Solving Infinite-dimensional Optimization Problems by Polynomial Approximation

Olivier Devolder[1], François Glineur[1], and Yurii Nesterov[1]

ICTEAM & IMMAQ, Université catholique de Louvain,
CORE, Voie du Roman Pays, 34, Louvain-la-Neuve, B-1348, Belgium,
{Olivier.Devolder,Francois.Glineur,Yurii.Nesterov}@uclouvain.be[†]

Summary. We solve a class of convex infinite-dimensional optimization problems using a numerical approximation method that does not rely on discretization. Instead, we restrict the decision variable to a sequence of finite-dimensional linear subspaces of the original infinite-dimensional space and solve the corresponding finite-dimensional problems in a efficient way using structured convex optimization techniques. We prove that, under some reasonable assumptions, the sequence of these optimal values converges to the optimal value of the original infinite-dimensional problem and give an explicit description of the corresponding rate of convergence.

1 Introduction

Optimization problems in infinite-dimensional spaces, and in particular in functional spaces, were already considered in the 17^{th} century: the development of the calculus of variations, motivated by physical problems, focused on the development of necessary and sufficient optimality conditions and finding closed-form solutions. Much later, the advent of computers in the mid-20^{th} century led to the consideration of finite-dimensional optimization from an algorithmic point of view, with linear and nonlinear programming. Finally, a general theory of optimization in normed spaces began to appear in the 70's [8, 2], leading to a more systematic and algorithmic approach to infinite-dimensional optimization.

Nowadays, infinite-dimensional optimization problems appear in a lot of active fields of optimization, such as PDE-constrained optimization [7], with applications to optimal control, shape optimization or topology optimization. Moreover, the generalization of many classical finite optimization problems to a continuous time setting lead to infinite-dimensional problems.

[†] The first author is a F.R.S.-FNRS Research Fellow. This text presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the authors.

From the algorithmic point of view, these problems are often solved using discretization techniques (either discretization of the problem or discretization of the algorithm). In this work, we consider a different method of resolution that does not rely on discretization: instead, we restrict the decision variables to a sequence of finite-dimensional linear subspaces of the original infinite-dimensional space, and solve the corresponding finite-dimensional problems.

2 Problem class and examples

Consider a normed vector space $(X, \|.\|_X)$ of infinite dimension and its topological dual $(X', \|.\|_{X'})$ equipped with the dual norm. We focus on the following class of convex infinite-dimensional optimization problems:

$$P^* = \inf_{x \in X} \langle c, x \rangle \text{ s.t } \langle a_i, x \rangle = b_i \quad \forall i = 1, \dots, L \text{ and } \|x\|_X \leq M \tag{P}$$

where $c \in X', a_i \in X'$, $M \in \mathbb{R}_{++}$, $b_i \in \mathbb{R}$ for all $i = 1, \dots, L$ (L is finite) and P^* denotes the optimal objective value. This problem class, with a linear objective, linear equalities and a single nonlinear inequality bounding the norm of the decision variable, is one of the simplest that allows us to outline and analyze our approximation technique. Nevertheless, it can be used to model many applications, among which the following simple continuous-time supply problem, which we describe for the sake of illustration.

A company buys a specific substance (for example oil or gas) in continuous time. Assume that this substance is made of L different constituents and that its composition continuously changes with time. In the same way, the price of this substance follows a market rate and therefore also changes in continuous time.

The finite time interval $[0, T]$ represents one production day. Assume that, for each constituent i, a specific daily demand b_i must be satisfied at the end of the day. We want to compute a purchase plan $x(t)$, i.e. the quantity of substance to buy at each time t, such that it meets the daily demands for a minimal total cost. For this application, the decision functional space X can be taken as the space of continuous functions on $[0, T]$ (see also Section 5 for other examples of suitable functional spaces). Denoting the price of the substance at time t by $\gamma(t)$, the amount of constituent i in the substance at time t by $\alpha_i(t)$, we obtain the following infinite-dimensional problem

$$\inf_{x \in X} \int_0^T \gamma(t) x(t) dt \text{ s.t } \int_0^T \alpha_i(t) x(t) dt = b_i \ \forall i \text{ and } 0 \leq x(t) \leq K \ \forall t \in T$$

where we also impose a bound for the maximal quantity that we can buy at each moment of time. The objective function and equality constraints are linear, so that we only need to model the bound constraints as a norm constraint. This is easily done with a linear change of variable: letting $x(t) = \frac{1}{2}K + \bar{x}(t)\ \forall t$, the bound constraint becomes $\|\bar{x}\|_\infty \leq \frac{1}{2}K$, which now fits the format of (P).

