Gram-Schmidt Orthogonalization Process

P. Sam Johnson

November 16, 2014

Definition

Let V be an inner product space, $x, y \in V$. Let A, B be subsets of V.

$\langle x, y \rangle = 0$ *(we write $x \perp y$)*	*x and y are* **orthogonal** *to each other*
$x \perp y$ for every pair of distinct vectors x, y in A	*A is* **orthogonal**
A is orthogonal and every vector in A has norm 1	*A is* **orthonormal**
every vector in A is orthogonal to every vector in B	*A is* **orthogonal** *to B*

Definition

Let S be a subspace of an inner product space. We say that B is an **orthogonal basis** *(resp. an* **orthonormal basis***) of S if B is a basis of S and B is an orthonormal (resp. an orthonormal) set.*

Theorem

Every inner product space has an orthogonal (orthonormal) basis.

Proof. Start by selecting any nonzero vector v_1 in V. If V contains a nonzero vector v_2 that is orthogonal to v_1, put it in the basis. If V contains a nonzero vector v_3 that is orthogonal to v_1 and v_2, put it in the basis.

Proceed in this way. The chosen points $v_1, v_2, \ldots$ will be mutually orthogonal. The generated set is an orthogonal set, which is also a linearly independent. Thus, if V is n dimensional, the selection process certainly must stop after n steps.

If each vector v_i is normalized, then the set is an **orthonormal** basis for V. Normalizing a vector v means replacing v by $v/\|v\|$. The norm of a vector is derived from the inner product : $\|x\| = \sqrt{\langle x, x \rangle}$.

A concrete realilzation of a process similar to the one just described is the **Gram-Schmidt process**. It operates in any finite dimensional inner product space and produces an orthonormal basis.

Little information about Erhard Schmidt

Erhard Schmidt (1876-1959) was another important mathematician who serves as a professor of mathematics in several German universities. His advisor was David Hilbert (who formulated the theory of Hilbert spaces). Schmidt became an expert in the eigen functions that arise in the study of integral equations and partial differntial equations, and he was one of the first to make use of infinite dimensional vector spaces in his work.

Erhard Schmidt

He introduced the notation $\|.\|$ for the magnitude of a vector, $\langle x, y \rangle$ for the inner product. He proved the Phythagorean theorem in abstract inner product spaces and many other results in this subject while it was in its infancy and new to almost all mathematicians. In a 1907 paper, Schmidt described what is now called the **Gram-Schmidt process**.

Little information about Jorgen Pedersen Gram

Jorgen Pedersen Gram (1850-1916) published his first important mathematical paper while still a university student! Rather than teaching mathematics at a university he became a research mathematician employed by an insurance company. He published papers, gave lectures, and won awards for his mathematical research. At the age of 65, Gram was killed after being struck by a bicycle.

Jorgen Pedersen Gram

Gram-Schmidt Algorithm

Suppose $\{v_1, v_2, \ldots, v_n\}$ is a basis of an inner product space V. For the frst step, we define w_1 to be the normalized version of v_1 ; that is, $w_1 = v_1/\|v_1\|$.

For an inductive definition, suppose that we have constructed an orthonormal system $w_1, w_2, \ldots, w_{k-1}$ whose span is the same as span $\mathrm{span}\{v_1, v_2, \ldots, v_{k-1}\}$. To get w_k, subtract from v_k its projection on the span of $\{w_1, w_2, \ldots, w_{k-1}\}$, and then normallize it.

The formula for this process is

$$w_k = \frac{v_k - \sum_{j=1}^{k-1} \langle v_k, w_j \rangle w_j}{\| v_k - \sum_{j=1}^{k-1} \langle v_k, w_j \rangle w_j \|} \quad (k = 2, 3, \ldots, n).$$

In this algorithm, the vectors are normalized as we go along. The new basis has the property that for each $k \leq n$,

$$\mathrm{span}\{w_1, w_2, \ldots, w_k\} = \mathrm{span}\{v_1, v_2, \ldots, v_k\}.$$

Unnormalized Gram-Schmidt Algorith

Theorem

Let $\{v_1, v_2, \ldots, v_n\}$ *be a basis of a subspace* S *of an inner product space* V. *Define* $z_1, z_2, \ldots, z_k, \ldots, z_n$ *inductively by :*

$$z_k = v_k - \sum_{j=1}^{k-1} \frac{\langle v_k, z_j \rangle}{\langle z_j, z_j \rangle} z_j \quad (k = 1, 2, \ldots, n).$$

