Mathematics & Statistics
Auburn University, Alabama, USA

Dec 17, 2010

QR decomposition: History and its Applications

Tin-Yau Tam 譚天祐

学术报告
厦门大学数学科学学院

email: tamtiny@auburn.edu Website: www.auburn.edu/∼tamtiny

1. QR decomposition

Recall the QR decomposition of $A \in \mathrm{GL}_n(\mathbb{C})$:

$$A = QR$$

where $Q \in \mathrm{GL}_n(\mathbb{C})$ is unitary and $R \in \mathrm{GL}_n(\mathbb{C})$ is upper Δ with positive diagonal entries. Such decomposition is unique. Set

$$a(A) := \mathrm{diag}\,(r_{11}, \ldots, r_{nn})$$

where A is written in column form

$$A = (a_1 | \cdots | a_n)$$

Geometric interpretation of $a(A)$:
r_{ii} is the distance (w.r.t. 2-norm) between a_i and span $\{a_1, \ldots, a_{i-1}\}$, $i = 2, \ldots, n$.

Example:

$$\begin{pmatrix} 12 & -51 & 4 \\ 6 & 167 & -68 \\ -4 & 24 & -41 \end{pmatrix} = \begin{pmatrix} 6/7 & -69/175 & -58/175 \\ 3/7 & 158/175 & 6/175 \\ -2/7 & 6/35 & -33/35 \end{pmatrix} \begin{pmatrix} 14 & 21 & -14 \\ 0 & 175 & -70 \\ 0 & 0 & 35 \end{pmatrix}.$$

• QR decomposition is the matrix version of the Gram-Schmidt orthonormalization process.

• QR decomposition can be extended to rectangular matrices, i.e., if $A \in \mathbb{C}_{m\times n}$ with $m \geq n$ (tall matrix) and full rank, then

$$A = QR$$

where $Q \in \mathbb{C}_{m\times n}$ has orthonormal columns and $R \in \mathbb{C}_{n\times n}$ is upper Δ with positive "diagonal" entries.

2. QR history

• When Erhard Schmidt presented the formulae on p. 442 of his

E. Schmidt, Zur Theorie der linearen und nichtlinearen Integralgleichungen. I. Teil: Entwicklung willkülicher Funktionen nach Systemen vorgeschriebener, Math. Ann., **63** (1907) 433–476.

he said that essentially the same formulae were in

J. P. Gram, Ueber die Entwickelung reeler Funtionen in Reihen mittelst der Methode der kleinsten Quadrate, Jrnl. für die reine und angewandte Math. **94** (1883) 71–73.

• Modern writers, however, distinguish the two procedures, sometimes using the term "Gram-Schmidt" for the Schmidt form and "modified Gram-Schmidt" for the Gram version.

• But Gram-Schmidt orthonormalization appeared earlier in the work of Laplace and Cauchy.

• In the theory of semisimple Lie grougs, Gram-Schmidt process is extended to the Iwasawa decomposition $G = KAN$.

• A JSTOR search: the term "Gram-Schmidt orthogonalization process" first appears on p.57 of

Y. K. Wong, An Application of Orthogonalization Process to the Theory of Least Squares," Annals of Mathematical Statistics, **6** (1935), 53–75.

• In 1801 Gauss predicted the orbit of the steroid Ceres using the method of least squares. Since then, the principle of least squares has been the standard procedure for the analysis of scientific data.

• Least squares problem

$$Ax \approx b$$

i.e., finding $\hat{x}$ that would yield

$$\min_x \|Ax - b\|_2.$$

The solution is characterized by $r \perp R(A)$, where $r = b - Ax$ is the residual vector, or equivalently, given by the normal equation

$$A^*Ax = A^*b.$$

• With $A = QR$ and $A \in \mathbb{C}_{m\times n}$ tall with full rank

$$R^*Rx = (QR)^*QRx = R^*Q^*b \Leftrightarrow Rx = Q^*b.$$

There are several methods for computing the QR decomposition:

- GS or modified GS,
- Givens rotations (real A), or
- Householder reflections.

