ANZIAM J. 42 (E) ppE27–E68, 2001

Duality and randomization in nonlinear programming

Jean B. Lasserre*

(Received 4 May 2000, revised 2 Nov 2000)

Abstract

We consider the NLP optimization problem

$$\mathrm{P} \mapsto \inf_x \{f_0(x) \mid f_i(x) \le b_i,\, i = 1, \ldots, m\}$$

and discuss the duality gap between P and

$$\mathrm{D} \mapsto \sup_{\lambda \ge 0} \inf_x \left\{ f_0(x) + \sum_{i=1}^m \lambda_i [f_i(x) - b_i] \right\}.$$

*LAAS-CNRS, 7 Avenue du Colonel Roche, 31077 Toulouse Cédex, France. mailto:lasserre@laas.fr

[0]See http://anziamj.austms.org.au/V42/E017 for this article and ancillary services,

The convex problem D is in fact the dual of a "relaxed" version of P via "randomization" which permits to give a simple interpretation for the presence or absence of a duality gap in the general case. Several particular cases are also discussed and the case of homogeneous functions is given special attention.

Contents

1 Introduction

Consider the standard Nonlinear Programming Problem (NLP)

$$\mathrm{P} \mapsto \inf_x \{f_0(x) \mid f_i(x) \leq b_i,\ i = 1, 2, \ldots, m\}$$

where $f_i : R^n \to R$, $i = 0, 1, \ldots, m$, are all continuous functions. It is well-known that

$$\sup \mathrm{D} \leq \inf \mathrm{P} = \inf \mathrm{P}^* \tag{1}$$

where

$$\mathrm{P}^* \mapsto \inf_x \sup_{\lambda \geq 0} \left\{ f_0(x) + \sum_{i=1}^m \lambda_i [f_i(x) - b_i] \right\} \tag{2}$$

and

$$\mathrm{D} \mapsto \sup_{\lambda \geq 0} \inf_x \{f_0(x) + \sum_{i=1}^m \lambda_i [f_i(x) - b_i]\} \tag{3}$$

When permitted, the interchange of "inf" and "sup" operators (that is, the equivalence of (2) and (3)) characterizes the absence of a duality gap and optimal solutions (if any) are saddle points of the Lagrangian (see e.g. [8], [10]). Except for the convex case, i.e. when all the f_i are convex (and under a constraint qualification), there is no systematic approach to decide whether or not there is a duality gap. Hence, identifying instances for which the absence of a duality gap is guaranteed is a challenging problem. This is not surprising for P and D are very different in nature. Solving P is in general a difficult *global* optimization problem whereas D is a *convex* problem, easier to solve in principle. In particular cases, some authors have obtained interesting results (see e.g. quadratic and pure quadratic problems in [2], [9], [13] and the references therein).

The goal of this paper is to provide some insights on this duality gap issue. We show that both D and P* are in fact derived from "relaxed" versions of P. D is obtained as the dual of a linear "randomized" version PP of P, the analogue of the familiar "relaxed control" procedure in control (see e.g. [3], [11], [15]), that yields the concept of "generalized solutions" (or "Young measures") to the original problem, whereas P* is obtained by embedding P into a larger (but equivalent) problem. In turn, PP is also the dual of D and under weak conditions, there is no duality gap between D and PP. In particular, it is shown that to every optimal solution λ of the dual problem D, corresponds an optimal "randomized" solution μ to the primal problem PP (μ may be interpreted as a "generalized" solution to P). In this approach, the absence of a duality gap between P and D in the convex case under Slater's constraint qualification appears as an immediate consequence of Jensen's inequality.

In a sense, PP is a "regularized" (or "convex") version of P, pretty much like f^{**} is a regularized version of a nonconvex function f with Legendre-Fenchel transformed f^*. This "randomization" (or "convexification") procedure has been successfully applied to some problems in Control Theory, Economics (see e.g. [11], [15]). Here, we emphasize the *duality* point of view. More precisely, it is shown that under weak conditions, there is no duality gap between PP and D. In addition, when both D and PP are solvable, there is a probability distribution μ on the the set $X(\lambda)$ of global minima of the Lagrangian $L(., \lambda)$ (with λ any optimal solution of D), such that $\int f_i d\mu \leq b_i$ $i = 1, 2, \ldots, m$ and $\int f_0 d\mu = \max \mathrm{D}$. In other words, μ is an optimal solution of a "relaxed" version of P, i.e., when the constraints and the objective are "averaged out" with respect to a probability measure. In general, this "averaging procedure" yields generalized solutions with a strictly better cost than usual solutions. An even finer characterization of optimal generalized solutions is obtained via Caratheodory Theorem when the set $X(\lambda)$ is compact.

As a Corollary, there is no duality gap between D and P if and only if all the global minimizers of P belong to $X(\lambda)$ and checking the absence of a duality gap reduces to check whether there is a feasible solution of P in $X(\lambda)$ such that a complementary condition holds. For instance, when the f_i's are all quadratic, first solve D, obtain λ and then check whether there is a feasible solution x^* of P that solves a quadratic system of equations/inequations in a space of smaller dimension (cf. Corollary 4 of Section 2). This is in contrast to the *trial and error* method which consists in first finding a Karush-Kuhn-Tucker point (x^*, λ^*) of P and checking afterwards whether x^* is also a global minimizer of the Lagrangian $L(., \lambda^*)$.

Some other particular cases are also investigated. For instance, if $X(\lambda)$ is a singleton then solving the nonconvex problem P reduces to solving the convex problem D. The general quadratic case is also investigated. The dual D is an LMI problem whose dual D* is a well-known relaxation of P (see Boyd and Vandenberghe [6]). The optimal values of PP and D* coincide and optimal solutions of PP with finite support provide a natural interpretation of the optimal solutions of D* in terms of "randomization".

Finally, the case of homogeneous functions is considered. The dual D takes a particular form and the solvability of PP is obtained under weak conditions. In addition, when $X(\lambda)$ is a one-dimensional cone, then $\min \mathrm{P} = \max \mathrm{D}$ and solving P reduces to solving a convex problem (an LMI problem in case all the f_i's are quadratic forms).

2 On the duality gap

Consider the following optimization problem in $X := R^n$:

$$P \mapsto f^* := \inf \{f_0(x) \mid f_i(x) \le b_i, \quad i = 1, 2, \ldots, m\} \tag{4}$$

where $f_i : X \to R$, $i = 0, 1, \ldots, m$ are all continuous functions. Denote by S the feasible set of P, i.e.,

$$S := \{x \in X \mid f_i(x) \le b_i \quad i = 1, 2, \ldots, m\} . \tag{5}$$

Assumption A: *$\forall i = 1, 2, \ldots, m$, there exists $x_i \in S$ such that $f_i(x_i) < b_i$.*

Note that Assumption A is just Slater's constraint qualification whenever the functions f_i are convex $i = 1, 2, \ldots, m$.

2.1 The dual D

Consider now the following "relaxation" PP of P:

$$\mathrm{PP} \begin{cases} \displaystyle\inf_{\mu \in \mathcal{M}(X), \mu \ge 0} \int f \, d\mu & \\ \displaystyle\int [f_i(x) - b_i] \, \mu(dx) \le 0, & i = 1, 2, \ldots, m, \\ \mu(X) = 1 & \end{cases} \tag{6}$$

where $\mathcal{M}(X)$ is the Banach space of signed Borel measures on $\mathcal{B}$ (the Borel σ-field on X), equipped with the total variation norm.