Then $z_1, z_2, \ldots, z_n$ *is an* **orthogonal basis** *of* S.
An orthonormal basis of V *can be obtained by normalizing the* z_i*'s.*

Starting from any basis of an inner product space V, we can construct an orthonormal basis by the Gram-Schmidt process: **Every finite-dimensional inner product space has an orthonormal basis.**

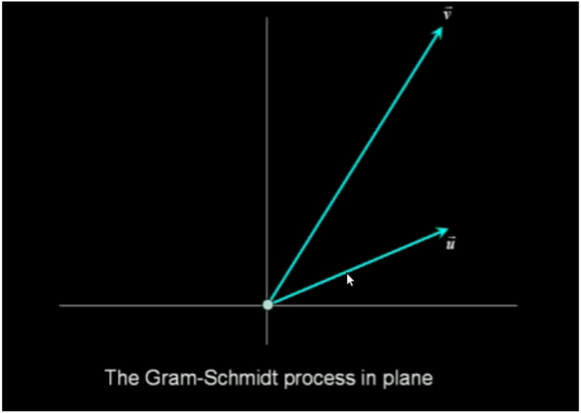
An Advantage in Unnormalized Gram-Schmidt Algorithm

The main difference between the algorithms for the w_k (normalized) and z_k (unnormalized) is that the vectors w_k are normalized after each step, where the z_k are not. Hence, they remain unnormalized! Avoiding the calculation of square root is another advantage.

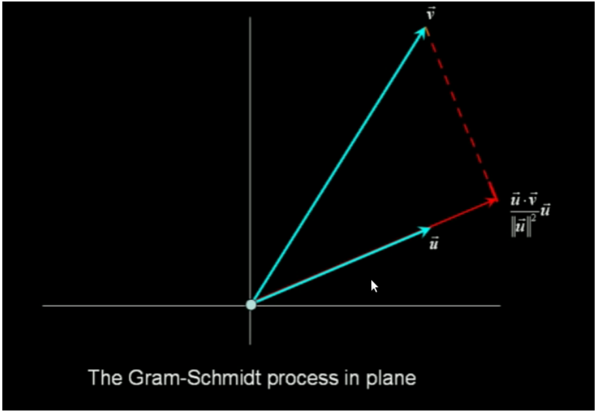
For hand calculations, it is easier to construct an orthonormal basis by first constructing an orthogonal basis and then normalizing the vectors all at once at the end.

Next few slides are showing the Gram-Schmidt orthogonalization process in plane and space.

