,  $i = 1, \dots, n$ ,  $s = 1, \dots, S$ , real  $\tau$  and  $z \in \mathbb{R}^N$ .

- E.G., in the case of single-load obstacle-free Truss design ( $Pr^+$ ) becomes

$$\begin{aligned}
& \tau \rightarrow \min \\
& \begin{pmatrix} 2\tau & q^1 & \cdots & q^n \\ q^1 & t_1 & & \\ \vdots & & \ddots & \\ q^n & & & t_n \end{pmatrix} \succeq 0; \\
& \sum_{i=1}^n q^i b_i = f; \\
& t_i \geq 0, \quad i = 1, \dots, n; \\
& \sum_{i=1}^n t_i \leq w.
\end{aligned}$$

Since by Lemma on Schur Complement for  $t_i \geq 0$ , one has

$$\begin{pmatrix} 2\tau & q^1 & \cdots & q^n \\ q^1 & t_1 & & \\ \vdots & & \ddots & \\ q^n & & & t_n \end{pmatrix} \succeq 0 \Leftrightarrow \frac{1}{2} \sum_{i=1}^n \frac{(q^i)^2}{t_i} \leq \tau,$$

the problem can be rewritten as

$$\frac{1}{2} \sum_{i=1}^n \frac{(q^i)^2}{t_i} \rightarrow \min \mid t_i \geq 0, \sum_{i=1}^n t_i \leq w, \sum_{i=1}^n q^i b_i = f.$$

- We can analytically carry out partial minimization in  $t_i$ :

$$\sum_{i=1}^n \frac{(q^i)^2}{t_i} \rightarrow \min \mid t_i \geq 0, \sum_i t_i \leq w$$

$$\Downarrow$$

$$t_i = w \frac{|q^i|}{\sum_{\ell} |q^{\ell}|}$$

thus coming to the problem in  $q$ -variables:

$$\frac{1}{2w} \left( \sum_i |q^i| \right)^2 \rightarrow \min \mid \sum_i q^i b_i = f$$

$$\Updownarrow$$

$$\underbrace{\sum_i |q^i| \rightarrow \min \mid \sum_i q^i b_i = f}_{\text{LP!}}$$

# OPTIMAL CONTROL EXAMPLE

$t_0 =$  initial time,  $t_1 > t_0$  final time

$x(t) \in \mathbb{R}^n$  state vector at time  $t$

$u(t) \in \mathbb{R}$  control function,  $u \in L_2[t_0, t_1]$

$A(t)$   $n \times n$  matrix,  $b \in \mathbb{R}^n$ ,  $c \in \mathbb{R}^n$

## PRIMAL PROBLEM

$$\min \left\{ J = \frac{1}{2} \int_{t_0}^{t_1} u^2(t) dt \right\}$$

subject to

$$(1) \quad X(t_1) \geq c$$

$$(2) \quad \dot{x}(t) = A(t)x(t) + b(t)u(t).$$

Let  $\{X^1(t), \dots, X^n(t)\}$  be a set of linearly independent solutions of the homog. system

$$\dot{x}(t) = A(t)x(t)$$

set

$$X(t) = [X^1(t), \dots, X^n(t)] \quad n \times n \text{ fundamental matrix}$$

Then, solution of (2) is given by

$$x(t) = X(t)X^{-1}(t_0)x(t_0) + X(t) \int_{t_0}^t X^{-1}(s)b(s)ds.$$

Define

$$\Phi(t_1, t_2) = X(t_1)X^{-1}(t_2).$$

Then, primal problem can be written

$$\min \frac{1}{2} \int_{t_0}^{t_1} u^2(t)dt$$

subject to

$$\underbrace{\int_{t_0}^{t_1} \phi(t_1, t)b(t)u(t)dt}_{Ku} \geq \underbrace{c - \Phi(t_1, t_0)x_0}_d$$