Obviously, PP is a *relaxation* of P via *randomization* since for every admissible point $x \in S$, the Dirac measure $\mu := \delta_x$ concentrated at the point x, is admissible in PP with corresponding value $f(x)$ and therefore, inf PP $\leq f^*$. In fact, this relaxation is a "convexification" of P, the analogue of the "relaxed controls" procedure in control (see e.g. [3], [11], [15] and the references therein). The advantage of doing this is that the linear problem PP admits a natural "dual" linear problem D introduced below:

$$\mathrm{D} \mapsto \sup_{\lambda \geq 0, \gamma} \left\{ \gamma \,\middle|\, f_0(x) + \sum_{i=1}^{m} \lambda_i [f_i(x) - b_i] \geq \gamma, \quad \forall x \in X \right\}, \tag{7}$$

or equivalently

$$\mathrm{D} \mapsto \sup_{\lambda \geq 0} \inf_{x} \left\{ f_0(x) + \sum_{i=1}^{m} \lambda_i [f_i(x) - b_i] \right\} \tag{8}$$

which the usual "dual" considered in NLP. In fact D and PP are dual of each other.

More precisely, with $w := 1 + \max_{i=1,\ldots,m} |f_i|$, and following Anderson and Nash [1], let $(\mathcal{X}, \mathcal{Y})$ and $(\mathcal{Z}, \mathcal{W})$ be two dual pairs of vector spaces where

$$\mathcal{X} := \left\{ \mu \in \mathcal{M}(X) \,\middle|\, \int w \, d|\mu| < \infty, \; i = 0, 1, \ldots, m \right\},$$

(where $|\mu|$ denotes the total variation of μ), and

$$\mathcal{Y} := \left\{ f : R^n \to R \,\middle|\, \sup_{x \in R^n} \frac{|f(x)|}{w(x)} < \infty \right\},$$

whereas, $\mathcal{W} = \mathcal{Z} := R^m$. The pair $(\mathcal{X}, \mathcal{Y})$ is in duality via

$$\langle \mu, h \rangle = \int h \, d\mu \,, \quad \mu \in \mathcal{X},\ h \in \mathcal{Y} \,.$$

Introducing the linear maps

$$\begin{array}{rcl} T : \mathcal{X} & \longrightarrow & \mathcal{Z} \\ \mathcal{Y} & \longleftarrow & \mathcal{W} : \widehat{T} \end{array}$$

where

$$\mu \mapsto T\mu \ := \begin{bmatrix} \int (f_1 - b_1) d\mu \\ \vdots \\ \int (f_m - b_m) d\mu \end{bmatrix} \quad \mu \in \mathcal{X},$$

and

$$\lambda \mapsto \widehat{T}\lambda \ := \sum_{i=1}^{m} \lambda_i [f_i - b_i] \quad \lambda \in \mathcal{Z},$$

the linear program PP reads

$$\mathrm{PP} \mapsto \inf_{\mu \in \mathcal{X}, \mu \geq 0} \{ \langle f_0, \mu \rangle \mid T\mu \leq 0;\ \langle 1, \mu \rangle = 1 \},$$

whereas the linear program D reads

$$\mathrm{D} \mapsto \sup_{(\lambda, \gamma) \in \mathcal{Z} \times R, \lambda \geq 0} \{ \gamma \mid \gamma - \widehat{T}(\lambda) \leq f_0 \},$$

which is just (7). It is straightforward to check that $R + \widehat{T}(\mathcal{Z}) \subseteq \mathcal{Y}$ so that the linear map $T_1 : \mathcal{X} \to \mathcal{Z} \times R$ defined by $\mu \mapsto T_1\mu := (T\mu, \langle 1, \mu\rangle)$ is continuous for the respective weak topologies $\sigma(\mathcal{X}, \mathcal{Y})$ and $\sigma(\mathcal{Z} \times R, \mathcal{W} \times R)$ (see e.g. [7]).

On the other hand, consider now the other following equivalent version P^* of P,

$$\mathrm{P}^* \mapsto \inf_{u \geq 0, x} \{u f_0(x) \mid u[f_i(x) - b_i] \leq 0,\ i = 1, 2, \dots, m;\ u = 1\}.$$

Equivalently,

$$\mathrm{P}^* \mapsto \inf_x \inf_{u \geq 0} \{u f_0(x) \mid u[f_i(x) - b_i] \leq 0,\ i = 1, 2, \dots, m;\ u = 1\},$$

and noting from LP duality that the dual of the LP problem

$$\inf_{u \geq 0} \{u f_0(x) \mid u[f_i(x) - b_i] \leq 0,\ i = 1, 2, \dots, m;\ u = 1\},$$

is just

$$\sup_{\lambda \geq 0, \gamma} \left\{ \gamma \;\middle|\; f_0(x) + \sum_{i=1}^m \lambda_i [f_i(x) - b_i] \geq \gamma \right\},$$

one obtains the the following equivalent formulation

$$\mathrm{P}^* \mapsto \inf_x \sup_{\lambda \geq 0} \left\{ f_0(x) + \sum_{i=1}^m \lambda_i [f_i(x) - b_i] \right\},$$

which is (2). We now introduce the following condition:

Condition B: *There exist nonnegative coefficients* λ_i, $i = 1, 2, \ldots, m$, *such that*

$$\inf_{x \in X} \left\{ f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right\} > -\infty \,. \tag{9}$$

We first consider the solvability of D and the absence of a duality gap between D and PP, under Assumption A and condition B.

Theorem 1 *Let Assumption A and Condition B hold, then*

$$\sup D = \max D = \inf PP \,. \tag{10}$$

The proof is relegated to Appendix A. Of course, the main issue of interest is to determine when $\inf \mathrm{PP} = \inf \mathrm{P} = f^*$ holds or to gain some insight into the presence of a duality gap, i.e. when $\inf \mathrm{PP} < \inf \mathrm{P}$.

For convex problems recall that under Slater's condition, $\inf \mathrm{PP} = \inf \mathrm{P}$ (see e.g. [10]). However, we provide below a simple proof of this result that simply uses Jensen's inequality.

2.2 The convex case

In this section we assume that f_i are convex $i = 0, 1, \ldots, m$. Assumption A becomes Slater's constraint qualification under which $\max\ \mathrm{D} = \inf \mathrm{P}$, a well-known result.

(i) The fact that $\inf \mathrm{PP} = \inf \mathrm{P}$ is in fact an immediate consequence of Jensen's inequality. In case X would be a compact convex subset of R^n, then $\int x\, d\mu =: E_\mu(x) \in X$ is well defined and Jensen's inequality applies, that is,

$$\int f\, d\mu \geq f(E_\mu(x))\,,$$

for every continuous convex function f (see e.g. Perlman [12]). Therefore, to every admissible solution μ of PP, one may associate an admissible solution $E_\mu(x)$ of P with a lower cost.

When $X := R^n$, the proof also uses a compactness argument. For the problems PP_i defined in the proof of Theorem 1, μ_i defines a probability measure on K_i with corresponding expectation operator E_{μ_i}. As K_i is compact and can be chosen convex, the random vector $x \in K_i$ is μ-integrable, and since K_i is convex, Jensen's inequality is valid and yields

$$\int f_k\, d\mu_i \geq f_k(E_{\mu_i}(x))\,, \qquad k = 0, 1, \ldots, m\,.$$

In particular,

$$0 \geq \int (f_k(x) - b_k)\, \mu_i(dx) \geq f_k(E_{\mu_i}(x)) - b_k\,, \qquad k = 1, 2, \ldots, m\,,$$

so that the point $y := E_{\mu_i}(x)$ is feasible in P with value $f_0(y) \leq \gamma_i$. Therefore, $\inf \mathrm{P} = \inf \mathrm{PP}_i$ for all i. As $\inf \mathrm{PP}_i \to \inf \mathrm{PP} = \max\ \mathrm{D}$ we obtain $\inf \mathrm{P} = \max \mathrm{D}$.

