SHORT-MEMORY LINEAR PROCESSES AND ECONOMETRIC APPLICATIONS

SHORT-MEMORY LINEAR PROCESSES AND ECONOMETRIC APPLICATIONS

KAIRAT T. MYNBAEV

International School of Economics
Kazakh-British Technical University
Almaty, Kazakhstan

WILEY

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data:

Mynbaev, K. T. (Kairat Turysbekovich)
Short-memory linear processes and econometric applications / Kairat T. Mynbaev.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-92419-8
1. Linear programming. 2. Econometric models. 3. Regression analysis.
4. Probabilities. I. Title.
T57.74.M98 2011
519.7′2—dc22

2010040947

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

To my teacher Mukhtarbai Otelbaev,
from whom I learnt the best I know.

CONTENTS

LIST OF TABLES

PREFACE

1 RED LIGHT

There are no new econometric models in this book. You will not find real-life applications or tests of economic theories either.

2 GREEN LIGHT

The book concentrates on the methodology of asymptotic theory in econometrics. Specifically, central limit theorems (CLTs) for weighted sums of short-memory processes are obtained. They are applied to several well-known econometric models to demonstrate how their asymptotic behavior can be studied, what kind of assumptions are (in)appropriate and how probabilistic convergence statements are applied. Currently, no monographs or textbooks are devoted specifically to econometric models with deterministic regressors. The field is considered rather narrow by some specialists because the first thing they think about is polynomial trends. Indeed, polynomial trends are not widely used in econometrics. However, some other types of regressors fall into the classes of deterministic regressors considered in the literature; for example, some spatial matrices and seasonal dummies. This makes deterministic explanatory variables more important than commonly thought. Besides, on the level of CLTs deterministic weights are of interest in themselves. There is a monograph by Taylor (1978) devoted exclusively to such theorems.

3 THE ESSENCE

By and large, CLTs here are based on only one global idea: how sequences of discrete objects (vectors and matrices) can be approximated with functions of a continuous argument (defined on the segment [0, 1]). Stated in this general way, the idea is as old as calculus. The novelty here consists in application of the idea to weighted sums of linear processes

$$\sum_{t=1}^{n} w_{nt} u_t, \tag{0.1}$$

where $w_n = (w_{n1}, \ldots, w_{nn})$, $n = 1, 2, \ldots$ is a sequence of deterministic vector weights and

$$u_t = \sum_{j=-\infty}^{\infty} \psi_{t-j} e_j, \quad t = 0, \pm 1, \pm 2, \ldots \tag{0.2}$$

is a short-memory linear process. Anybody with a little experience in probabilities, statistics, and econometrics can confirm that statements on convergence in distribution of such weighted sums have many applications. As it turned out, the main difficulties in proving precise CLTs lay in the theory of functions. Hence, attempts to obtain general CLTs for sums of type Eq. (0.1) by researchers with backgrounds other than the theory of functions yielded results less satisfactory than those published in my paper (Mynbaev, 2001) on L_p-approximable sequences.

My interest in CLTs for Eq. (0.1) arose from the necessities of regression analysis. In the asymptotic theory of regressions with deterministic regressors, sequences of regressors can be approximated by functions of a continuous argument. The structure of the corresponding estimators allows for application of CLTs for Eq. (0.1). As I was developing applications, I needed various additional properties of L_p-approximable sequences. They are distributed throughout the book and, taken together, constitute a complete toolkit accompanying the main CLTs.

In the econometrics context, two other definitions of deterministic regressors are suggested in the literature. A purely algebraic definition (based on recursion) was proposed by Johansen (2000) and developed further by Nielsen (2005) to study strong consistency of ordinary least squares (OLS) estimators. Nielsen's result, given in Chapter 8, shows that such regressors are asymptotically polynomial functions multiplied by oscillating (trigonometric) functions. The Johansen–Nielsen approach and L_p-approximability complement each other.

Phillips (2007) has defined regressors in terms of slowly varying (SV) functions (which is a functional–theoretical construction). Slow variation is a limit property at infinity and, in general, has nothing to do with L_p-approximability, which is a limit property distributed over the segment [0, 1]. However, special sequences arising from SV functions in the regression context are all L_p-approximable.

4 STANDING PROBLEMS

About half of the results contained in the book were obtained after I started writing it. The theory has grown to the extent that no single person can embrace all the ramifications.