Disadvantage of GS: sensitive to rounding error (orthogonality of the computed vectors can be lost quickly or may even be completely lost) $\rightarrow$ modified Gram-Schmidt.

Idea of modified GS: do the projection step with a number of projections which will be against the errors introduced in computation.

Example:

$$A = \begin{pmatrix} 1+\epsilon & 1 & 1 \\ 1 & 1+\epsilon & 1 \\ 1 & 1 & 1+\epsilon \end{pmatrix}$$

with very small ϵ such that $3+2\epsilon$ will be computed accurately but $3+2\epsilon+\epsilon^2$ will be computed as $3+2\epsilon$. For example, $|\epsilon| < 10^{-10}$. Then

$$Q = \begin{pmatrix} \frac{1+\epsilon}{\sqrt{3+2\epsilon}} & \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{3+2\epsilon}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{3+2\epsilon}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}$$

and $\cos\theta_{12}$ and $\cos\theta_{13} \approx \pi/2$ but and $\cos\theta_{23} \approx \pi/3$.

See http://www-db.stanford.edu/TR/CS-TR-69-122.html for a heuristic analysis of why Gram-Schmidt not stable.

Computing QR by Householder reflections

A Householder reflection is a reflection about some hyperplane. Consider

$$Q_v = I - 2vv^*, \quad \|v\|_2 = 1.$$

Q_v sends v to $-v$ and fix pointwise the hyperplane $\perp$ to v. Householder reflections are Hermitian and unitary.

Let $e_1 = (1, 0, ..., 0)^T$. Recall

$$A = (a_1 | \cdots | a_n) \in \mathbf{GL}_n(\mathbb{C})$$

If $\|a_1\|_2 = \alpha_1$, set

$$u = a_1 - \alpha_1 e_1, \quad v = \frac{u}{\|u\|_2}, \quad Q_1 = I - 2vv^*$$

so that

$$Qa_1 = (\alpha_1, 0, \cdots, 0)^T = \alpha_1 e_1.$$

Then

$$Q_1 A = \begin{pmatrix} \alpha_1 & * & \dots & * \\ 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{pmatrix}$$

After t iterations of this process, $t \le n-1$,

$$R = Q_t \cdots Q_2 Q_1 A$$

is upper Δ. So, with

$$Q = Q_1 Q_2 \cdots Q_t$$

$A = QR$ is the QR decomposition of A.

• This method has greater numerical stability than GS.

• On the other hand, GS produces the q_j vector after the jth iteration, while Householder reflections produces all the vectors only at the end.

3. An Asymptotic result

Theorem 3.1. (Huang and Tam 2007) Given $A \in \mathrm{GL}_n(\mathbb{C})$. Let $A = Y^{-1}JY$ be the Jordan decomposition of A, where J is the Jordan form of A,

$$\operatorname{diag} J = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$$

satisfying $|\lambda_1| \geq \cdots \geq |\lambda_n|$. Then

$$\lim_{m\to\infty} a(A^m)^{1/m} = \operatorname{diag}(|\lambda_{\omega(1)}|, \ldots, |\lambda_{\omega(n)}|),$$

where the permutation ω is uniquely determined by the Gelfand-Naimark decomposition of $Y = L\omega U$:

$$\operatorname{rank} \omega(i|j) = \operatorname{rank}(Y)(i|j), \quad 1 \leq i, j \leq n.$$

Here $\omega(i|j)$ denotes the submatrix formed by the first i rows and the first j columns of ω, $1 \leq i, j \leq n$.

• Gelfand-Naimark decomposition of $Y = L\omega U$ is different from the Guassian decomposition $Y = P^T LU$ obtained by Gaussian elimination with row exchanges.

• None of P, U and L in the Gaussian decomposition is unique.