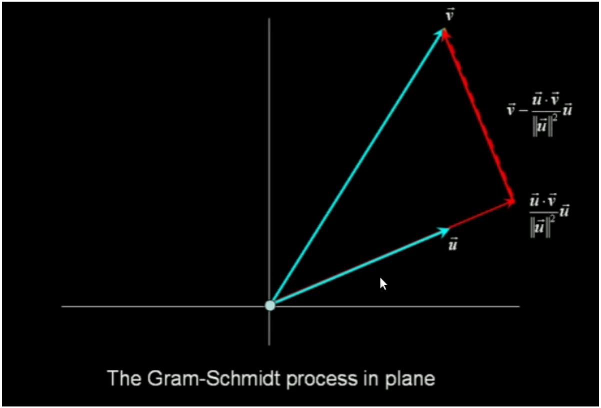
Gram-Schmidt process in plane



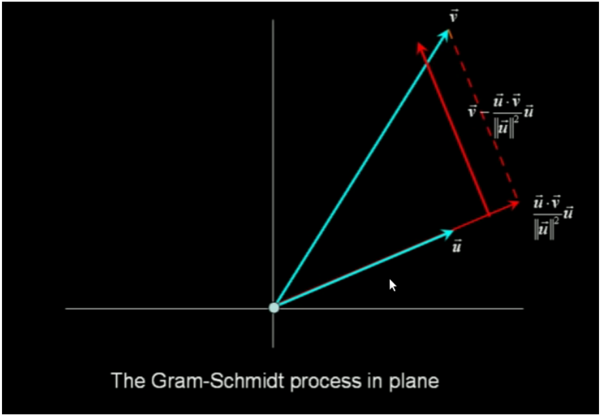
Gram-Schmidt process in plane



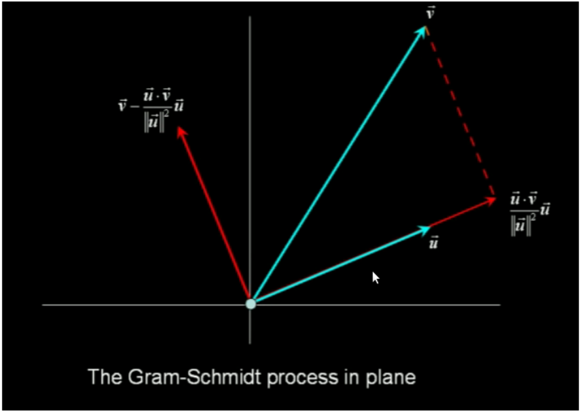
Gram-Schmidt process in plane



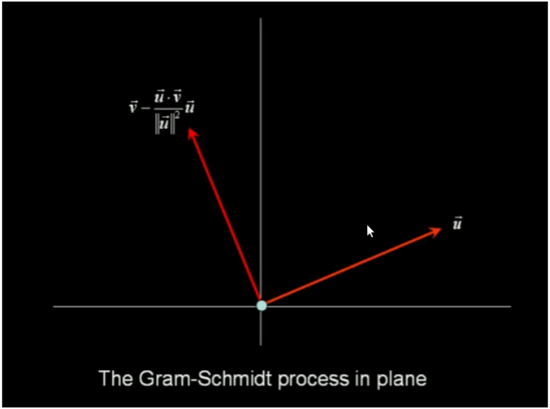
Gram-Schmidt process in plane



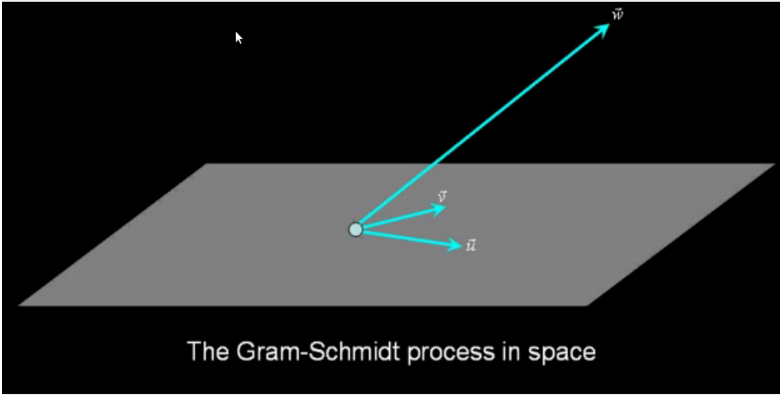
Gram-Schmidt process in plane



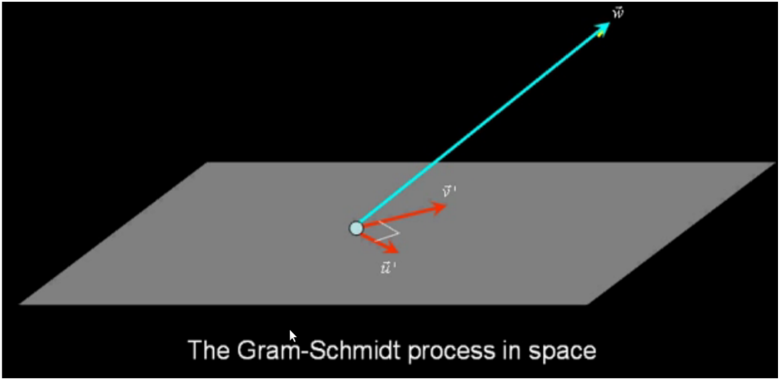
Gram-Schmidt process in plane



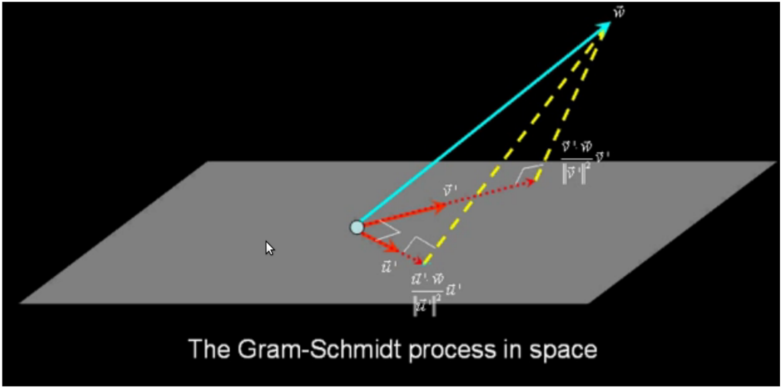
Gram-Schmidt process in space



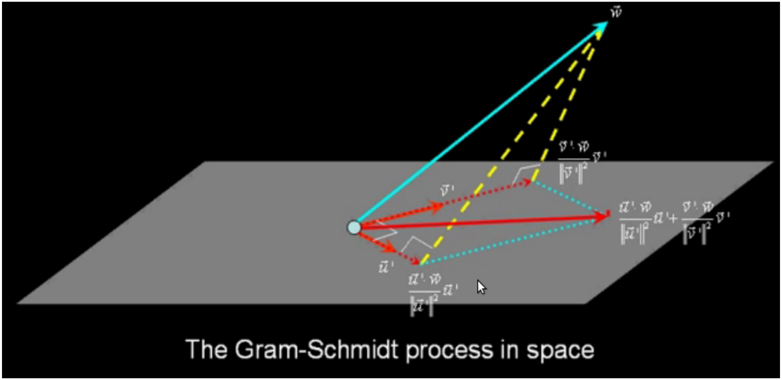
Gram-Schmidt process in space



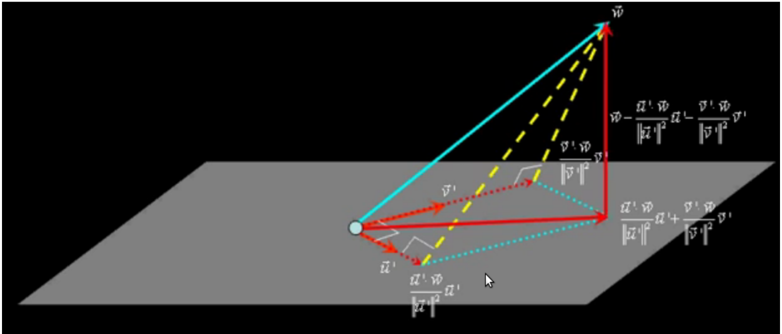
Gram-Schmidt process in space



Gram-Schmidt process in space

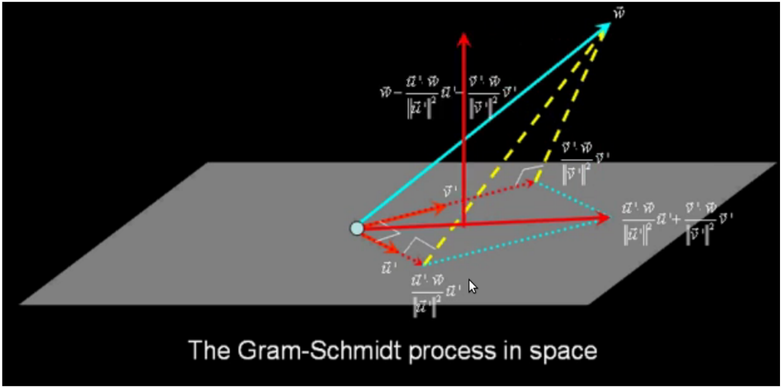


Gram-Schmidt process in space

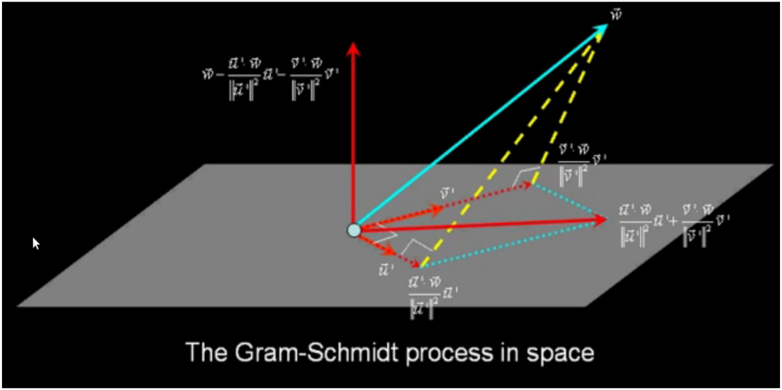


The Gram-Schmidt process in space

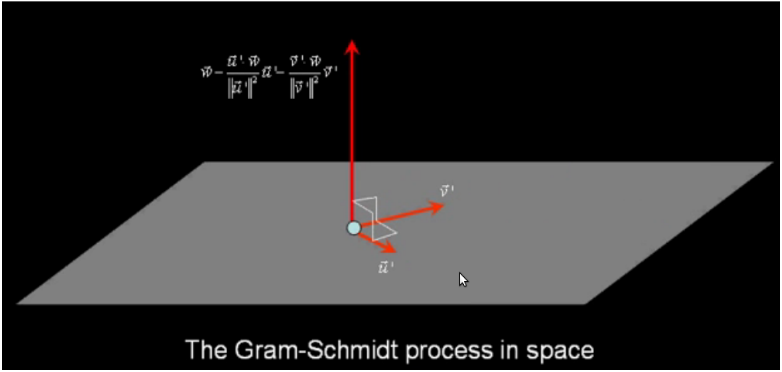
Gram-Schmidt process in space



Gram-Schmidt process in space



Gram-Schmidt process in space



Generalized Gram-Schmidt Process