This model can also be used to compute how to modify an existing purchase plan when changes in the demands occur. Denote the modification of the daily demand for the constituent i by Δb_i and the change in our purchase quantity at time t by $\tilde{x}(t)$, and assume we do not want to modify the existing planning too much, so that we impose the constraint $\|\tilde{x}\| \leq M$ for a given norm. We obtain the following infinite-dimensional problem:

$$\inf_{x \in X} \int_0^T \gamma(t)\tilde{x}(t)dt \text{ s.t } \int_0^T \alpha_i(t)\tilde{x}(t)dt = \Delta b_i \ \forall i = 1, \ldots, L \text{ and } \|\tilde{x}\| \leq M$$

which also belongs to problem class (P). Finally, note that this problem class allows the formulation of continuous linear programs (CLPs, see [2]), such as

$$\inf_{x \in X} \int \gamma(t)x(t)dt \text{ s.t } \int \alpha_i(t)x(t)dt = b_i \ \forall i \text{ and } x(t) \geq 0 \ \forall t$$

provided we know an upper bound K on the supremum of $x(t)$, so that the nonnegativity constraint can be replaced by $0 \leq x(t) \leq K \ \forall t$, which can be rewritten using the infinity norm with a linear change of variables as in first example above.

3 Finite-dimensional approximations

We propose to approximate infinite-dimensional problem (P) by a sequence of finite-dimensional approximations. Let $\{p_1, \ldots, p_n, \ldots\} \subset X$ be an infinite family of linearly independent elements of X and denote by $X_n = \text{span}\{p_1, \ldots, p_n\}$, the finite-dimensional linear subspace generated by the first n elements of this family.

Replacing the infinite-dimensional space X in (P) by X_n, we obtain the following family of problems with optimal values P_n^*

$$P_n^* = \inf_{x \in X_n} \langle c, x \rangle \text{ s.t } \langle a_i, x \rangle = b_i \ \forall i = 1, \ldots, L \text{ and } \|x\|_X \leq M \ . \qquad (\text{P}_n)$$

Expressing function x as $x = \sum_{i=1}^n x_i p_i$ and denoting the finite vector of variables x_i by $\mathbf{x}$ leads to the following family of equivalent finite-dimensional formulations

$$P_n^* = \inf_{\mathbf{x} \in \mathbb{R}^n} \langle c^{(n)}, \mathbf{x} \rangle \text{ s.t } \langle a_i^{(n)}, \mathbf{x} \rangle = b_i \ \forall i = 1, \ldots, L \text{ and } \Big\| \sum_{i=1}^n x_i p_i \Big\|_X \leq M \ ,$$

where $c^{(n)}$ and $a_i^{(n)}$ are vectors in $\mathbb{R}_n$ whose components are defined by $[c^{(n)}]_j = \langle c, p_j \rangle$ and $[a_i^{(n)}]_j = \langle a_i, p_j \rangle \ \forall j = 1, \ldots, n$ and $\forall i = 1, \ldots, L$.

For our approach to be effective, these problems must be solvable by existing algorithms for finite-dimensional optimization. In particular, we would like to ensure that the bounded norm inequality can be handled by existing efficient optimization methods (the other components of the problem, namely the linear objective and linear equalities, are usually easily handled). We now list some explicit situations where this is indeed the case.

1. The easiest case corresponds to situations where X is a Hilbert space. Indeed, if we choose in that case $\{p_1, \ldots, p_n\}$ to be an orthonormal basis of X_n, we have that $\|\sum_{i=1}^n x_i p_i\|_X = \|\mathbf{x}\|_2$, where the last norm is the standard Euclidean norm of $\mathbb{R}^n$. The bounded norm inequality becomes a simple convex quadratic constraint, hence the approximation problem (P_n) can easily be solved (in fact, it admits a solution in closed form).

In the rest of this list, we focus on situations where functional space X is a Lebesgue or Sobolev space (see Section 5 for some properties) and where the basis elements $p_1, p_2, \ldots$ are polynomials (hence the title of this work), because this leads in many situations to problems that can be efficiently solved. We can take for example the monomial basis $X_n = \operatorname{span}\{1, t, \ldots, t^{n-1}\}$, which means that variable x in problem (P_n) can be written $x(t) = \sum_{i=0}^{n-1} x_i t^i$ and becomes a polynomial of degree $n-1$.