1. Linear processes (0.2), depending on the rate at which the numbers ψ_j vanish at infinity, are classified as follows. Processes for which

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty \tag{0.3}$$

are called short-memory processes. Processes for which the series in Eq. (0.3) diverges, but

$$\sum_{j=-\infty}^{\infty} \psi_j^2 < \infty$$

are called long-memory processes. My CLT for weighted sums Eq. (0.1) holds in the case of short-memory processes. The existing CLTs for long-memory ones, as deep as they are, leave some questions open.

2. The main advantage of representing sequences of vectors with the help of functions of a continuous argument is that the limit expressions in asymptotic distributions involve integrals of those functions. Thus, they are amenable to further analysis, which I call analysis at infinity. For this reason alone, when my definition of L_p-approximability does not fit practical situations (and there is at least one, in spatial econometrics), developing a more suitable definition may be better than relinquishing the concept altogether.
3. The name of the book reflects its coverage rather than its potential. There are two important directions in which it can be extended. One is nonparametric and nonlinear estimation, where even my CLT will suffice for the beginning. Another is the case of stochastic regressors. In this case Anderson and Kunitomo (1992) impose conditions on separate parts of the OLS estimator that allow them to prove its convergence. As an alternative, I would embed enough structure in the stochastic regressors to be able to derive convergence of separate parts of the OLS estimator. The structure entailed by L_p-approximability in the deterministic case may guide the choice for the stochastic case.

5 REVIEW BY CHAPTERS

Chapter 1 is a collection of general ideas and preliminaries from probability theory and functional analysis. It also contains a discussion of L_p-approximability and its advantages. The first nontrivial application is to the convergence in distribution of the fitted value for the linear regression. This convergence looks to some econometricians so incredible that an anonymous referee of *Econometric Theory* said that my paper was “full of mistakes.” Naturally, the paper was rejected and, not so naturally, the result was not published in journals. Thus this book was written. The discussion of issues related to normalization of regressors draws from folklore and should be in the core of any course on asymptotic theory in econometrics.

Chapter 2 covers the nonstochastic part of my paper (Mynbaev, 2001). Readers with taste for mathematical precision will find it illuminating that L_p-approximability (which relates sequences of vectors to functions defined on [0, 1]) can be characterized intrinsically (in terms of sequences of vectors themselves). This is evidence of a well-balanced definition. On a more practical note, such results and their by-products make sure that the ensuing CLTs are the most precise and general.

The main CLTs are proved in Chapter 3, which is based on Mynbaev (2001) but I would like to acknowledge the influence of Nabeya and Tanaka (1990) who paved the way to treating convergence of quadratic forms. This is where the theory of integral operators is needed and introduced first.

There are many CLTs and weak laws of large numbers (WLLN) out there. The reader will notice that when the innovations e_j in linear processes [Eq. (0.2)] are martingale differences (m.d.'s), the McLeish CLT (McLeish, 1974) and the Chow-Davidson WLLN (Davidson, 1994) are absolutely sufficient for the purposes of Chapter 3. I dare to suggest trying these tools first in all other problems with linear processes involving m.d.'s.

Serious applications (to static models) start in Chapter 4. Phillips (2007) developed a nice scheme of investigating asymptotic properties of regressions with regressors such as $\log s$, $\log(\log s)$, their reciprocals and so on. Chapter 4 follows this scheme, while the underlying central limit results are derived from my CLT. This is possible because Phillips' specification of the weights in Eq. (0.1) is a special case of L_p-approximable sequences. One of the methodological conclusions of this chapter is that direct derivation of a CLT given in Chapter 3 is better than recourse to Brownian motion used by Phillips.

Chapter 5 demonstrates what can be done with the help of L_p-approximability in the theory of spatial models. This research started with a joint paper (Mynbaev and Ullah, 2008) in which we showed that the OLS estimator for a purely spatial model is not asymptotically normal. In Mynbaev (2010), this result is extended to a mixed spatial model. Spatial models are peculiar in many respects, a full discussion of which would be too technical for a preface. It is worth stating here only the most general methodological conclusion. When studying the asymptotic behavior of a new model, never presume it is of a certain class. Otherwise, you will be bound to use specific techniques that will take you to a particular result so you will not see the general picture.

In the 1980s, Lai and Wei in a series of papers (Lai and Wei, 1982, 1983a, 1983b, 1985) obtained outstanding results on strong consistency of the OLS estimator for the linear model, with and without autoregressive terms. Reading those papers is a thankless task because the solution to a large problem is divided into publishable articles and the times of publication of the articles are not the best reflection of the logic of the solution. Chapter 6 is an attempt to expound Lai and Wei's theory coherently.