• But ω and diag U are unique in the Gelfand-Naimark decomposition.

H. Huang and T.Y. Tam, An asymptotic behavior of QR decomposition, Linear Algebra and Its Applications, 424 (2007) 96-107.

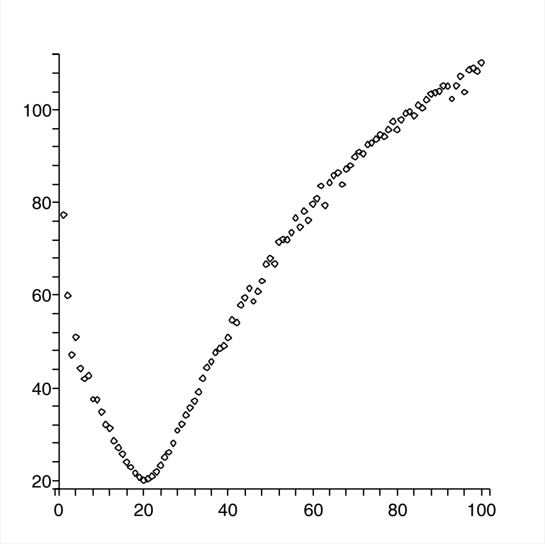
H. Huang and T.Y. Tam, An asymptotic result on the a-component in Iwasawa decomposition, Journal of Lie Theory, 17 (2007) 469-479.

Numerical experiments:

Computing the discrepancy between

$$[a(A^m)]^{1/m} \quad \text{and} \quad |\lambda(A)|$$

of randomly generated $A \in \mathrm{GL}_n(\mathbb{C})$.



The graph of

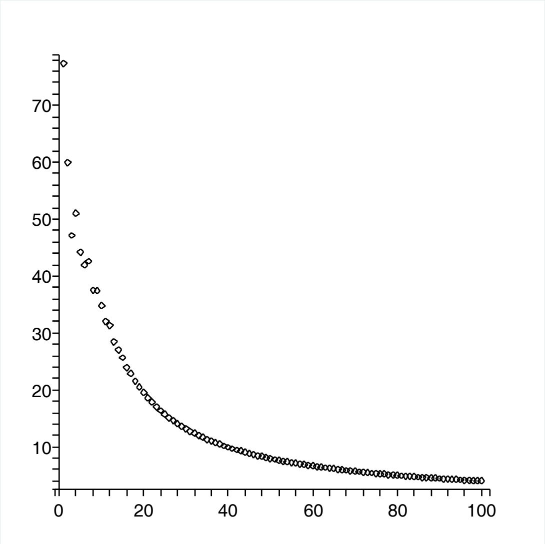
$$\|[a(A^m)]^{1/m} - \mathrm{diag}\,(|\lambda_1|, \ldots, |\lambda_n|)\|_2$$

versus m ($m = 1, \ldots, 100$)

If we consider

$$|a_1(A^m)^{1/m} - |\lambda_1(A)||$$

instead of $\|[a(A^m)]^{1/m} - \text{diag}\,(|\lambda_1|, \ldots, |\lambda_n|)\|_2$ for the above example, convergence occurs.



The graph of

$$|a_1(A^m)^{1/m} - |\lambda_1(A)||$$

versus m ($m = 1, \ldots, 100$)

4. QR iteration

• Because of Abel's theorem (1824), the roots of a general fifth order polynomial cannot be solved by radicals. Thus the computation of eigenvalues of $A \in \mathbb{C}_{n\times n}$ has to be approximative.

• Given $A \in \mathrm{GL}_n(\mathbb{C})$, define a sequence $\{A_k\}_{k\in\mathbb{N}}$ of matrices with

$$A_1 := A = Q_1 R_1$$

and if $A_k = Q_k R_k$,, $k = 1, 2, \ldots$ then

$$A_{k+1} := R_k Q_k = Q_k^* Q_k R_k Q_k = Q_k^* A_k Q_k$$

So the eigenvalues are fixed in the process.