2. Let $[a, b]$ denote a bounded, semi-infinite or an infinite interval. When $X = L^\infty([a, b])$, the norm inequality $\|x\|_\infty \leq M$ can be formulated as $-M \leq x(t) \leq M \; \forall t \in [a, b]$, which is equivalent to requiring positivity of both polynomials $x(t) + M$ and $M - x(t)$ on interval $[a, b]$. This in turn can be formulated as a semidefinite constraint, using the sum of squares approach (see e.g. [10]). Therefore, problems (P_n) can be efficiently solved as a semidefinite programming problem, using interior-point methods with polynomial-time worst-case algorithmic complexity.
3. When X is the Sobolev space $W^{k,\infty}([a, b])$, we have that constraint $\|x\|_{k,\infty} \leq M$ is equivalent to $-M \leq x^{(l)}(t) \leq M \forall t \in [a, b]$ and $\forall l \leq k$, where $x^{(l)}(t)$ is the l^{th} derivative of $x(t)$, whose coefficients depend linearly on those of vector $\mathbf{x}$. Therefore, as in the previous case, we solve the corresponding (P_n) as a semidefinite programming problem.
4. In the case of $X = L^q([a, b])$ where q is an *even* integer, we use Gaussian quadrature to obtain an suitable finite-dimensional representation of the constraint $\|x\|_q = (\int_a^b |x(t)|^q dt)^{1/q} \leq M$. We use the following result (see e.g. [6]):

   Theorem 1. *Given an integer m, there exists a set of m abscissas $\{z_1, z_2, \ldots, z_m\}$ and a set of m positive weights $\{w_1, w_2, \ldots, w_m\}$ such that the quadrature formula $\int_a^b f(x)dx \approx \sum_{i=1}^m w_i f(z_i)$ is exact for all polynomials of degree less or equal to $2m - 1$.*

   We now use the fact that, since $x(t)$ is a polynomial of degree at most $n - 1$, $|x(t)|^q$ is a polynomial of degree at most $q(n - 1)$, so that we can choose $m = \frac{1}{2}q(n-1) + 1$ and have $\int_a^b |x(t)|^q dt = \sum_{i=1}^{\frac{1}{2}q(n-1)+1} w_i \lambda_i^q$ where $\lambda_i = x(z_i)$; note that quantities λ_i depend linearly on the coefficients $\mathbf{x}$ of polynomial $x(t)$. The bound constraint can now be written as $\sum_{i=1}^{\frac{1}{2}q(n-1)+1} w_i \lambda_i^q \leq M^q$, which is a structured convex constraint on vector of variables $\mathbf{x}$. Because a self-concordant barrier is known for this set [9, Ch. 6], it can be solved in polynomial time with an interior-point method. The same kind of approach can be used to obtain an explicit translation in finite dimension of the polynomial approximation when $X = W^{k,q}([a, b])$ for even integers q.

Now that we know how to solve problems (P_n) efficiently, we show in the next section that, under some reasonable assumptions, the sequence of their optimal values P_n^* converges to the optimal value of the original infinite-dimensional problem P^* when $n \to +\infty$.

4 Convergence of the approximations

The optimal values of problems (P) and (P_n) clearly satisfy $P^* \leq P_n^*$; moreover, $P_{n+1}^* \leq P_n^*$ holds for all $n \geq 0$. In order to prove that P_n^* converges to P^*, we also need to find an upper bound on the difference $P_n^* - P^*$. Our proof requires the introduction of a third problem, a relaxation of problem (P_n) where the equality constraints are only satisfied approximately. More specifically, we define the linear operator $A : X \to \mathbb{R}^L$ by $[Ax]_i = \langle a_i, x\rangle$, form the vector $b = (b_1, b_2, \ldots, b_L)$ and impose that the norm of the residual vector $Ax - b$ is bounded by a positive parameter ϵ:

$$P_{n,\epsilon}^* = \inf_{x\in X_n} \langle c, x\rangle \text{ s.t } \|Ax - b\|_q \leq \epsilon \text{ and } \|x\|_X \leq M \qquad (\mathrm{P}_{n,\epsilon})$$

(we equip $\mathbb{R}^L$, the space of residuals, with the classical q-norm $\|.\|_q$ norm and define the conjugate exponent q' by $\frac{1}{q} + \frac{1}{q'} = 1$). We clearly have $P_{n,\epsilon}^* \leq P_n$. Our proof of an upper bound for the quantity $P_n^* - P^* = (P_n^* - P_{n,\epsilon}^*) + (P_{n,\epsilon}^* - P^*)$ proceeds in two steps: we first prove an upper bound on $P_{n,\epsilon}^* - P^*$ for a specific value of ϵ depending on n, and then use a general regularity theorem to establish a bound on the difference $P_n^* - P_{n,\epsilon}^*$.

We use the following notations: for $x \in X$, an element of best approximation of x in X_n is denoted by $P_{X_n}(x) = \arg\min_{p\in X_n} \|x - p\|_X$ (such kind of elements exists as X is a normed vector space and X_n is a finite-dimensional linear subspace see e.g. section 1.6 in [4]; in case it is not unique, it is enough to select one of these best approximations in the following developments), while the corresponding best approximation error is $E_n(x) = \min_{p\in X_n} \|x - p\|_X = \|x - P_{X_n}(x)\|_X$.