## COMPUTING THE LAGRANGIAN DUAL

Dual objective function

$$\begin{aligned} g(y) &= \min_{u \in L_2[t_0, t_1]} \int_{t_0}^{t_1} \frac{1}{2} u^2(t)dt + y^T (d - Ku(t))dt \\ &= y^T d + \min_{u(\cdot)} \int_{t_0}^{t_1} \left[ \frac{1}{2} u^2(t) - y^T \Phi(t_1, t)b(t)u(t) \right] dt. \end{aligned}$$

Pointwise minimization inside integral

$$\bar{u}(t) = y^T \Phi(t_1, t)b(t)$$

hence,

$$g(y) = y^T d - \underbrace{\frac{1}{2} \int_{t_0}^t [y^T \Phi(t_1, t) b(t) b(t)^T \Phi(t_1, t)^T y] dt}_{y^T Q y}$$

where  $Q$  is the N.S.D.  $n \times n$  matrix

$$Q = -\frac{1}{2} \int_{t_0}^{t_1} \Phi(t_1, t) b(t) b(t)^T \Phi(t_1, t)^T dt .$$

The Dual Problem is

$$\boxed{\max_{y \geq 0} \{y^T d + y^T Q y\}}$$

A quadratic programming (finite dimensional) with  $Q$  as above and  $d = c - \phi(t_1, t_0)x_0$ .

## EXAMPLE: A NONCONVEX QUADRATIC PROBLEM

$$(A) \quad \min_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} z^T Q z + c^T z : z^T z < 1 \right\}$$

$Q = n \times n$  symmetric indefinite

eigenvalues  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$  ( $\lambda_1 < 0$ )

orthonormal e-vectors  $u_1 \leq u_2 \leq \dots \leq u_n$

$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ ;  $P = (u_1, \dots, u_n)$

change of variables:  $x = Pz$  and using  $Q = P^T \Lambda P$ ,  
 $P^T P = I$ ,

problem (A) converted to:

$$(P) \quad \min_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} \sum \lambda_i x_i^2 + \bar{c}^T x : x^T x \leq 1 \right\}, \quad \bar{c} = Pc.$$

Dual of (P) via lagrangian duality:

$$(D) \quad \max_{\mu \geq 0} \left\{ -\frac{1}{2} \left( \sum \frac{\bar{c}_i^2}{\lambda_i + \mu} + \mu \right) : \lambda_i + \mu \geq 0, \forall i \right\}$$

a concave program

computing the *dual of the dual*.

Rewriting (D) as

$$(\hat{D}) \quad \left\{ \begin{array}{l} \max_{y \in \mathbb{R}^n, \mu \in \mathbb{R}} \left\{ -\frac{1}{2} \left( \sum \frac{\bar{c}_i^2}{y_i} + \mu \right) \right\} \\ \text{subject to} \\ -y_i + \lambda_i + \mu = 0 \quad \leftarrow \text{multiplier } \frac{u_i}{2} \\ y_i \geq 0, \mu \geq 0 \end{array} \right.$$

The (lagrangian) dual of  $(\hat{D})$  is

$$(DD) \quad \boxed{\begin{array}{l} \min \left\{ \frac{1}{2} \sum \lambda_i u_i - \sum |\bar{c}_i| \sqrt{u_i} \right\} \\ \text{subject to} \\ \sum u_i \leq 1, u_i \geq 0 \end{array}}$$

$(DD)$  is, of course, a convex program.

What is the relation between  $(P)$  and  $(DD)$ ?

$$(P) \quad \boxed{\begin{array}{l} \min \sum \left( \frac{1}{2} \lambda_i x_i^2 + \bar{c}_i^T x_i \right) \\ \sum x_i^2 \leq 1 \end{array}}$$

**Theorem 1** *The nonconvex program (P) is equivalent to the convex program (DD).*

$$\{u_i^* : i = 1, \dots, n\} \text{ solves } (DD) \text{ iff}$$
$$\left\{ x_i^* = -(\text{sign } \bar{c}_i) \sqrt{u_i^*}, i = 1, \dots, n \right\} \text{ solves } (P).$$

**Proof:**

$$\min(P) \leq \max(D) = \min(DD)$$

but  $x_i^* = -(\text{sign } \bar{c}_i) \sqrt{u_i^*}$  is feasible to (P) and objective function of (P) evaluated at  $x^*$  is equal to  $\min(DD)$ .