• One hopes to have "some sort of convergence" of the sequence $\{A_k\}_{k\in\mathbb{N}}$ so that the "limit" would provide the eigenvalue approximation of A.

Theorem 4.1. (Francis (1961/62), Kublanovskaja (1961), Huang and Tam (2005)) Suppose that the moduli of the eigenvalues $\lambda_1, \ldots, \lambda_n$ of $A \in \mathrm{GL}_n(\mathbb{C})$ are distinct:

$$|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n| \ (> 0).$$

Let

$$A = Y^{-1}\mathrm{diag}\,(\lambda_1, \ldots, \lambda_n)Y.$$

Assume

$$Y = L\omega U,$$

where ω is a permutation, L is lower Δ and U is unit upper Δ. Then

1. the strictly lower Δ part of A_k converge to zero.
2. $\mathrm{diag}\, A_k \to \mathrm{diag}\,(\lambda_{\omega(1)}, \ldots, \lambda_{\omega(n)})$.

H. Huang and T.Y. Tam, On the QR iterations of real matrices, Linear Algebra and Its Applications, 408 (2005) 161-176.

- It rates as one of the most important algorithmic developments of the past century
- Parlett (2000): "The QR algorithm solves the eigenvalue problem in a very satisfactory way ... What makes experts in matrix computations happy is that this algorithm is a genuinely new contribution to the field of numerical analysis and not just a refinement of ideas given by Newton, Gauss, Hadamard, or Schur."
- Higham (2003): "The QR algorithm for solving the nonsymmetric eigenvalue problem is one of the jewels in the crown of matrix computations."

A common misconception

• Horn and Johnson's "Matrix Analysis" (p.114):

Under some circumstances (for example, if all the eigenvalues of A_0 has distinct absolute values), the QR iterates A_k will converge to an upper triangular matrix as $k \to \infty$....

• Quarteroni, Sacco, and Saleri's "Numerical Mathematics" (p.201-202):

Let $A \in \mathbb{R}^{n\times n}$ be a matrix with real eigenvalues such that $|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n|$. Then

$$\lim_{k\to\infty} T^{(k)} = \begin{pmatrix} \lambda_1 & t_{12} & \dots & t_{1n} \\ 0 & \lambda_2 & t_{23} & \dots \\ 0 & \cdots & \ddots & \\ 0 & \cdots & & \lambda_n \end{pmatrix}$$

• Wikipedia: Under certain conditions, the matrices A_k converge to a triangular matrix, the Schur form of A.

D. Serre's "Theory of Matrices" (p.176-177):

Let us recall that the sequence A_k is not always convergent. For example, if A is already triangular, its QR factorization is $Q = D$, $R = D^{-1}A$, with $d_j = a_{jj}/|a_{jj}|$. Hence, $A_1 = D^{-1}AD$ is triangular, with the same diagonal as that of A. By induction, A_k is triangular, with the same diagonal as that of A so that $A_k = D^{-k}AD^k$.... Hence, the part above the diagonal of A_k does not converge.

Summing up, a convergence theorem may concern only the diagonal of A_k and what is below it.

More on QR algorithm

• The QR algorithm is numerically stable because it proceeds by orthogonal/unitary similarity transforms.