(ii) The fact that there is no duality gap between PP and D also follows from Slater's constraint qualification which is an "interior point" condition for absence of a duality gap (for instance, invoke Theorem 3.13, p. 55 in Anderson and Nash [1]).

(iii) Assume that x^* is an optimal solution of P. With λ^* an optimal solution of D we obtain,

$$f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* [f_i(x^*) - b_i] \geq \max \mathrm{D} = f_0(x^*)\,,$$

which yields

$$\sum_{i=1}^{m} \lambda_i^* [f_i(x^*) - b_i] = 0\,,$$

since $f_i(x^*) \leq b_i$ for all i. Hence, x^* is a saddle point of the Lagrangian $L(., \lambda^*)$. In addition, if the f_i are all differentiable,

$$\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) = 0\,, \tag{11}$$

so that (x^*, λ^*) is a Karush-Kuhn-Tucker (KKT) point of P.

In the general (non convex) case, $\sup \mathrm{D} \leq \inf \mathrm{P}$. The case where $\max \mathrm{D} = \min \mathrm{PP}$ is of particular interest for we are then able to provide some further insight on a possible duality gap between P and D. It first requires the solvability of PP.

Theorem 2 *Let Assumption A and Condition B hold. In addition, assume that*

(a) $f_m(x)$ *is inf-compact, that is, the levels sets* $\{x \mid f_m(x) \le r\}$ *are compact for every* $r \in R$*;*

(b) $|f_i(x)|/(M + f_m(x)) \to 0$ *as* $|x| \to \infty$*, for all* $i = 0, 1, \ldots, m-1$*, and some* $M > 0$*.*

Then PP is solvable and $\max D = \min PP$*.*

The proof is relegated to Appendix B. The index m in Conditions (a)–(b) in Theorem 2 is arbitrary and can be any in the set $\{0, \ldots, m\}$. These conditions (a)–(b) are particular cases of the *property (γ)* stated in Balder [3] in a more general context, that is, (a)–(b) imply that f_i, $i = 0, 1, \ldots, m-1$ have the property (γ) with respect to f_m.

Observe that the conditions (a)–(b) rule out the cases where all the f_i are of same nature (e.g. linear, quadratic, ...). In this case, ad hoc conditions must be found to ensure solvability of PP.

2.3 On the duality gap

We now provide, proved in Appendix C, several characterizations of the optimal solutions of the relaxed version PP of P, when they exist. This will help understand the presence of a duality gap between P and D.

Theorem 3 *Let Assumption A hold and Condition B be true, then*

(a) *for every solution μ^* of PP and every solution λ^* of D, one has:*

$$f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) = \min_{y \in X} \left\{ f_0 + \sum_{i=1}^m \lambda_i^* f_i \right\} \quad \mu^* \ a.e. \tag{12}$$

and

$$\lambda_i^* \left[\int f_i \, d\mu^* - b_i \right] = 0 \,, \quad i = 1, 2, \ldots, m \,. \tag{13}$$

In addition, if the f_i are all differentiable, then

$$\nabla f_0(x) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x) = 0 \,, \quad \mu^* a.e. \tag{14}$$

Hence, μ^* is concentrated on the global minimizers of the Lagrangian $L(., \lambda^*)$.

(b) *If the set of global minimizers $X(\lambda^*)$ of the Lagrangian $L(., \lambda^*)$ is compact, then every optimal solution μ^* of PP is a convex combination of at most $s+1$ Dirac measures δ_{x_k}, $x_k \in X(\lambda^*), k = 1, \ldots, s+1$, where s is the number of active constraints in PP at μ^*, and*

$$\nabla f_0(x_k) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x_k) = 0 \quad k = 1, \ldots, s+1.$$

When PP is solvable, Theorem 3 provides some insight into the absence of a duality gap when P is solvable, i.e. when $\max \mathrm{D} < \min \mathrm{P} = f_0(x^*)$ where x^* is a global minimizer of P. In this case, there is a probability distribution μ^* on the set of global minimizers of the Lagrangian $L(., \lambda^*)$ such that "averaging out" with respect to μ^* yields a better result, i.e., the average values $\int f_i d\mu^*$ satisfy

$$\int f_0 d\mu^* < f(x^*) \, ; \quad \int f_i d\mu^* \leq b_i \, , \quad i = 1, 2, \ldots, m \, .$$

Note in parenthesis that if $\max \mathrm{D} < \min \mathrm{P}$ and (x^*, λ) is a KKT point of P with x^* a global minimizer, then x^* is *not* a global minimizer of the Lagrangian $L(., \lambda)$. Indeed, otherwise with $\gamma := f_0(x^*)$, the point (λ, γ) would be feasible for D, so that $\max \mathrm{D} = \min \mathrm{P}$, a contradiction. One retrieves the well-known result that if x^* and λ^* are optimal solutions of P and D respectively, then there is no duality gap if and only if (x^*, λ^*) is a saddle point of the Lagrangian.

The following result characterizes the absence of a duality gap.

Corollary 4 *Let Assumption A and Condition B hold. Let (λ^*, γ^*) be an optimal solution of D. Then $\max D = \min P$ if and only if there exists $x^* \in X(\lambda^*) \cap S$ such that*

$$\lambda_i^* [f_i(x^*) - b_i] = 0 \, , \quad i = 1, 2, \ldots, m \, . \tag{15}$$

Proof:

The if part Let γ^* be the optimal value of D. As $x^* \in X(\lambda^*)$, x^* is a global minimizer of $L(., \lambda^*)$ so that

$$f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* [f_i(x^*) - b_i] = \gamma^* = \max \ \mathrm{D} = f_0(x^*)\,,$$

and since x^* is feasible, we have $\max \mathrm{D} = \min \mathrm{P}$, the desired result.

The only if part Let x^* be a global minimizer of P. If $\max \mathrm{D} = \min \mathrm{P}$, then we also have $\min \mathrm{PP} = \min \mathrm{P}$ and the Dirac measure δ_{x^*} is solution of PP. Therefore, from Theorem 3(a) it follows that $\delta_{x^*}(X(\lambda^*)) = 1$, that is, $x^* \in X(\lambda^*)$ (hence $x^* \in X(\lambda^*) \cap S$) and (15) follows from (13).

♠

Corollary 4 is useful in global optimization. A "trial and error" method to find a global minimizer of P first computes a local minimizer x^* of P with associated KKT multiplier λ^*, and then tries to check whether x^* is a global minimizer, that is, if it is also a global minimizer of $L(., \lambda^*)$. On the contrary, Corollary 4 suggests to *first* solve D (a convex problem) and *then* check whether there exists some $x^* \in X(\lambda^*) \cap S$ that satisfies (15). In some cases, the latter problem also reduces to a convex problem (see next Section on the general quadratic case).

If we now denote by Δ the set of optimal solutions λ of D, we obtain as a corollary:

Corollary 5 *Let Assumption A and condition B hold. If PP is solvable and*

$$\bigcap_{\lambda \in \Delta} X(\lambda) = \{x^*\} \tag{16}$$

then x^* is a global minimizer of P and $\max D = \min P$.