## EXAMPLE: STATISTICAL INFORMATION THEORY

$X$  – random variable (nondegenerate)

$B$  – support of  $X$

$f_X$  – density of  $X$

$D$  – class of density functions with support  $B$   
which are absolutely contin. w.r.t.  
a nonnegative measure  $dt$

$A_i(t)$  – summable functions

$$(P) \quad \inf_{f \in D} \left\{ I(f, f_X) = \int_B f(t) \log \frac{f(t)}{f_X(t)} dt \right\}$$

s.t.

$$\int f(t) A_i(t) dt = \theta_i, \quad i = 1, \dots, m$$

$\infty$ -dimensional problem\*

Fundamental Problem in *Statistical Information Theory Application* in Traffic Engineering, Accounting, Marketing, Signal Processing, Statistics ...

---

\*  $D \in L^p(\Omega, \mathcal{F}, P)$  linear space of measurable real-valued function  $f : \Omega \rightarrow \mathbb{R}$ ,  $\|f\|_p < \infty$

$\|f\|_p = \left( \int_{\Omega} |f(\omega)|^p dP(\omega) \right)^{1/p}$ ,  $(1 < p < \infty)$

## The Dual Problem

$$(D) \quad \sup_{\pi \in \mathbb{R}^m} \left\{ \sum \theta_i \pi_i - \log \int_B f_X(t) e^{\sum A_i(t) \pi_i} dt \right\}$$

**Unconstrained!      Finite Dimensional!**

## Duality Theorem

- (i)  $\inf(P) = \sup(D)$
- (ii)  $\inf(P) = \min(P)$
- (iii)  $\sup(D) = \max(D)$  iff  $(P)$  is superconsistent
- (iv)  $\sup(D) < \infty \Leftrightarrow (P)$  is feasible  
 $\Rightarrow$
- (v) If  $\pi^*$  solves  $(D)$ , then

$$f^*(t) = \frac{f_X(t) e^{\sum A_i(t) \pi_i^*}}{\int_B f_X(t) e^{\sum A_i(t) \pi_i^*} dt}$$

solves  $(P)$

\* If  $f_X(t)$  is not a density function, but just a positive summable function, then dual problem:

$$\sup_{\pi \in \mathbb{R}^m} \left\{ \sum \theta_i \pi_i - \int_B f_X(t) e^{\sum A_i(t) \pi_i - 1} dt \right\}$$

## Examples

$$(1) \inf \left\{ \int_a^b f(t) \log f(t) dt : \int f(t) dt = 1 \right\}$$

$$\text{sol. } f^*(t) = \begin{cases} 1/(b-a) & a \leq t \leq b \\ 0 & \text{otherwise} \end{cases} \quad \text{UNIFORM}$$

$$(2) \inf_{f \in D} \left\{ \int_{-\infty}^{\infty} f(t) \log f(t) dt : \int_{-\infty}^{\infty} t^2 f(t) dt = \sigma^2 \right\}$$

$$\text{sol. } f^*(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} \quad \text{NORMAL}$$

$$(3) \inf_{f \in D} \int_0^{\infty} f(t) \log \frac{f(t)}{t^{K-1}} dt$$

$$\int t f(t) dt = K/\lambda$$

$$\text{sol. } f^*(t) = \frac{t^{K-1} e^{-\lambda t}}{\Gamma(K)} \quad \text{GAMMA}$$

## LIST OF P.D.F.'s DERIVED FROM THE ABOVE DUALITY

### DISCRETE R.V.

uniform

geometric

binomial

Poisson

log Series

Truncated Geometric

⋮

### CONTINUOUS R.V.

Normal

Laplace

Generalized Cauchy

exponential

gamma

beta

log normal

⋮

### MULTIVARIATE R.V.

Multi Normal

Multi log Normal

Dirichlet

Multivariate Beta of 2nd kind

(Generalized) multivariate logistic

⋮