Let $x_1, x_2, \ldots, x_s$ be a given vectors in V, not necessarily basis.

1. **Step 1:** Set $k = 1$.
2. **Step 2:** Compute $z_k = x_k - \sum_{j=1}^{k-1} \frac{\langle x_k, y_j \rangle}{y}_j$.
3. **Step 3:** Compute $y_k := \frac{z_k}{\|z_k\|}$ or 0 according as $z_k \neq 0$ or $z_k = 0$.
4. **Step 4:** If $k < s$, increase k by 1 and go to Step 2. Otherwise go to Step 5.
5. **Step 5:** For $i = 1, 2, \ldots, s$, the set B_i of all non-null vectors **among** $y_1, y_2, \ldots, y_i$ is an orthonormal basis of the span S_i of $\{x_1, x_2, \cdots, x_i\}$.

If $x_1, x_2, \ldots, x_\ell$ form an orthonormal set then $y_j = x_j$ for $j = 1, 2, \ldots, \ell$.

Theorem

Let S be a subspace of a finite-dimensional inner product space V. Any orthonormal subset of S can be extended to an orthonormal basis of S.

Proof. Let $A = \{x_1, x_2, \ldots, x_\ell\}$ be an orthonormal subset of S. Extend A to a spanning set $\{x_1, x_2, \ldots, x_\ell, x_{\ell+1}, \ldots, x_s\}$ of S by **appending a basis**. Applying the generallized