Proof: From Theorem 3, let μ^* be an optimal solution of PP. Since, μ^* is concentrated on $X(\lambda)$ for every $\lambda \in \Delta$, from (16) we must have $\mu^* = \delta_{x^*}$, with δ_{x^*} the Dirac measure at x^*. But this implies that x^* is feasible for P and $\min \mathrm{PP} = f(x^*) \leq \inf \mathrm{P}$, and thus, $\max\ \mathrm{D} = \min \mathrm{P}$. ♠

2.4 The (general) quadratic case

Consider the case where the f_i's are quadratic functions, i.e.,

$$x \mapsto f_i(x) \ := \ x'Q_i x + 2c_i' x\,, \quad i = 0, 1, \ldots, m\,,$$

where c' denotes the transpose of a vector c. Denote also by $\langle A, B \rangle$ the usual scalar product trace(AB) for real-valued symmetric matrices.

Then D has the form

$$\left\{\begin{array}{rcl} & \max\limits_{\lambda\geq 0,\gamma} \gamma & \\ Q_0+\sum\limits_{i=1}^m \lambda_i Q_i & \succeq & 0 \\ x'\left(Q_0+\sum\limits_{i=1}^m \lambda_i Q_i\right)x+2\left(c_0+\sum\limits_{i=1}^m \lambda_i c_i\right)' x & \geq & \gamma+\sum\limits_{i=1}^m \lambda_i b_i \end{array}\right.$$

(the first constraint is necessary for $\sup \mathrm{D} > -\infty$) or, equivalently,

$$\max_{\lambda\geq 0,\gamma}\left\{\gamma \;\middle|\; \begin{bmatrix} Q_0+\sum_{i=1}^m \lambda_i Q_i & c_0+\sum_{i=1}^m \lambda_i c_i \\ (c_0+\sum_{i=1}^m \lambda_i c_i)' & -\gamma-\sum_{i=1}^m \lambda_i b_i \end{bmatrix} \succeq 0\right\}$$

an LMI problem whose dual D^*

$$\mathrm{D}^* \mapsto \left\{\begin{array}{rrcl} \min & \langle Q_0, Y\rangle + 2c_0'x & & \\ \text{s.t.} & \langle Q_i, Y\rangle + 2c_i'x & \leq & b_i\,, \quad i=1,2,\ldots,m \\ & \begin{bmatrix} Y & x \\ x' & 1 \end{bmatrix} & \succeq & 0 \end{array}\right.$$

is a well-known relaxation of P (see e.g. Boyd and Vandenberghe [6] and the references therein). Under Assumption A and Condition B, $\max \mathrm{D} = \inf \mathrm{PP}$. Therefore, if $\max \mathrm{D} = \min \mathrm{D}^*$, PP and D^* have same value. In fact, the equivalence of D^* and PP follows from the fact that to each positive semidefinite matrix $\begin{bmatrix} 1 & x' \\ x & Y \end{bmatrix}$, one may associate a probability measure μ with first moment vector x and second-order moment matrix Y, that is,

$$x_i = \int z_i \, d\mu \,, \quad \forall i = 1, \ldots, n\,; \quad Y(i,j) = \int z_i z_j \, d\mu \,, \quad \forall i,j = 1, \ldots, n\,,$$

for some probability μ. Conversely, for every probability measure μ, its vector x of first-order moments and its matrix Y of second-order moments must satisfy $\begin{bmatrix} 1 & x' \\ x & Y \end{bmatrix} \succeq 0$. Therefore, the above formulation D^* is just an equivalent formulation of PP since for every $i = 0, \ldots, m$, $\int f_i \, d\mu$ only involves the first and second-order moments of μ.

Let λ^* be an optimal solution of D and let $\{u_1, \ldots, u_p\}$ be an orthonormal basis of $N := \mathrm{Ker}(Q_0 + \sum_{i=1}^m \lambda_i^* Q_i)$. Then, $x \in X(\lambda^*)$ if and only if $x = z_0 + \sum_{j=1}^p y_j u_j$, where z_0 solves

$$\left[Q_0 + \sum_{i=1}^m \lambda_i^* Q_i \right] z_0 = - \left[c_0 + \sum_{i=1}^m \lambda_i^* c_i \right] . \tag{17}$$

Let z_0 be a particular solution of (17). Let H_i be the real-valued (p,p)-symmetric matrix defined by $H_i(jk) := u_j' Q_i u_k$, $1 \le j,k \le p$, and let d_i be the p-vector $d_i(j) := (c_i + 2Q_i z_0)' u_j$, $1 \le j \le p$. Then, invoking Corollary 4, $\max \mathrm{D} = \min \mathrm{P}$ if and only if the following systems of quadratic equations/inequations in R^p:

$$y' H_i y + d_i' y \;=\; b_i - z_0' Q_i z_0 - c_i' z_0 \quad (\lambda_i^* > 0) \tag{18}$$
$$y' H_i y + d_i' y \;\le\; b_i - z_0' Q_i z_0 - c_i' z_0 \quad (\lambda_i^* = 0) \tag{19}$$

has a solution.

Let $m = 2$, $c_i = 0$, $i = 0, 1, 2$ (the pure quadratic case with 2 constraints) so that $z_0 = 0$, and let λ^* be such that $\lambda_i^* > 0$ for all $i = 1, 2$. As $d_i = 0$ for all $j = 1, 2$, checking (18)–(19) is equivalent to checking whether the linear system

$$\langle H_i, Z \rangle = b_i\,, \quad i = 1, 2,$$

has a positive semidefinite solution Z, an LMI (convex) problem (see e.g. Corollary 20.3 in Barvinok [4]).

3 The homogeneous case

In this section we specialize the results to the case where the f_i are all (positively) homogeneous polynomials with the same degree p, that is, for every scalar $\lambda > 0$,

$$f_i(\lambda x) = \lambda^p f_i(x)\,, \quad \forall x \in X\,, i = 0, 1, \ldots, m\,.$$

(In fact, it is also true for an arbitrary scalar λ.) A particular case of interest is when $p = 2$, i.e., when the f_i are all quadratic forms $x \mapsto f_i(x) = x'Q_i x$ for some real-valued (n, n)-symmetric matrices Q_i, $i = 0, 1, \ldots, m$.

We first obtain a simplified expression of D, the solvability of PP (remember that the assumptions in Theorem 2 are not satisfied here) and show that there exist optimal solutions of PP with finite support.

Theorem 6 *Let Assumption A and Condition B hold. Then PP is solvable so that* $\max D = \min PP$. *The dual D is:*

$$\max_{\lambda \geq 0} \left\{ -\sum_{i=1}^{m} \lambda_i b_i \,\middle|\, f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \geq 0, \quad x \in X \right\}. \tag{20}$$

In addition, let λ^* *be an optimal solution of D. Then, there exists an optimal solution* μ^* *of PP which is a convex combination of at most* $m+1$ *Dirac measures at points* x_j *that verify*

$$\nabla \left(f_0 + \sum_{i=1}^{m} \lambda_i^* f_i \right)(x_j) = 0\,, \quad j = 1, 2, \ldots, m+1\,. \tag{21}$$

The proof is relegated to Appendix D. We also obtain the following result that takes advantage of the fact that $X(\lambda^*)$ is a cone. Indeed, by homogeneity, the ∇f_i are all (positively) homogeneous polynomials of degree $p-1$, so that

$$\nabla \left(f_0 + \sum_{i=1}^{m} \lambda_i^* f \right)(x) = 0 \Rightarrow \nabla \left(f_0 + \sum_{i=1}^{m} \lambda_i^* f \right)(\alpha x) = 0\,,$$

for every scalar $\alpha > 0$ (and in fact, every scalar α).