4.1 Upper bound on $P_{n,\epsilon}^* - P^*$

Assume that problem (P) is solvable. This is true for example if X is a reflexive Banach space or the topological dual of a separable Banach space (see [12] and Section 5 for further comments on this issue). Let x_{opt} be an optimal solution to (P) and let us consider $P_{X_n}(x_{\mathrm{opt}})$, its best approximation in X_n (note that if (P) is not solvable, we can consider for all $\mu > 0$ a μ-solution x_μ of this problem, i.e. a feasible solution such that $< c, x_\mu > \leq P^* + \mu$, and replace P^* by $P^* + \mu$ in the following developments).

First, $\|P_{X_n}(x_{\mathrm{opt}})\|_X$ can be bigger than $\|x_{\mathrm{opt}}\|_X$, and does not necessarily satisfy the norm inequality constraint, but we have $\|P_{X_n}(x_{\mathrm{opt}})\|_X \leq \|x_{\mathrm{opt}}\|_X + \|x_{\mathrm{opt}} - P_{X_n}(x_{\mathrm{opt}})\|_X \leq M + E_n(x_{\mathrm{opt}})$. Therefore, if we choose $\lambda = \frac{M}{M+E_n(x_{\mathrm{opt}})}$ and $\overline{p} = \lambda P_{X_n}(x_{\mathrm{opt}})$, we obtain $\|\overline{p}\|_X \leq M$. Moreover, we have $\|\overline{p} - x_{\mathrm{opt}}\|_X \leq 2E_n(x_{\mathrm{opt}})$ because we can write $\|\overline{p} - x_{\mathrm{opt}}\|_X \leq \|\overline{p} - P_{X_n}(x_{\mathrm{opt}})\|_X + \|P_{X_n}(x_{\mathrm{opt}}) - x_{\mathrm{opt}}\|_X$, and $\|\overline{p} - P_{X_n}(x_{\mathrm{opt}})\|_X = \|(\lambda - 1)P_{X_n}(x_{\mathrm{opt}})\|_X \leq (1 - \lambda)(M + E_n(x_{\mathrm{opt}})) \leq E_n(x_{\mathrm{opt}})$.

On the other hand, we have for all $i = 1, \ldots, L$ that $|\langle a_i, \overline{p}\rangle - b_i| = |\langle a_i, \overline{p} - x_{\mathrm{opt}}\rangle| \leq \|a_i\|_{X'} \|\overline{p} - x_{\mathrm{opt}}\|_X \leq \|a_i\|_{X'} 2E_n(x_{\mathrm{opt}})$. Therefore, choosing $\epsilon(n) = 2E_n(x_{\mathrm{opt}})\big(\sum_{i=1}^L \|a_i\|_{X'}^q\big)^{1/q}$, we obtain that $\overline{p}$ is feasible for the problem ($\mathrm{P}_{n,\epsilon(n)}$). Similarly, we have $|\langle c, x_{\mathrm{opt}}\rangle - \langle c, \overline{p}\rangle| \le \|c\|_{X'} 2E_n(x_{\mathrm{opt}})$, and we have proved the following lemma:

Lemma 1. *For* $\epsilon(n) = 2E_n(x_{opt})\left(\sum_{i=1}^{L} \|a_i\|_{X'}^q\right)^{1/q}$, *the optimal values of problems* (P) *and* ($\mathrm{P}_{n,\epsilon(n)}$) *satisfy* $P^*_{n,\epsilon(n)} - P^* \le \|c\|_{X'} 2E_n(x_{opt})$.

4.2 Upper bound on $P^*_n - P^*_{n,\epsilon}$

We first introduce a general regularity theorem that compares the optimal value of a problem with linear equality constraints with the optimal value of its relaxation, and then apply it to the specific pair of problems (P_n) and ($\mathrm{P}_{n,\epsilon}$).

Regularity Theorem

Let $(Z, \|.\|_Z)$ and $(Y, \|.\|_Y)$ be two normed vector space, $A : Z \to Y$ be a linear operator, $Q \subset Z$ be a convex bounded closed subset of Z with nonempty interior, $b \in Y$ and $\mathcal{L} = \{z \in Z : Az = b\}$. We denote the distance between a point z and subspace $\mathcal{L}$ by $d(z, \mathcal{L}) = \inf_{y\in\mathcal{L}} \|z - y\|_Z$.

Lemma 2. *Assume that there exists a point* $\hat{z} \in Z$ *such that* $A\hat{z} = b$ *and* $B(\hat{z}, \rho) \subset Q \subset B(\hat{z}, R)$ *for some* $0 < \rho \le R$. *Then, for every point* $z \in Q$ *such that* $d(z, \mathcal{L}) \le \delta$, *there exists* $\tilde{z} \in \mathcal{L} \cap Q$ *such that* $\|z - \tilde{z}\|_Z \le \delta\left(1 + \frac{R}{\rho}\right)$.