Gram-Schmidt process to $\{x_1, x_2, \ldots, x_s\}$, get $\{y_1, y_2, \ldots, y_s\}$. Then the non-null vectors among $y_1, y_2, \ldots, y_s$ form an orthonormal basis of S which contains $A = \{x_1, x_2, \ldots, x_\ell\}$.

We note that the orthonormal basis obtained by the Gram-Schmidt process from $x_1, x_2, \ldots, x_\ell$ may be quite different from that obtained from generallized Gram-Schmidt process (a rearrangement of $x_1, x_2, \ldots, x_\ell$).

Exercises

1. Consider $\mathbb{R}^4$ with the usual inner product. Extend

$$\left\{ \frac{1}{\sqrt{3}}(1,0,1,-1)^T, \frac{1}{\sqrt{7}}(-2,1,1,-1)^T \right\}$$

to an orthonormal basis by the method of the preceding theorem.

2. Consider the inner product $\langle x, y \rangle = y^T A x$ on $\mathbb{R}^3$ where

$$A = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 1 & 0 \\ -1 & 0 & 3 \end{pmatrix}.$$

Find an orthonormal basis B of $S := \{(x_1, x_2, x_3) : x_1 + x_2 + x_3 = 0\}$ and then extend it to an orthonormal basis $\mathbb{C}$ of $\mathbb{R}^3$.