Theorem 7 *Assume that* f_i *are all (positively) homogeneous polynomials with degree* p, *and let Assumption A and Condition B hold. Let* λ^* *be an optimal solution to the convex problem D such that* $X(\lambda^*)$ *is a one-dimensional* *cone, then*

$$\max D = \min PP = \min P\,. \tag{22}$$

In addition, let $x_0 \in X(\lambda^*)$ *, then* $x^* := \alpha x_0$ *with* $\alpha^p = b_i / f_i(x_0)$ *(with* $\lambda_i^* > 0$*) is a global minimizer of* P*.*

Proof: From Theorem 6, we already know that both D and PP are solvable so that $\max \mathrm{D} = \min \mathrm{PP}$. Let x_0 be an arbitrary solution in $X(\lambda^*)$. From (12) in Theorem 3, every solution μ^* of PP is a probability measure on $X(\lambda^*)$, that is, for every $f \in L_1(\mu)$,

$$\int f(x)\,\mu(dx) = \int f(\alpha x_0)\nu(d\alpha),$$

for some probability measure ν on the real line.

From the homogeneity of f_i, $i = 0, 1, \ldots, m$, we have

$$\int f_i\,d\mu^* = f_i(x_0)\int \alpha^p\,d\nu \quad i = 0, 1, \ldots, m.$$

Therefore, from (13)

$$\int \alpha^p d\nu = \frac{b_i}{f_i(x_0)} \quad \text{whenever } \lambda_i^* > 0\,.$$

In addition, as μ^* is admissible

$$\begin{aligned} b_k &\geq \int f_k(x)\, d\mu^* \\ &= f_k(x_0) \int \alpha^p\, d\nu\,, \quad \forall k = 1, 2, \ldots, m\,. \end{aligned}$$

But then the point $x^* := \alpha_0 x_0$ with

$$\alpha_0^p = \frac{b_i}{f_i(x_0)} \text{ for some } \lambda_i^* > 0\,,$$

satisfies

$$f_k(x^*) = f_k(x_0)\alpha_0^p = \int f_k(x)\, d\mu^*\,, \quad \forall k = 0, 1, \ldots, m\,,$$

which proves that x^* is feasible and $f_0(x^*) = \min \mathrm{PP}$, so that we may conclude that x^* is a global minimizer of P. ♠

In fact, Theorem 7 can be improved. In view of Corollary 5, if Δ denotes the set of optimal solutions of D, Theorem 7 is valid if only $\cap_{\lambda\in\Delta} X(\lambda)$ is one-dimensional.

3.1 Example: Pure quadratic optimization

Consider the following (pure) quadratic optimization problem

$$\mathrm{P} \mapsto f^* = \min_x \{x' Q_0 x \mid x' Q_i x \leq b_i,\ i = 1, 2, \ldots, m\} \tag{23}$$

where all the Q_i, $i = 0, 1, \ldots, m$ are real-valued (n, n)-symmetric matrices. We make the following assumption:

Assumption C:

(i) Q_i *is positive semi-definite* $i = 1, 2, \ldots, m$ *;*

(ii) Q_0 *has a unique negative eigenvalue (counting its multiplicity) and, whenever* $x \neq 0$, $Q_0 x = 0 \Rightarrow Q_i x \neq 0$ *for all* $i = 1, 2, \ldots, m$ *;*

(iii) There exist nonnegative coefficients λ_i, $i = 1, 2, \ldots, m$, *such that* $Q_0 + \sum_{i=1}^m \lambda_i Q_i$ *is positive semi-definite;*

(iv) Slater's condition holds, i.e., $x_0' Q_i x_0 < b_i$, $i = 1, 2, \ldots, m$ *for some* x_0 *.*

From Theorem 6, the dual problem D associated with P reads:

$$\mathrm{D} \mapsto \max_{\lambda \geq 0, \gamma} \left\{ \gamma \,\middle|\, x' Q_0 x + \sum_{i=1}^m \lambda_i \, x' Q_i x \geq \gamma + \sum_{i=1}^m \lambda_i b_i, \quad \forall x \in X \right\}. \tag{24}$$

or, equivalently,

$$\mathrm{D} \mapsto \max_{\lambda \geq 0} \left\{ -\sum_{i=1}^m \lambda_i b_i \,\middle|\, Q_0 + \sum_{i=1}^m \lambda_i Q_i \succeq 0 \right\}, \tag{25}$$

(with "$A \succeq B$" standing for $A - B$ positive semidefinite) which is called an "LMI" problem and can be solved efficiently via interior point methods (see e.g. [6] and the references therein).

Remark 8 *This is a particular case of the general quadratic case considered in Section 2.4, and D has a dual D^* which reads*

$$D^* \mapsto \min_{Y \in S^n;\, Y \succeq 0} \{\langle Q_0, Y\rangle \mid \langle Q_i, Y\rangle \leq b_i \quad i = 1, 2, \ldots, m\} \tag{26}$$

where S^n denotes the space of real-valued (n, n)-symmetric matrices. D^* is a well-known relaxation of P. The relationship between D^* and PP which are both duals of D with same value, is as follows. Let μ^* be an optimal solution of PP, and let $Y \in R^{n\times n}$ be its matrix of second-order moments, that is,

$$Y(i, j) := \int z_i z_j \, d\mu^* \quad \forall i, j = 1, \ldots, n.$$

Then, obviously, $Y \succeq 0$ and D^* is an equivalent formulation of PP for $\int f_i d\mu^* = \langle Q_i, Y\rangle$ for every $i = 1, \ldots, n$. Conversely, every solution $Y \succeq 0$ of D^* can be decomposed into a convex combination $\sum_k \alpha_k x_k x_k'$ of rank-one matrices $x_k x_k'$ (with $\sum_k \alpha_k = 1$). The probability measure $\mu := \sum_k \alpha_k \delta_{x_k}$ solves PP.

The following result is an application of Theorem 7 in the pure quadratic case.

Theorem 9 *Under Assumption C,* $\max D = f^* = f_0(x^*)$ *for every global optimal solution x^* of P.*

Proof: One may check that C(iii)–(iv) imply that Assumption A and Condition B hold. Moreover, from C(i)–(ii), the null-space N of $Q_0 + \sum_{i=1}^m \lambda_i^* Q_i$ is at most one-dimensional. The conclusion follows from Theorem 7. ♠

Hence computing a global optimal solution for the non-convex problem P reduces to solving the convex "LMI" problem D. In fact, P is a "hidden" convex problem.

References

[1] E.J. Anderson and P. Nash. *Linear Programming in Infinite-Dimensional Spaces*, John Wiley & Sons, Chichester, 1987. E34, E39, E58

[2] A. Ben Tal and M. Teboulle. Hidden convexity in some nonconvex quadratically constrained quadratic programming, *Math. Progr. 72*, pp. 51–63, 1996. E30

[3] E.J. Balder. *Lectures on Young Measures*, Cahiers de Mathématiques de la Décision, CEREMADE, Université Paris IX-Dauphine, 1994. E30, E34, E40

[4] A. Barvinok. *Convexity, Duality, and Optimization*, Lecture Notes, Department of Mathematics, The University of Michigan, 1998. E47, E64

[5] P. Billingsley. *Convergence of Probability Measures*, Wiley, New York, 1968. E61

[6] L. Vandenberghe and S. Boyd. Semidefinite Programming, *SIAM Review 38*, pp. 49–95, 1996. E32, E45, E52

[7] B.D. Craven and J.J. Koliha, Genralizations of Farkas' theorem, *SIAM J. Math. Anal. 8*, pp. 983–997, 1977. E36

[8] J.B. Hiriart-Urruty. *Optimisation et Analyse Convexe*, PUF, Paris, 1998. E30