Proof. Denote $Q_z = \mathrm{conv}(z, B(\hat{z}, \rho)) \subset Q$. The support function of this set is $\sigma_{Q_z}(s) = \sup_{y\in Q_z}\langle s, y\rangle = \max\{\langle s, z\rangle, \langle s, \hat{z}\rangle + \rho\|s\|_{Z'}\}$. Let π be any element of best approximation of the point z into $\mathcal{L}$. Define $\alpha = \frac{\rho}{\rho+\delta}$ and consider $\tilde{z} = \alpha\pi + (1-\alpha)\hat{z}$. Then we have for any $s \in Z'$ that

$$\begin{aligned}\langle s, \tilde{z}\rangle &= \alpha\langle s, z\rangle + (1-\alpha)\langle s, \hat{z}\rangle + \alpha\langle s, \pi - z\rangle \\ &\le \alpha\langle s, z\rangle + (1-\alpha)\left[\langle s, \hat{z}\rangle + \frac{\alpha\delta}{1-\alpha}\|s\|_{Z'}\right] \\ &= \alpha\langle s, z\rangle + (1-\alpha)\left[\langle s, \hat{z}\rangle + \rho\|s\|_{Z'}\right] \le \sigma_{Q_z}(s)\end{aligned}$$

and hence $\tilde{z} \in Q_z \subset Q$. Since we also have $\tilde{z} \in \mathcal{L}$, it remains to note that

$$\begin{aligned}\|z - \tilde{z}\|_Z &\le \|z - \pi\|_Z + \|\pi - \tilde{z}\|_Z = \delta + (1-\alpha)\|\pi - \hat{z}\|_Z \\ &\le \delta + (1-\alpha)(\|\pi - z\|_Z + \|z - \hat{z}\|_Z) \le \delta + (1-\alpha)(\delta + R) \\ &= \delta\left(1 + \frac{R+\delta}{\rho+\delta}\right) \le \delta\left(1 + \frac{R}{\rho}\right)\end{aligned}$$

□

We consider now the following optimization problem:

$$g^* = \inf_{z\in Z}\langle c, z\rangle \text{ s.t } Az = b \text{ and } z \in Q \tag{G}$$

and its relaxed version

$$g^*_\epsilon = \inf_{z\in Z}\langle c, z\rangle \text{ s.t } \|Az - b\|_Y \le \epsilon \text{ and } z \in Q. \tag{G$_\epsilon$}$$

The following Regularity Theorem links the optimal values of these two problems.

Theorem 2 (Regularity Theorem). *Assume that*

A1. (G_ϵ) *is solvable,*
A2. *there exists* $\hat{z} \in Z$ *s.t* $A\hat{z} = b$ *and* $B(\hat{z}, \rho) \subset Q \subset B(\hat{z}, R)$ *for* $0 < \rho \leq R$,
A3. *the operator* $A : Z \to Y$ *is non degenerate, i.e. there exists a constant* $\sigma > 0$ *such that* $\|Az - b\|_Y \geq \sigma d(z, \mathcal{L})\ \forall z \in Z$.

Then $g^* \geq g^*_\epsilon \geq g^* - \frac{\epsilon \|c\|_{Z'}}{\sigma}\left(1 + \frac{R}{\rho}\right)$.

Proof. The first inequality is evident. For the second one, consider z^*_ϵ, an optimal value of the problem (G_ϵ). Since $d(z^*_\epsilon, \mathcal{L}) \leq \delta := \frac{\epsilon}{\sigma}$, in view of Lemma 2, there exists a point $\tilde{z} \in \mathcal{L} \cap Q$ such that $\|z^*_\epsilon - \tilde{z}\|_Z \leq \delta\left(1 + \frac{R}{\rho}\right)$. Therefore, we can conclude $g^*_\epsilon = \langle c, z^*_\epsilon \rangle = \langle c, \tilde{z} \rangle + \langle c, z^*_\epsilon - \tilde{z} \rangle \geq g^* - \|c\|_{Z'}\, \delta\left(1 + \frac{R}{\rho}\right)$ □

Satisfying the hypotheses of the Regularity Theorem

We want to apply the Regularity Theorem to the pair of problems (P_n) and $(\mathrm{P}_{n,\epsilon})$. First, we note that, as X_n is finite-dimensional, the set $\{x \in X_n : \|x\|_X \leq M, \|Ax - b\|_q \leq \epsilon\}$ is compact. As the functional c is continuous, we conclude that problem $(\mathrm{P}_{n,\epsilon})$ is solvable, i.e. hypothesis A1 is satisfied.