QR-decomposition

Let A be an $n \times s$ matrix with rank p. Let $y_1, y_2, \ldots, y_s$ be the vectors obtained when generalized Gram-Schmidt process is applied to the columns of A. For each $k = 1, 2, \ldots, s$,

$$z_k = A_{*k} - \sum_{j=1}^{k-1} \langle A_{*k}, y_j \rangle y_j = \|z_k\| y_k$$

and, $y_k := \frac{z_k}{\|z_k\|}$ or 0 according as $z_k \neq 0$ or $z_k = 0$.

Hence k-th column of A is a linear combination of $y_1, y_2, \ldots, y_s$. That is,

$$A_{*k} = \langle A_{*k}, y_1 \rangle y_1 + \langle A_{*k}, y_2 \rangle y_2 + \cdots + \langle A_{*k}, y_{k-1} \rangle y_{k-1} + \|z_k\| y_k.$$

QR-decomposition

$$A = [y_1 \ y_2 \ \cdots y_s] \begin{bmatrix} \|z_1\| & \langle A_{*2}, y_1\rangle & \langle A_{*3}, y_1\rangle & \cdots & \langle A_{*s}, y_1\rangle \\ 0 & \|z_2\| & \langle A_{*3}, y_2\rangle & \cdots & \langle A_{*s}, y_2\rangle \\ 0 & 0 & \|z_3\| & \cdots & \langle A_{*s}, y_3\rangle \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \|z_3\| \end{bmatrix}.$$

Let U be the $s \times s$ upper triangular matrix (u_{ik}) where

$$u_{ik} = \begin{cases} \langle A_{*k}, y_i\rangle & \text{if } i < k \\ \|z_k\| & \text{if } i = k \\ 0 & \text{otherwise} \end{cases}$$

Then $A = PU$.

Also if $\underline{Q \text{ is the submatrix of } P}$ formed by the non-null columns (the columns of Q form an orthonormal basis, $Q^*Q = I_p$) and $\underline{R \text{ the submatrix of } U}$ formed by the corresponding rows, then (Q, R) is a rank-factorization of A and $Q^*Q = I_p$.

When A is of full column rank $(Q, R) = (P, U)$ is known as a QR-**decomposition of** A.

Uniqueness. QR-factorization is unique if **we insist that the diagonal elements of** R **are real and positive**, i.e., if A is of full column rank, then there exist unique matrices Q and R such that $A = QR, Q^*Q = I$, R is upper triangular and $r_{ii} > 0$ for all i.

Exercises

1. Let x, y, u and v belong to $\mathbb{R}^n$. Then show that

$$\langle x + iy, u + iv \rangle := u^T x + v^T y$$

   is an inner product on the vector space $\mathbb{C}^n$ over $\mathbb{R}$.
   What is its connection with the canonical inner product on $\mathbb{C}^n$?

2. Show that $\langle x, y \rangle = 0$ for all y iff $x = 0$.

Example on the space of random variables

Let V be the vector space of all real-valued random variables with mean 0 and finite variance, defined on a fixed probability space. Let $F = \mathbb{R}$ and define $\langle x, y \rangle$ to be the covariance between x and y.

- An orthogonal set is a set of pairwise uncorrelated random variables. They form an orthonormal set if, further, each of them has unit variance.
- Suppose $A = \{x_1, x_2, \ldots, x_k\}$ be an orthogonal set (not a basis) of non-null vectors in V. Then for any $x \in V$,

$$z := x - \sum_{j=1}^{k} \frac{\langle x, x_j \rangle}{\langle x_j, x_j \rangle} x_j,$$

  the **residual of** x **with respect to** A. The sum $\sum_{j=1}^{k} \frac{\langle x, x_j \rangle}{\langle x_j, x_j \rangle}$ is the linear regression of x on $x_1, x_2, \ldots, x_k$.

References

- S. Kumaresan, "*Linear Algebra - A Geometric Approach*", PHI Learning Pvt. Ltd., 2011.
- A. Ramachandra Rao and P. Bhimasankaram, "*Linear Algebra*", Hindustan Book Agency, 2000.
- Ward Cheney and David Kincaid, "*Linear Algebra - Theory and Applications*", Jones & Bartlett, 2010.