[9] J.B. Hiriart-Urruty. Conditions for global optimality 2, *J. Global Optim.*, to appear. E30

[10] L.S. Lasdon, *Optimization Theory for large Systems*, The Macmillan Company, New York, 1970. E30, E37

[11] E. Mascolo and L. Migliacco, Relaxation methods in Control Theory, *Appl. Math. Optim. 20*, pp. 97–103, 1989. E30, E31, E34

[12] M.D. Perlman. Jensen's inequality for a convex vector-valued function on a infinite dimensional space, *J. Multivariate Anal. 4*, pp. 52–65, 1974. E38

[13] B.T. Polyak. Convexity of quadratic transformations and its use in control and optimization, *J. Optim. Th. Appl. 99*, pp. 553–583, 1998. E30

[14] F. Topsoe. *Topology and Measure*, Lecture Notes in Mathematics, Springer, 1970. E64

[15] J. Warga, *Optimal Control of Differential and Functional Equations*, Academic Press, New York, 1972. E30, E31, E34

A Proof of Theorem 1

(i) We first prove the solvability of D. In view of Assumption A, P is of course feasible and therefore we have $\sup \mathrm{D} < +\infty$. In addition, since the feasible set of D is not empty, we also have $\sup \mathrm{D} > -\infty$. Hence, consider a maximizing sequence $\{\gamma^k, \lambda_i^k \geq 0\}$, i.e., a sequence such that

$$f_0(x) + \sum_{i=1}^{m} \lambda_i^k [f_i(x) - b_i] \geq \gamma^k \quad \forall x \in X. \tag{27}$$

and

$$\gamma^k \uparrow \sup \mathrm{D}. \tag{28}$$

In particular, with x_j as in Assumption A,

$$f_0(x_j) + \sum_{i=1}^{m} \lambda_i^k [f_i(x_j) - b_i] \geq \gamma^k,$$

which implies that λ_j is bounded, otherwise

$$f_0(x_j) + \sum_{i=1}^{m} \lambda_i^k [f_i(x_j) - b_i] \to -\infty,$$

in contradiction with $\gamma_k \uparrow \sup\ \mathrm{D}$. Hence, with a similar argument, all the λ_i are nonnegative and bounded above.

Therefore, there exist nonnegative scalars γ^*, λ_i^*, $i = 1, 2, \ldots, m$, and a subsequence $\{k_j\}$ such that, as $j \to \infty$,

$$\gamma^{k_j} \to \gamma^* \text{ and } \lambda_i^{k_j} \to \lambda_i^*, \quad i = 1, 2, \ldots, m,$$

so that, $\gamma^* = \sup\ \mathrm{D}$, and obviously (fix x and let $j \to \infty$ in (27))

$$f_0(x) + \sum_{i=1}^{m} \lambda_i^* [f_i(x) - b_i] \geq \gamma^* \quad \forall x \in X. \tag{29}$$

Equivalently,

$$\inf_x \left\{ f_0(x) + \sum_{i=1}^{m} \lambda_i^* [f_i(x) - b_i] \right\} = \gamma^* = \sup \mathrm{D}. \tag{30}$$

This proves that λ^* solves D, i.e., $\sup \mathrm{D} = \max \mathrm{D}$.

(ii) We now show that there is no duality gap between PP and D. Under Slater's condition, i.e., when there is some $x_0 \in S$ such that $f_i(x_0) < b_i$

for all $i = 1, 2, \ldots, m$ the absence of a duality gap follows from e.g. Theorem 3.13, p. 55 in Anderson and Nash. Indeed, Slater's condition is just the interior point condition needed in that theorem. However, under the weaker Assumption A, such an interior point may not exist.

As $X := R^n$ is locally compact separable, consider a nondecreasing sequence of compact sets $K_i \uparrow X$, $i = 0, 1, \ldots$, such that K_0 contains $x_j, j = 1, 2, \ldots, m$, with x_j as in Assumption A. Let D_i be the optimization problem

$$\mathrm{D}_i \mapsto \sup_{\lambda \geq 0, \gamma} \left\{ \gamma \,\middle|\, f_0(x) + \sum_{k=1}^{m} \lambda_k [f_k(x) - b_k] \geq \gamma \quad \forall x \in K_i \right\},$$

which is the dual of the optimization problem

$$\mathrm{PP}_i \mapsto \left\{ \begin{array}{rcl} \multicolumn{3}{l}{\inf_{\mu \in \mathcal{M}(K_i), \mu \geq 0} \int f_0 \, d\mu} \\ \int [f_k(x) - b_k] \mu(dx) & \leq & 0, \; k = 1, 2, \ldots, m \\ \mu(K_i) & = & 1 \end{array} \right.$$

In view of Assumption A and B and the fact that K_i is compact, both D_i and PP_i are consistent with finite value. In addition, from the compactness of K_i, and the fact that the restrictions to K_i of $f_k, k = 0, 1, \ldots, m$ are continuous, there is no duality gap between D_i and PP_i, and PP_i is solvable. Indeed, the feasible set of PP_i is a compact subset of $\mathcal{M}(K_i)$ for the weak* topology $\sigma(\mathcal{M}(K_i), C(K_i))$ (with $C(K_i)$ the space of continuous functions on K_i), and $\int f_0 d\mu$ is a continuous linear functional for that topology. Thus, PP_i is solvable. For the absence of a duality gap it suffices to prove that the set $H \subset R^{m+2}$, defined by

$$H := \left\{ \left(T\mu + v, \langle 1, \mu \rangle, \int f_0 \, d\mu \right) \mid v \in (R^m)^+, \ \mu \geq 0, \ \mu \in \mathcal{M}(K_i) \right\}$$

is closed (see e.g. Theorem 3.9 in [1]), which also follows from compactness arguments.

In addition, with similar arguments as in (i), D_i is also solvable. Therefore, let $\mu_i \in \mathcal{M}(K_i)$ be an optimal solution of PP_i and let $\{\lambda^i \geq 0, \gamma^i\}$ be an optimal solution of D_i. From the absence of a duality gap between PP_i and DD_i, we have

$$\inf \mathrm{PP}_i = \min \mathrm{PP}_i = \max \mathrm{D}_i \geq \sup \mathrm{D} = \max \mathrm{D} \,. \tag{31}$$

Let $i \to \infty$ so that $\gamma^i \downarrow \gamma^* \geq \max\ \mathrm{D}$. It suffices to prove that $\gamma^* = \max\ \mathrm{D}$ since then, the sequence μ_i which satisfies

$$\int f_0 \, d\mu_i = \gamma^i \downarrow \gamma^* \leq \inf \mathrm{PP},$$

will be a minimizing sequence of PP with limit value γ^*.

Consider the sequence of $\{\lambda^i\}$. As $x_j \in K_i$ for all i, we have

$$f_0(x_j) + \sum_{k=1}^{m} \lambda_k^i [f_k(x_j) - b_k] \geq \gamma^i \geq \gamma^* \,, \quad j = 1, 2, \ldots, m. \tag{32}$$

Therefore, for each $k = 1, 2, \ldots, m$, the sequence $\{\lambda_k^i\}$ is bounded otherwise (32) yields a contradiction. Therefore, there is a subsequence $\{i_p\}$ and coefficients $\lambda_k^*, k = 1, 2, \ldots, m$ such that, as $p \to \infty$,

$$\lambda_k^{i_p} \to \lambda_k^*, \quad k = 1, 2, \ldots, m.$$

Now fix $x \in X$. As $K_i \uparrow X$, there is some p_0 such that $x \in K_{i_p}$ for every $p \geq p_0$, and therefore

$$f_0(x) + \sum_{k=1}^{m} \lambda_k^{i_p}[f_k(x) - b_k] \geq \gamma^{i_p}.$$

Letting $p \to \infty$ yields

$$f_0(x) + \sum_{k=1}^{m} \lambda_k^*[f_k(x) - b_k] \geq \gamma^*.$$

As x was arbitrary, (λ^*, γ^*) is feasible for D. This and $\gamma^* \geq \max \mathrm{D}$ proves that (λ^*, γ^*) is an optimal solution of D.