In order to prove hypothesis A2, we assume that there exists $\hat{x} \in X_n$ such that $A\hat{x} = b$ and $\|\hat{x}\|_X < M$ (a kind of Slater condition). If we denote $\mathcal{R}_n = \min_{x \in X_n, Ax = b} \|x\|_X$, $\tilde{x} = \arg\min_{x \in X_n, Ax=b} \|x\|_X$ and $Q_n = \{x \in X_n : \|x\|_X \leq M\} = B_{X_n}(0, M)$, we have : $B_{X_n}(\tilde{x}, M - \mathcal{R}_n) \subset Q_n \subset B_{X_n}(\tilde{x}, 2M)$.

Regarding hypothesis A3, denote $\mathcal{L}_n = \{x \in X_n : Ax = b\}$ and write

$$\begin{aligned} d(x, \mathcal{L}_n) &= \min_{u \in X_n, Au = b} \|x - u\|_X = \min_{\lambda \in \mathbb{R}^n, A^{(n)}\lambda = b} \left\| x - \sum_{i=1}^{n} \lambda_i p_i \right\|_X \\ &= \min_{u \in X_n} \max_{y \in \mathbb{R}^L} \left[\|x - u\|_X + \langle y, -Au + b \rangle\right] \end{aligned}$$

where we defined $[A^{(n)}]_{i,j} = \langle a_i, p_j \rangle$. Since a linearly constrained optimization problem in $\mathbb{R}^n$ with convex objective function always admits a zero duality gap (see e.g. [3]), we have

$$\begin{aligned} d(x, \mathcal{L}_n) &= \max_{y \in \mathbb{R}^L} \min_{u \in X_n} \left[\|x - u\|_X + \langle y, -Au + b \rangle\right] \\ &= \max_{y \in \mathbb{R}^L} \left(\langle y, b - Ax \rangle + \min_{u \in X_n} \|x - u\|_X + \langle y, A(x - u) \rangle \right). \end{aligned}$$

Consider now the Lagrangian dual functional $\gamma(y) = \min_{u \in X_n} \|x - u\|_X + \langle y, A(x - u) \rangle$. If we define $A' : \mathbb{R}^L \to X'$ by $\langle y, Ax \rangle = \langle A'y, x \rangle\ \forall x \in X, \forall y \in \mathbb{R}^L$, we can check that $A'y = \sum_{i=1}^{L} y_i a_i$. Denoting $\|A'y\|_{X',n} = \sup_{w \in X_n} \frac{|\langle A'y, w \rangle|}{\|w\|_X}$, it follows from the definition of the dual norm that $\gamma(y) = 0$ if $\|A'y\|_{X',n} \leq 1$ and $-\infty$ otherwise. We conclude that

$$\begin{aligned} d(x, \mathcal{L}_n) &= \max_{\{y \in \mathbb{R}^L \text{ s.t. } \|A'y\|_{X',n} \le 1\}} \langle y, b - Ax \rangle \\ &\le \max_{\{y \in \mathbb{R}^L \text{ s.t. } \|A'y\|_{X',n} \le 1\}} \|y\|_{q'} \|b - Ax\|_q \ . \end{aligned}$$

Therefore, choosing a $\sigma_n > 0$ such that $\frac{1}{\sigma_n} = \max_{\{y \in \mathbb{R}^L \text{ s.t. } \|A'y\|_{X',n} \le 1\}} \|y\|_{q'}$ ensures degeneracy of A, and we have

Lemma 3. *If $\sigma_n = \min_{\{y \in \mathbb{R}^L, \|y\|_{q'}=1\}} \|A'y\|_{X',n}$ is strictly positive then operator $A : X_n \to \mathbb{R}^L$ is non-degenerate with constant σ_n.*

Remark 1. If X is a Hilbert space and if we work with the Euclidean norm for $\mathbb{R}^L$, we can obtain a more explicit non-degeneracy condition, by identifying all $x' \in X'$ with the corresponding element of X given by the Riesz representation theorem such that X' is identified with X. Suppose $\{p_1, \ldots, p_n\}$ is an orthonormal basis of X_n. Using $A^{(n)}$ as defined above, we have $\sup_{w \in X_n} \frac{|\langle A'y, w \rangle|}{\|w\|_X} = \sup_{\mathbf{w} \in \mathbb{R}^n} \frac{|\langle A^{(n)T} y, \mathbf{w} \rangle|}{\|\mathbf{w}\|_2} = \|A^{(n)T} y\|_2$. Furthermore, if $\cup X_n$ is dense in X, $\|A^{(n)T} y\|_2^2 = \sum_{j=1}^n \left(\langle \sum_{i=1}^L a_i y_i, p_j \rangle \right)^2$ converges to $\|A^T y\|_X^2 = \left\| \sum_{i=1}^L a_i y_i \right\|_X^2 = \lambda_{\min}(AA^T)$ when n tends to infinity. Operator $AA^T : \mathbb{R}^L \to \mathbb{R}^L$ is positive semidefinite and corresponds to a matrix with components $[AA^T]_{i,j} = \langle a_i, a_j \rangle$. It is therefore enough to assume it is nonsingular or, equivalently, the linear independence of all a_i in $X' = X$, to show that there exists N such that for all $n \ge N$, $\sigma_n > 0$.