Chapter 7 contains a treatment of two nonlinear estimators: nonlinear least squares (NLS) and maximum likelihood (ML). The choice of the models is explained by the fact that in both cases the explanatory variables are deterministic. The first part of the chapter covers the Phillips (2007) result for the model $y_s = \beta s^{\gamma} + u_s$. The second part is my extension to unbounded explanatory variables of the approach to binary logit models suggested by Gouriéroux and Monfort (1981).

Finally, Chapter 8 contains a study of algebraic properties of L_p-approximable sequences of matrix-valued functions and a study of a different type of deterministic trends from Nielsen (2005). The applications to vector autoregressions (VARs) with deterministic trends are left out.

6 EXPOSITION

The book is analytical in nature, meaning that there is a lot of formula manipulation. Most calculations are detailed so they can be followed without a pen and paper. To simplify the reader's job, all meaningful parts of proofs are given in separate statements. Because of this, some proofs look longer than they are. Commuters who need to do their reading in buses and trains will benefit from such exposition.

Only the core theoretical results are collected in Chapters 2 and 3. All others are given immediately before they are applied (including some CLTs). Thus, application-specific properties of L_p-approximable sequences, as well as parts of the theory of integral operators, are scattered throughout the book.

If someone were to lecture using this book, I have imagined how clumsy it would be to say, "Let us recall the function defined by Eq. (9) in Lecture 3." For this reason I have tried to give names not only to final statements, but also to auxiliary objects, such as lemmas, functions and operators. In most cases the names reflect the roles performed by such objects. Thus, you will see βad and γood coefficients, a chain product, annihilation lemma, balancer, cutter and the like. However, in a couple of cases descriptive names would be too long, and the names I give reflect the look, not the role. There is a proξy and proXy, an awkward aggregate, genie (because of G_n) and so on.

No subsection contains more than one statement. Therefore statements are referred to by the section they are in. Thus, Lemma 3.1.2 means the statement from subsection 3.1.2, even though the name 'Lemma' may not be there. Equation numbering follows the Wiley standard: Eq. (7.1) means equation 1 from Chapter 7. To make the book self-contained, most preliminaries are given in the book. All calculations are detailed with extensive cross-referencing.

7 SUGGESTIONS FOR READING

The variety and depth of mathematical theories used by econometricians can be a serious obstacle for novices. Davidson (1994) has done an excellent job in gathering in one place the required minimum, from measure theory to stochastic processes. For me, this is the most important book I have read in the past 10 years, and I recommend it for preliminary or concurrent reading.

A partial excuse for the limited coverage of the existing literature is that during the four years that I was working on the book I did not receive any support, financial or otherwise, and did not have access to a good library, except when I traveled to international conferences. At the final stage, when the book was in production, I received useful references from some colleagues. Regarding weighted sums and their applications in econometrics, Jonathan B. Hill suggests reading Čížek (2008), Goldie and Smith (1987), Hahn *et al.* (1987), Hill (2010, 2011) and references therein. Jan Mielniczuk, who contributed a lot to the theory of long-memory processes not covered here, proposes reading Wu (2005) and Wu and Min (2005) for the most recent developments in the area. M. H. Pesaran was kind enough to provide references Chudik *et al.* (2010), Holly *et al.* (2008) and Pesaran and Chudik (2010) for spatial models, vector autoregressions and panel data models.

Personally, I find nothing more gratifying than reading applied econometric papers because they abound in new ideas. Sometimes they also show how things should not be done. See more about this in Chapter 1.

Kairat T. Mynbaev
Almaty, Kazakhstan

ACKNOWLEDGMENTS

I am grateful to Carlos Martins Filho for his encouragement for this book and many other projects. The folks at John Wiley & Sons have been highly efficient in preparing the book for publication. The production process surely involved many people of whom I would like to especially thank Susanne Steitz-Filler, Christine Punzo, Jacqueline Palmieri, and Nick Barber (Books Manager, Techset Composition Ltd).

CHAPTER 1

INTRODUCTION TO OPERATORS, PROBABILITIES AND THE LINEAR MODEL

THIS CHAPTER has a little bit of everything: normed and Hilbert spaces, linear operators, probabilities, including conditional expectations and different modes of convergence, and matrix algebra. Introduction to the OLS method is given along with a discussion of methodological issues, such as the choice of the format of the convergence statement, choice of the conditions sufficient for convergence and the use of L_2-approximability. The exposition presumes that the reader is versed more in the theory of probabilities than in functional analysis.