B Proof of Theorem 2

PP is consistent since P is. Therefore, consider a minimizing sequence $\{\mu_n\}$ in $\mathcal{M}(X)$, that is, from $\mu_n(X) = 1$ and $\mu_n \geq 0$, a sequence in $\mathcal{P}(X)$ that satisfies

$$\int f_0\, d\mu_n \downarrow \inf \mathrm{PP}; \int f_i d\mu_n \leq b_i\,, \quad i = 1, 2, \ldots, m\,.$$

As f_m is inf-compact, there is some $M > 0$ such that $x \mapsto w(x) := M + f_m(x) \geq 1$ for all $x \in X$. Consider the associated subsequence of measures

$$\varphi_n(B) := \int_B w\, d\mu_n\,, \quad B \in \mathcal{B},\ n = 1, 2, \ldots\,.$$

The constraints $\int f_i d\mu_n \leq b_i$, $i = 1, 2, \ldots, m$ read

$$\int \frac{f_i}{w}\, d\varphi_n \leq b_i\,, \quad i = 1, 2, \ldots, m,$$

and $\int f_0 d\mu_n \downarrow \inf \mathrm{PP}$ reads $\int \frac{f_0}{w} d\varphi_n \downarrow \inf \mathrm{PP}$. As the f_i are all continuous and $w \geq 1$, under Condition (b), the functions $f_i/w \in C_0(X)$, $i = 0, 1, \ldots, m-1$, where $C_0(X)$ is the Banach space of continuous functions that vanish at infinity, equipped with the sup-norm, and whose topological dual is $\mathcal{M}(X)$.

From $\int w d\mu_n = \int f_m d\mu_n + M \leq b_m + M$, it follows that the sequence of measures $\{\varphi_n\}$ is norm-bounded, so that by weak* sequential compactness of the unit ball of $\mathcal{M}(X)$, there is a measure φ and a subsequence $\{\varphi_{n_k}\}$ such that

$$\lim_{k \to \infty} \int h\, d\varphi_{n_k} = \int h\, d\varphi\,, \quad \forall h \in C_0(X)\,,$$

and therefore, in particular, as $f_i/w \in C_0(X)$, $\forall i \neq m$,

$$\int \frac{f_i}{w} d\varphi = \lim_{k\to\infty} \int \frac{f_i}{w}\, d\varphi_{n_k} \leq b_i\,, \quad i = 1, 2, \ldots, m-1\,. \tag{33}$$

In addition,

$$\int \frac{f_0}{w}\, d\varphi = \lim_{k\to\infty} \int \frac{f_0}{w}\, d\varphi_{n_k} = \inf \mathrm{PP}. \tag{34}$$

Introduce the measure μ^* defined by $\mu^*(B) := \int_B w^{-1} d\varphi$, $B \in \mathcal{B}$, so that (33)–(34) read

$$\int f_i d\mu^* \leq b_i\,, \quad i = 1, 2, \ldots, m-1\,; \int f_0\, d\mu^* = \inf \mathrm{PP}. \tag{35}$$

We next prove that μ^* is a probability measure. As w is inf-compact, $w^{-1} \in C_0(X)$, so that

$$\begin{aligned} \mu^*(X) &= \int w^{-1} d\varphi \\ &= \lim_{k\to\infty} \int w^{-1} d\varphi_{n_k} \\ &= \lim_{k\to\infty} \mu_{n_k}(X) = 1. \end{aligned}$$

Therefore, by the Portmanteau Theorem (see e.g. Billingsley [5]), the sequence φ_{n_k} converges "weakly" to φ and not only "weak*", that is,

$$\int h\, d\varphi_{n_k} \to \int h\, d\varphi\,, \quad \forall h \in C_b(X).$$

It remains to prove that $\int f_m d\mu^* \leq b_m$. This follows from the definition of w. Indeed, as $f_m w^{-1} \in C_b(X)$, one has

$$\begin{aligned}\int f_m\, d\mu^* = \int f_m w^{-1}\, d\varphi &= \lim_{k\to\infty} \int f_m w^{-1}\, d\varphi_{n_k} \\ &= \lim_{k\to\infty} \int f_m\, d\mu_{n_k} \leq b_m\,.\end{aligned}$$

The latter combined with (35) implies that μ^* is admissible for PP and $\int f_0 d\mu^* = \inf \mathrm{PP}$, which proves that PP is solvable.

C Proof of Theorem 3

(a) Let (λ^*, γ^*) be an optimal solution of D and let μ^* be an optimal solution of PP. From the absence of a duality gap between PP and D we have:

$$\gamma^* = \int f_0\, d\mu^*.$$

In addition, from

$$0 \leq f_0(x) + \sum_{i=1}^{m} \lambda_i^* [f_i(x) - b_i] - \gamma^*\,, \quad \forall x \in X\,,$$

and $\int(f_i(x)-b_i)d\mu^* \leq 0$ for all $i=1,2,\ldots,m$, integration w.r.t. μ^* yields

$$0=\sum_{i=1}^m \lambda_i^* \int (f_i(x)-b_i)\,d\mu^*.$$

Consequently, from the nonnegativity of each term, one has

$$\lambda_i^* \int (f_i\,d\mu^* - b_i)\,d\mu^* = 0\,, \quad i=1,2,\ldots,m,$$

i.e. (13) holds and

$$0=\int \left[f_0(x)+\sum_{i=1}^m \lambda_i^*[f_i(x)-b_i]-\gamma^*\right]\,d\mu^*.$$

This clearly implies that

$$0=f_0(x)+\sum_{i=1}^m \lambda_i^*[f_i(x)-b_i]-\gamma^*, \quad \mu^* \text{ a.e.}$$

which yields (12), and (14) follows whenever the f_i are all differentiable.

(b) Let $X(\lambda^*)$ be the compact set of global minimizers of $L(.,\lambda^*)$ and let μ^* be an optimal solution of PP. We have just seen that μ^* is a probability measure on $X(\lambda^*)$. Thus, we could have replaced X by $X(\lambda^*)$. The space $\mathcal{P}(X(\lambda^*))$ of probability measures on $X(\lambda^*)$ is compact and convex for the weak* topology of $\mathcal{M}(X(\lambda^*))$. By the Krein-Milman Theorem, $\mathcal{P}(X(\lambda^*))$ is the weak* closure of the convex hull of its extreme points which are Dirac measures (see e.g. [14]).

Let I^* be such that $i \in I^*$ if and only if $\int f_i d\mu^* = b_i$, i.e., I^* denotes the set of active constraint at μ^*.

The optimal value of PP is the same as if we had removed the constraints $i \notin I^*$ in PP and replaced the inequality by equality for $i \in I^*$. Indeed, if by removing the inactive constraints $i \notin I^*$, we obtain a strictly better solution ν (eventually with $\int f_i d\nu > b_i$ for some $i \notin I^*$) then there would be a convex combination $\mu := \alpha\nu + (1-\alpha)\mu^*$ with $\alpha > 0$, such that $\int f_i d\mu \leq b_i$ for all $i \notin I^*$ and with $\int f_0 d\mu < \int f_0 d\mu^*$, a contradiction.