We are now able to apply the Regularity Theorem to (P_n) and $(\mathrm{P}_{n,\epsilon})$.

Lemma 4. *Assume that*

1. *there exists $\hat{x} \in X_n$ such that $A\hat{x} = b$ and $\|\hat{x}\|_X < M$,*
2. *$\sigma_n = \min_{\{y \in \mathbb{R}^L, \|y\|_{q'}=1\}} \|A'y\|_{X',n} > 0$.*

Then the optimal values of the problems (P_n) and $(\mathrm{P}_{n,\epsilon})$ satisfy for all $\epsilon > 0$ $P_n^* - \frac{\epsilon \|c\|_{X'}}{\sigma_n} \left(1 + \frac{2M}{M - \mathcal{R}_n} \right) \le P_{n,\epsilon}^* \le P_n^*$ with $\mathcal{R}_n = \min_{x \in X_n, Ax = b} \|x\|_X$.

4.3 Convergence result

In order to combine the two bounds we have obtained, we need to assume that hypotheses of Lemmas 1 and 4 are satisfied for some values of n. In fact,

- If there exists N_1 such that $\mathcal{R}_{N_1} < M$ then $\mathcal{R}_n < M$ for all $n \ge N_1$.
- If there exists N_2 such that $\sigma_{N_2} > 0$ then $\sigma_n > 0$ for all $n \ge N_2$.

Therefore, we have proved the following convergence result:

Theorem 3. *Assume that*

1. *the infinite-dimensional problem* (P) *is solvable*

2. there exists N_1 *and* $\hat{x} \in X_{N_1}$ *such that* $A\hat{x} = b$ *and* $\|\hat{x}\|_X < M$
3. there exists N_2 *such that* $\sigma_{N_2} > 0$.

Then we have for all $n \geq N = \max\{N_1, N_2\}$ *that*

$$P^* \leq P_n^* \leq P^* + 2E_n(x_{opt}) \|c\|_{X'} \left(1 + \frac{\left(\sum_{i=1}^{L} \|a_i\|_{X'}^q\right)^{1/q}}{\sigma_n} \left(1 + \frac{2M}{M - \mathcal{R}_n}\right)\right)$$

where x_{opt} *is an optimal solution of* (P) *and* $\mathcal{R}_n = \min_{x \in X_n, Ax=b} \|x\|_X$.

To summarize, we have obtained a convergence result for our polynomial approximation scheme provided that $E_n(x_{\text{opt}})$, the best approximation error of the optimal solution of (P), converges to zero when n goes to infinity, which is is a natural condition from the practical point of view. This holds for example if the linear subspace $\text{span}\{p_1, \ldots, p_n, \ldots\} = \cup_n X_n$ is dense in X.

5 Specific classes of infinite-dimensional problems

To conclude, we provide a few examples of specific functional spaces X and comment on their solvability and the expected rate of convergence described by Theorem 3.

X is the Lebesgue space L^q

These functional spaces are suitable for use in the supply problems considered in Section 2. Let Ω be a domain of $\mathbb{R}^N$ and $1 \leq q \leq \infty$. Let X be the Lebesgue space $L^q(\Omega) = \{u \in \mathcal{M}(\Omega) : \int_\Omega |u(t)|^q \, dt < +\infty\}$ with norm $\|u\|_X = \|u\|_q = \left(\int_\Omega |u(t)|^q \, dt\right)^{1/q}$ in the case $1 \leq q < \infty$, and $\|u\|_X = \|u\|_\infty =$ ess $\sup_{t \in \Omega} |u(t)|$ when $q = \infty$. We take as linear and continuous functionals $c : L^q(\Omega) \to \mathbb{R},\ u \to \int_\Omega u(t)\gamma(t)dt$ and $a_i : L^q(\Omega) \to \mathbb{R},\ u \to \int_\Omega u(t)\alpha_i(t)dt$ where γ and $\alpha_i \in L^{q'}(\Omega)$ for all $i = 1, \ldots, L$.

Concerning the solvability of this problem, note that $L^q(\Omega)$ is reflexive for all $1 < q < \infty$ and that $L^\infty(\Omega) = (L^1(\Omega))'$ where $L^1(\Omega)$ is separable ([1]). Therefore, we can conclude that the infinite-dimensional problem has at least one optimal solution for all $1 < q \leq \infty$. Similar results can be otained if we consider the sequence space l^q.