1.1 LINEAR SPACES

In this book basic notions of functional analysis are used more frequently than in most other econometric books. Here I explain these notions the way I understand them—omitting some formalities and emphasizing the intuition.

1.1.1 Linear Spaces

The Euclidean space $\mathbb{R}^n$ is a good point of departure when introducing linear spaces. An element $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ is called a *vector*. Two vectors x, y can be added coordinate by coordinate to obtain a new vector

$$x + y = (x_1 + y_1, \ldots, x_n + y_n). \tag{1.1}$$

A vector x can be multiplied by a number $a \in \mathbb{R}$, giving $ax = (ax_1, \ldots, ax_n)$. By combining these two operations we can form expressions like $ax + by$ or, more generally,

$$a_1 x^{(1)} + \cdots + a_m x^{(m)} \tag{1.2}$$

where $a_1, \ldots, a_n$ are numbers and $x^{(1)}, \ldots, x^{(m)}$ are vectors. Expression (1.2) is called a *linear combination of vectors* $x^{(1)}, \ldots, x^{(m)}$ with coefficients $a_1, \ldots, a_n$. Generally, multiplication of vectors is not defined.

Short-Memory Linear Processes and Econometric Applications. Kairat T. Mynbaev

Here we observe the major difference between $\mathbb{R}$ and $\mathbb{R}^n$. In $\mathbb{R}$ both summation $a+b$ and multiplication ab can be performed. In $\mathbb{R}^n$ we can add two vectors, but to multiply them we use elements of another set – the set of real numbers (or scalars) $\mathbb{R}$.

Generalizing upon this situation we obtain abstract *linear* (or *vector*) *spaces*. The elements x, y of a linear space L are called *vectors*. They can be added to give another vector $x+y$. Summation is defined axiomatically and, in general, there is no coordinate representation of type (1.1) for summation. A vector x can be multiplied by a scalar $a \in \mathbb{R}$. As in $\mathbb{R}^n$, we can form linear combinations [Eq. (1.2)].

The generalization is pretty straightforward, so what's the big deal? You see, in functional analysis complex objects, such as functions and operators, are considered vectors or points in some space. Here is an example. Denote $C[0, 1]$ the set of continuous functions on the segment $[0, 1]$. The sum of two functions $F, G \in C[0, 1]$ is defined as the function $F+G$ with values $(F+G)(t) = F(t) + G(t)$, $t \in [0, 1]$ [this is an analog of Eq. (1.1)]. Continuity of F, G implies continuity of their sum and of the product aF, for a a scalar, so $C[0, 1]$ is a linear space.

1.1.2 Subspaces of Linear Spaces

A subset L_1 of a linear space L is called its *linear subspace* (or just a subspace, for simplicity) if all linear combinations $ax+by$ of any elements $x, y \in L_1$ belong to L_1. Obviously, the set $\{0\}$ and L itself are subspaces of L, called trivial subspaces. For example, in $\mathbb{R}^n$ the set $L_1 = \{x : c_1x_1 + \cdots + c_nx_n = 0\}$ is a subspace because if $x, y \in L_1$, then $c_1(ax_1 + by_1) + \cdots + c_n(ax_n + by_n) = 0$. Thus, in $\mathbb{R}^3$ the usual straight lines and two-dimensional (2-D) planes containing the origin are subspaces. All intuition we get from our day-to-day experience with the space we live in applies to subspaces. Geometrically, summation $x+y$ is performed by the parallelogram rule. Multiplying x by a number $a \neq 0$ we obtain a vector ax of either the same ($a > 0$) or opposite ($a < 0$) direction. Multiplying x by all real numbers, we obtain a straight line $\{ax : a \in \mathbb{R}\}$ passing through the origin and parallel to x. This is a particular situation in which it may be convenient to call x a point rather than a vector. Then the previous sentence sounds like this: multiplying x by all real numbers we get a straight line passing through the origin and the given point x.

For a given $x_1, \ldots, x_n$ its *linear span* $\mathfrak{M}$ is, by definition, the least linear space of L containing those points. In the case $n = 2$ it can be constructed as follows. Draw a straight line $L_1 = \{ax_1 : a \in \mathbb{R}\}$ through the origin and x_1 and another straight line $L_2 = \{ax_2 : a \in \mathbb{R}\}$ through the origin and x_2. Then form $\mathfrak{M}$ by adding elements of L_1 and L_2 using the parallelogram rule: $\mathfrak{M} = \{x + y : x \in L_1, y \in L_2\}$.