Let $H_i \subset \mathcal{M}(X(\lambda^*))$, $i \in I^*$ be the hyperplanes $\{\mu \mid \int f_i d\mu = b_i\}$ associated with the active constraints at μ^*. Since the f_i are continuous, the H_i are weak* closed and convex in $\mathcal{M}(X(\lambda^*))$ so that $\mathcal{P}(X(\lambda^*)) \cap [\cap_{i\in I^*} H_i]$ is a convex set in $\mathcal{M}(X(\lambda^*))$, compact for the weak* topology. In addition, the linear functional $\int f_0 d\mu$ is continuous in that topology. Therefore, it attains its minimum at an extreme point which, by Caratheodory Theorem (see e.g. [4, Th. 28.2, p. 66]), is a convex combination of at most $s+1$ extreme points of $\mathcal{P}(X(\lambda^*))$, i.e., $s+1$ Dirac measures on $X(\lambda^*)$.

D Proof of Theorem 6

From Theorem 1, D is solvable and there is no duality gap, that is $\max \mathrm{D} = \inf \mathrm{PP}$. Let (λ^*, γ^*) be an optimal solution of D so that $\gamma^* := \max \mathrm{D}$. As the f_i's are homogeneous polynomials, a minimizer x^* of $f_0 + \sum_i \lambda_i^*[f_i - b_i]$ satisfies

$$\nabla \left(f_0 + \sum_{i=1}^m \lambda_i^* f_i \right)(x^*) = 0\,,$$

and using the homogeneity of the f_i's (hence of the ∇f_i's), we obtain

$$\left(f_0 + \sum_{i=1}^m \lambda_i^* f_i \right)(x^*) = 0\,,$$

which in turn yields,

$$\gamma^* = -\sum_{i=1}^m \lambda_i^* b_i\,.$$

Hence, D simplifies to (20).

We now consider the solvability of PP. Let $K_n \uparrow X$ be the sequence of compact sets already considered in the proof of Theorem 1. We already know that

$$\max \mathrm{D}_n = \min \mathrm{PP}_n \downarrow \inf \mathrm{PP} = \max \mathrm{D}$$

(cf. (31)), and

$$\lambda^{n_k} \to \lambda^*, \quad \gamma_{n_k} \downarrow \gamma^* = \max \mathrm{D}$$

for some subsequence $\{\lambda^{n_k}, \gamma_{n_k}\}$ of optimal solutions to D_n. Now let x_n be a minimizer of $(f_0 + \sum_{i=1}^m \lambda_i^n f_i)$ on K_n. We must have

$$\left(f_0 + \sum_{i=1}^m \lambda_i^{n_k} f_i\right)(x_{n_k}) = \gamma_{n_k} + \sum_{i=1}^m \lambda_i^{n_k} b_i \geq \sum_{i=1}^m [\lambda_i^{n_k} - \lambda_i^*] b_i \,,$$

since $\gamma_{n_k} \downarrow \gamma^*$. In addition, given $\epsilon > 0$, from the convergence of λ^{n_k} to λ^* it follows that

$$\left(f_0 + \sum_{i=1}^m \lambda_i^{n_k} f_i\right)(x_{n_k}) \geq -\epsilon \,, \tag{36}$$

for all k sufficiently large, say $k \geq k_0$. Moreover, also from $\gamma_{n_k} \downarrow \gamma^*$, it follows that

$$\left(f_0 + \sum_{i=1}^m \lambda_i^{n_k} f_i\right)(x_{n_k}) \geq \left(f_0 + \sum_{i=1}^m \lambda_i^{n_{k+p}} f_i\right)(x_{n_{k+p}}) \geq -\epsilon \,, \quad \forall p = 1, 2, \ldots \tag{37}$$

and as $x_n \in K_{n+p}$ for all $p = 1, 2, \ldots$, it follows that

$$\left(f_0 + \sum_{i=1}^m \lambda_i^{n_{k+p}} f_i\right)(x_{n_k}) \geq \left(f_0 + \sum_{i=1}^m \lambda_i^{n_{k+p}} f_i\right)(x_{n_{k+p}}) \geq -\epsilon \,, \quad \forall p = 1, 2, \ldots \tag{38}$$

Now, fix $s \geq k_0$ and assume that

$$\left(f_0 + \sum_{i=1}^{m} \lambda_i^{n_s} f_i\right)(x_{n_s}) = -\delta < 0\,.$$

Take $y = \alpha x_{n_s}$ with $\alpha > 1$ sufficiently large to ensure $-\alpha^p\delta < -\epsilon$. From $K_n \uparrow X$, there is some p_0 such that $y \in K_{n_{s+p}}$ for all $p \geq p_0$ and thus,

$$\left(f_0 + \sum_{i=1}^{m} \lambda_i^{n_{s+p}} f_i\right)(y) \geq \left(f_0 + \sum_{i=1}^{m} \lambda_i^{n_{s+p}} f_i\right)(x_{n_{s+p}}) \quad p = 1, 2, \ldots\,. \tag{39}$$

Therefore,

$$\begin{aligned}
\left(f_0 + \sum_{i=1}^{m} \lambda_i^{n_{s+p}} f_i\right)(y) &= \alpha^p \left(f_0 + \sum_{i=1}^{m} \lambda_i^{n_{s+p}} f_i\right)(x_{n_s}) \\
&= -\alpha^p\delta \quad \text{[by homogeneity]} \\
&\geq \left(f_0 + \sum_{i=1}^{m} \lambda_i^{n_{s+p}} f_i\right)(x_{n_{s+p}}) \quad \text{[by (38)]} \\
&\geq -\epsilon \quad \text{[by (36)]},
\end{aligned}$$

a contradiction.

Therefore, there is a subsequence n_k such that for sufficiently large k,

$$f_0(x) + \sum_{i=1}^{m} \lambda_i^{n_k} f_i(x) \geq 0 \quad \forall x \in K_{n_k}.$$

But, as K_i has nonempty interior and $0 \in \mathrm{int} K_i$ for sufficiently large i, by homogeneity

$$f_0(x) + \sum_{i=1}^m \lambda_i^{n_k} f_i(x) \geq 0 \,, \quad \forall x \in X \,.$$

But then $\gamma_{n_k} = -\sum_{i=1}^m \lambda_i^{n_k} b_i$ and D_{n_k} is equivalent to solving

$$\max_{\lambda \geq 0} \left\{ -\sum_{i=1}^m \lambda_i b_i \,\middle|\, f_0 + \sum_{i=1}^m \lambda_i f_i \geq 0 \right\},$$

which is nothing less than solving D! Thus,

$$\max \mathrm{D}_{n_k} = \gamma_{n_k} = \gamma^* = \max \mathrm{D}, \quad \forall k \geq k_0.$$

As there is no duality gap between D_n and PP_n and both are solvable, from

$$\inf \mathrm{PP} = \max \mathrm{D} = \max \mathrm{D}_n = \min \mathrm{PP}_n,$$

it follows that PP is solvable (take as μ^* an optimal solution of PP_{n_k}).

Finally, as solving PP reduces to solving PP_n for some n sufficiently large, an optimal solution of PP is a probability measure μ^* on a compact set K_n. That there is an optimal solution μ^* which is a convex combination of at most $m+1$ Dirac measures at points x_i, $i = 1, 2, \ldots, m+1$ then follows with similar arguments as in the proof of Theorem 3(b). That the x_i's satisfy (21) follows from the fact that μ^* must be concentrated on $X(\lambda^*)$ (cf. (14) in Theorem 3).