•

$$A \to \text{ Hessenberg form } \to \text{ QR algorithm}$$

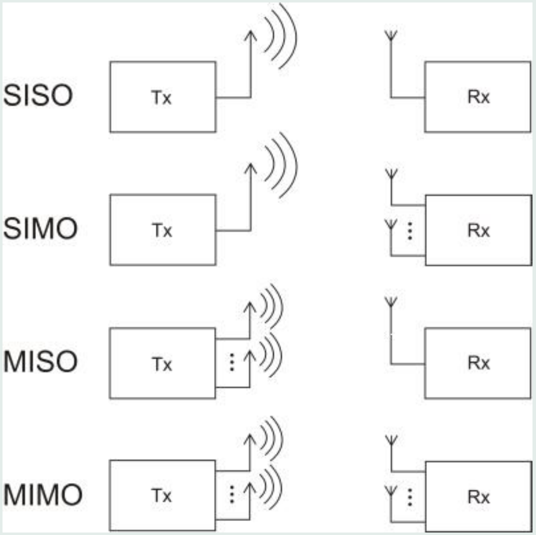
It would be cost effective if we convert A to an upper Hessenberg form, with a finite sequence of orthogonal similarities. Then determine the QR decomposition of an upper Hessenberg matrix costs $6n^2 + O(n)$ arithmetic operations.

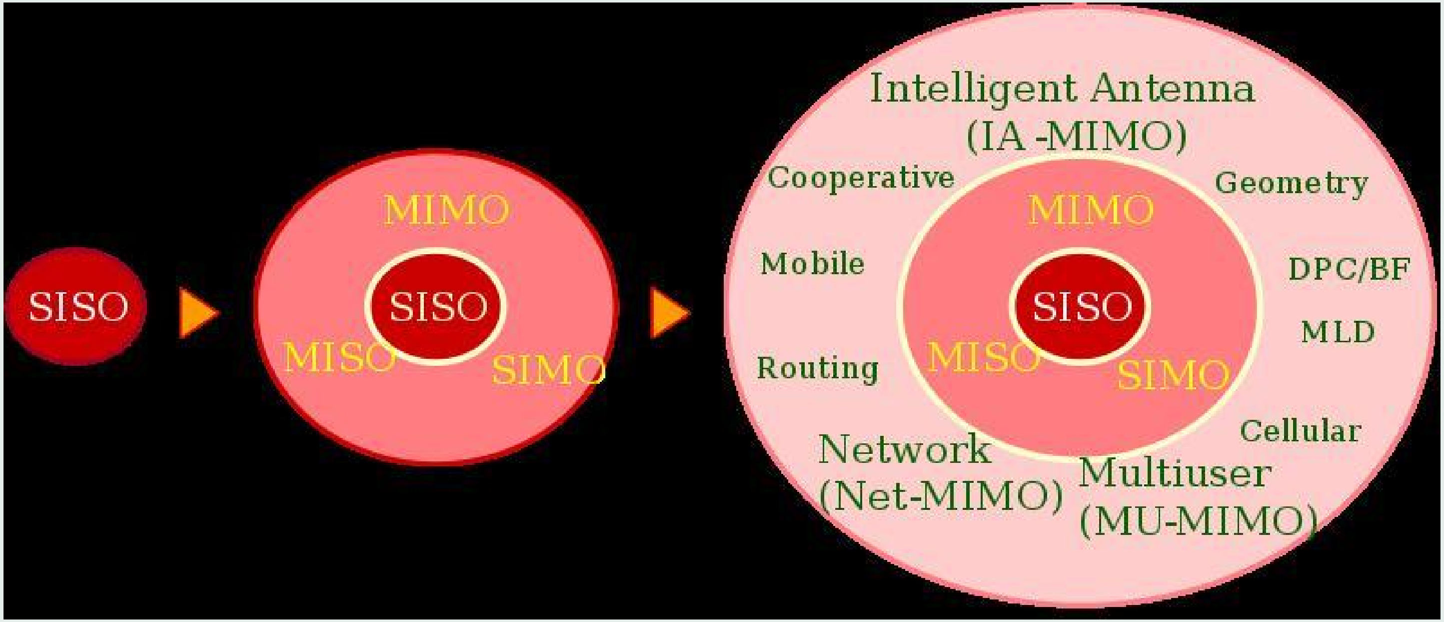
• A practical algorithm will use shifts to increase separation and accelerate convergence.

• The QR iteration is further extended to semisimple Lie groups:
H. Huang. R.R. Holmes and T.Y.Tam, Asymptotic behavior of Iwasawa and Cholesky iterations, manuscript.