1.1.3 Linear Independence

Vectors $x_1, \ldots, x_n$ are *linearly independent* if the linear combination $c_1x_1 + \cdots + c_nx_n$ can be null only when all coefficients are null.

EXAMPLE 1.1. Denote by $e_j = (0, \ldots, 0, 1, 0, \ldots, 0)$ (unity in the jth place) the jth *unit vector* in $\mathbb{R}^n$. From the definition of vector operations in $\mathbb{R}^n$ we see that

$c_1e_1 + \cdots + c_ne_n = (c_1, \ldots, c_n)$. Hence, the equation $c_1e_1 + \cdots + c_ne_n = 0$ implies equality of all coefficients to zero and the unit vectors are linearly independent.

If in a linear space L there exist vectors $x_1, \ldots, x_n$ such that

1. $x_1, \ldots, x_n$ are linearly independent and
2. any other vector $x \in L$ is a linear combination of $x_1, \ldots, x_n$,

then L is called *n-dimensional* and the system $\{x_1, \ldots, x_n\}$ is called its *basis*. If, on the other hand, for any natural n, L contains n linearly independent vectors, then L is called *infinite-dimensional*.

EXAMPLE 1.2. The unit vectors in $\mathbb{R}^n$ form a basis because they are linearly independent and for any $x \in \mathbb{R}^n$ we can write $x = (x_1, \ldots, x_n) = x_1e_1 + \cdots + x_ne_n$.

EXAMPLE 1.3. $C[0, 1]$ is infinite-dimensional. Consider monomials $x_j(t) = t^j$, $j = 0, \ldots, n$. By the main theorem of algebra, the equation $c_0x_0(t) + \cdots + c_nx_n(t) = 0$ with nonzero coefficients can have at most n roots. Hence, if $c_0x_0(t) + \cdots + c_nx_n(t)$ is identically zero on $[0, 1]$, the coefficients must be zero, so these monomials are linearly independent.

Functional analysis deals mainly with infinite-dimensional spaces. Together with the desire to do without coordinate representations of vectors this fact has led to the development of very powerful methods.

1.2 NORMED SPACES

1.2.1 Normed Spaces

The Pythagorean theorem gives rise to the Euclidean distance

$$\text{dist}(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \tag{1.3}$$

between points $x, y \in \mathbb{R}^n$. In an abstract situation, we can first axiomatically define the distance $\text{dist}(x, 0)$ from x to the origin and then the distance between any two points will be $\text{dist}(x, y) = \text{dist}(x - y, 0)$ (this looks like tautology, but programmers use such definitions all the time). $\text{dist}(x, 0)$ is denoted $\|x\|$ and is called a norm.

Let X be a linear space. A real-valued function $\|\cdot\|$ defined on X is called a *norm* if

1. $\|x\| \geq 0$ (*nonnegativity*),
2. $\|ax\| = |a|\|x\|$ for all numbers a and vectors x (*homogeneity*),
3. $\|x + y\| \leq \|x\| + \|y\|$ (*triangle inequality*) and
4. $\|x\| = 0$ implies $x = 0$ (*nondegeneracy*).

By homogeneity the norm of the null vector is zero:

$$\left\| \underset{\text{(vector)}}{0} \right\| = \left\| \underset{\text{(number)}}{0} \cdot \underset{\text{(vector)}}{0} \right\| = |0|\|0\| = 0.$$

Nondegeneracy makes sure that the null vector is the only vector whose norm is zero. If we omit the nondegeneracy requirement, the result is the definition of a *seminorm*.

Distance measurement is another context in which points and vectors can be used interchangeably. $\|x\|$ is a length of the vector x and a distance from point x to the origin.

In this book, the way norms are used for bounding various quantities is clear from the next two definitions. Let $\{X_i\}$ be a nested sequence of normed spaces, $X_1 \subseteq X_2 \subseteq \ldots$. Take one element from each of these spaces, $x_i \in X_i$. We say that $\{x_i\}$ is a *bounded* sequence if $\sup_i \|x_i\|_{X_i} < \infty$ and *vanishing* if $\|x_i\|_{X_i} \to 0$.

1.2.2 Convergence in Normed Spaces

A linear space X provided with a norm $\|\cdot\|$ is denoted $(X, \|\cdot\|)$. This is often simplified to X. We say that a sequence $\{x_n\}$ *converges* to x if $\|x_n - x\| \to 0$. In this case we write $\lim x_n = x$.