If Ω is a bounded interval $[a, b]$ and $X_n = \text{span}\{1, t, \ldots, t^{n-1}\}$, we have the following well-known results about the convergence of the best polynomial approximation error of a function $u \in X$, see e.g. [11, 5]:

- $E_n(u)_q \to 0$ iff $u \in L^q([a, b])$ for all $1 \leq q < \infty$
- $E_n(u)_\infty \to 0$ iff $u \in \mathcal{C}([a, b])$
- $E_n(u)_q = O(\frac{1}{n^r})$ if $u \in \mathcal{C}^{r-1,r-1}([a, b])$ for all $1 < q \leq \infty$

where $E_n(u)_q = \inf_{v \in X_n} \|u - v\|_q$ and $\mathcal{C}^{k,r} = \{u \in \mathcal{C}^k([a, b])$ s.t $u^{(r)}$ is Lipschitz continuous $\}$ with $r \leq k$.

Recall that these quantities, that describe the best approximation error of the optimal solution of (P), have a direct influence on the convergence rate of P_n^* to P^* (cf. Theorem 3).

X is the Sobolev space $W^{k,q}$

If we want to include derivatives of our variable in the constraints or in the objective, we need to work in Sobolev spaces. Let Ω be a domain of $\mathbb{R}^N$, $1 \leq q \leq \infty$ and $k \in \mathbb{N}$. For all multi-indices $(\beta_1, \ldots, \beta_N) \in \mathbb{N}^N$, we note $|\beta| = \sum_{i=1}^N \beta_i$ and $D^\beta u = \frac{\partial^{|\beta|} u}{\partial t_1^{\beta_1} \ldots \partial t_N^{\beta_N}}$ in the weak sense. We choose for X the Sobolev space $W^{k,q}(\Omega) = \{u \in \mathcal{M}(\Omega) : D^\beta u \in L^q(\Omega) \quad \forall 0 \leq |\beta| \leq k\}$ with the norm $\|u\|_X = \|u\|_{k,q} = \left(\sum_{0 \leq |\beta| \leq k} \left\|D^\beta u\right\|_q^q\right)^{1/q}$ in the case $1 \leq q < \infty$ and $\|u\|_{k,\infty} = \max_{0 \leq |\beta| \leq k} \left\|D^\beta u\right\|_\infty$ when $q = \infty$. Our linear and continuous functionals are $c : W^{k,q}(\Omega) \to \mathbb{R},\ u \to \sum_{0 \leq |\beta| \leq k} \int_\Omega D^\beta u(t) \gamma_\beta(t) dt$ and $a_i : W^{k,q}(\Omega) \to \mathbb{R},\ u \to \sum_{0 \leq |\beta| \leq k} \int_\Omega D^\beta u(t) \alpha_{i,\beta}(t) dt$ where γ_β and $\alpha_{i,\beta} \in L^{q'}(\Omega)$ for all $i = 1, \ldots, L$ and for all $0 \leq |\beta| \leq k$.

Since the space $W^{k,q}$ is reflexive for all $k \in \mathbb{N}$ and for all $1 < q < \infty$ [1], existence of an optimal solution to (P) is guaranteed. Furthermore, when Ω is a bounded interval $[a, b]$, it is well-known that the polynomials are dense in the Sobolev space $W^{k,q}([a, b])$ for all $k \in \mathbb{N}$ and for all $1 \leq q < \infty$. Therefore, Theorem 3 guarantees convergence of the polynomial approximation scheme in this case.

References

1. R.A. Adams (2003) Sobolev Spaces: 2nd edition. Academic Press.
2. E.J. Anderson and P. Nash (1987) Linear Programming in infinite-dimensional spaces. Wiley.
3. S. Boyd and L. Vandenbergh (2009) Convex Optimization (7th printing with corrections). Cambdrige University Press.
4. E.W. Cheney (1982) Introduction to Approximation Theory (second edition). AMS Chelsea Publishing.
5. Z. Ditzian and V. Totik (1987) Moduli of smoothness. Springer-Verlag.
6. A. Gil, J. Segura and N.M. Temme (2007) Numerical Methods for special functions. SIAM.
7. M. Hinze, R. Pinnau, M. Ulbrich and S. Ulbrich (2009) Optimization with PDE constraints. Springer-Verlag.
8. D.G. Luenberger (1969) Optimization by vector space methods. Wiley.
9. Y.E. Nesterov and A.S. Nemirovskii (1994) Interior-Point Polynomial Algorithms in Convex Programming. SIAM.
10. Y.E. Nesterov (2000) Squared functional systems and optimization problems. In High performance Optimization, pages 405-440. Kluwer Academic Publishers.
11. A.F. Timan (1963) Theory of approximation of functions of a real variable. Pergamon Press.
12. E. Zeidler (1985) Nonlinear Functional Analysis and its Applications (Part 3): Variational Methods and Optimization. Springer-Verlag.