Lemma

(i) Vector operations are continuous: if $\lim x_n = x$, $\lim y_n = y$ *and* $\lim a_n = a$, *then* $\lim a_n x_n = ax$, $\lim(x_n + y_n) = \lim x_n + \lim y_n$.

(ii) If $\lim x_n = x$, *then* $\lim \|x_n\| = \|x\|$ *(a norm is continuous in the topology it induces).*

Proof.

(i) Applying the triangle inequality and homogeneity,

$$\begin{aligned}\|a_n x_n - ax\| &\le \|(a_n - a)x\| + \|a_n(x_n - x)\| \\ &= |a_n - a|\|x\| + \|a_n\|\|x_n - x\| \to 0.\end{aligned}$$

Here we remember that convergence of the sequence $\{a_n\}$ implies its boundedness: $\sup|a_n| < \infty$.

(ii) Let us prove that

$$\big|\|x\| - \|y\|\big| \le \|x - y\|. \tag{1.4}$$

The proof is modeled on a similar result for absolute values. By the triangle inequality, $\|x\| \le \|x - y\| + \|y\|$ and $\|x\| - \|y\| \le \|x - y\|$. Changing the places of x and y and using homogeneity we get $\|y\| - \|x\| \le \|y - x\| = \|x - y\|$. The latter two inequalities imply Eq. (1.4).

Equation (1.4) yields continuity of the norm: $|\|x_n\| - \|x\|| \le \|x_n - x\| \to 0$. ■

We say that $\{x_n\}$ is a *Cauchy sequence* if $\lim_{n,m\to\infty}(x_n - x_m) = 0$. If $\{x_n\}$ converges to x, then it is a Cauchy sequence: $\|x_n - x_m\| \leq \|x_n - x\| + \|x - x_m\| \to 0$. If the converse is true (that is, every Cauchy sequence converges), then the space is called *complete*. All normed spaces considered in this book are complete, which ensures the existence of limits of Cauchy sequences.

1.2.3 Spaces l_p

A norm more general than (1.3) is obtained by replacing the index 2 by an arbitrary number $p \in [1, \infty)$. In other words, in $\mathbb{R}^n$ the function

$$\|x\|_p = \left(\sum_i |x_i|^p\right)^{1/p} \tag{1.5}$$

satisfies all axioms of a norm. For $p = \infty$, definition (1.5) is completed with

$$\|x\|_\infty = \sup_i |x_i| \tag{1.6}$$

because $\lim_{p\to\infty} \|x\|_p = \|x\|_\infty$. $\mathbb{R}^n$ provided with the norm $\|\cdot\|_p$ is denoted $\mathbb{R}^n_p$ $(1 \leq p \leq \infty)$.

The most immediate generalization of $\mathbb{R}^n_p$ is the *space* l_p of infinite sequences of numbers $x = (x_1, x_2, \ldots)$ that have a finite norm $\|x\|_p$ [defined by Eqs. (1.5) or (1.6), where i runs over the set of naturals $\mathbb{N}$]. More generally, the set of indices $I = \{i\}$ in Eq. (1.5) or Eq. (1.6) may depend on the context. In addition to $\mathbb{R}^n_p$ we use $\mathbb{M}_p$ (the set of matrices of all sizes).

The jth *unit vector* in l_p is an infinite sequence $e_j = (0, \ldots, 0, 1, 0, \ldots)$ with unity in the jth place and 0 in all others. It is immediate that the unit vectors are linearly independent and l_p is infinite-dimensional.

1.2.4 Inequalities in l_p

The triangle inequality in l_p $\|x + y\|_p \leq \|x\|_p + \|y\|_p$ is called the *Minkowski inequality*. Its proof can be found in many texts, which is not true with respect to another, less known, property that is natural to call *monotonicity* of l_p norms:

$$\|x\|_p \leq \|x\|_q \quad \text{for all } 1 \leq q \leq p \leq \infty. \tag{1.7}$$

If $x = 0$, there is nothing to prove. If $x \neq 0$, the general case can be reduced to the case $\|x\|_q = 1$ by considering the normalized vector $x/\|x\|_q$. $\|x\|_q = 1$ implies $|x_i| \leq 1$ for all i. Hence, if $p < \infty$, we have

$$\|x\|_p = \left(\sum_i |x_i|^p\right)^{1/p} \leq \left(\sum_i |x_i|^q\right)^{1/p} = \left(\sum_i |x_i|^q\right)^{1/q} = \|x\|_q\,.$$

If $p = \infty$, the inequality $\sup_i |x_i| \leq \|x\|_q$ is obvious.