5. Application in MIMO

In radio, multiple-input and multiple-output, or MIMO, is the use of multiple antennas at both the transmitter and receiver to improve communication performance.



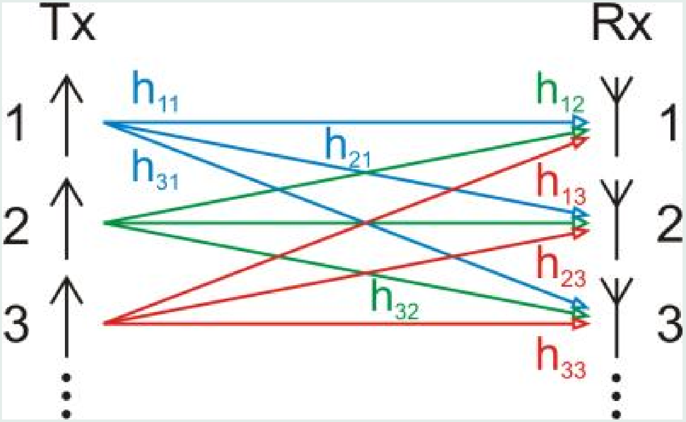


• MIMO is one of several forms of smart antenna technology.

• MIMO offers significant increases in data throughput and link range without additional bandwidth or transmit power.

• MIMO is an important part of modern wireless communication standards such as IEEE 802.11n (Wifi) and 4G.

• In MIMO systems, a transmitter sends multiple streams by multiple transmit antennas. The transmit streams go through a matrix channel which consists of all $N_t N_r$ paths between the N_t transmit antennas at the transmitter and N_r receive antennas at the receiver.



• Then, the receiver gets the received signal vectors by the multiple receive antennas and decodes the received signal vectors into the original information.

Mathematical description

Let $\mathbf{x} = (x_1, \ldots, x_n)^T$ be a vector to be transmitted over a noisy channel. Each x_i is chosen from a finite-size alphabet $\mathcal{X}$.

A general MIMO system is modeled as

$$\mathbf{r} = \mathbf{H}\mathbf{x} + \xi$$

• $\mathbf{H}$ is the $m \times n$ full column rank channel (tall, i.e., $m \geq n$) matrix (known to the receiver)
• $\xi = (\xi_1, \ldots, \xi_m)^T$ is a white Gaussian noise vector where $E(\xi\xi^*) = \sigma^2 I$
• $\mathbf{r} = (r_1, \ldots, r_m)^T$ is the observed received vector.
• Our task is to detect/estimate the vector $\hat{\mathbf{x}} = (\hat{x}_1, \ldots, \hat{x}_n)^T \in \mathcal{X}^n$ given the noisy observation $\mathbf{r}$.

• QR decomposition of the channel matrix $\mathbf{H}$ can be used to form the back-cancellation detector.

A. Successive Cancellation Detection Using QR Decomposition

(1) QR decomposition

Let $\mathbf{H} = \mathbf{QR}$ be the QR decomposition of the $m \times n$ channel matrix $\mathbf{H}$, i.e.,

- $\mathbf{Q} = m \times n$ matrix with orthonormal columns
- $\mathbf{R} = n \times n$ upper Δ matrix.

So

$$\mathbf{r} = \mathbf{Hx} + \xi \Rightarrow \mathbf{Q}^*\mathbf{r} = \mathbf{Rx} + \mathbf{Q}^*\xi$$

Set

$$\tilde{\mathbf{r}} := \mathbf{Q}^*\mathbf{r}, \quad \tilde{\xi} := \mathbf{Q}^*\xi$$

So

$$\begin{pmatrix} \tilde{r}_1 \\ \tilde{r}_2 \\ \vdots \\ \tilde{r}_n \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} \tilde{\xi}_1 \\ \tilde{\xi}_2 \\ \vdots \\ \tilde{\xi}_n \end{pmatrix}$$

(2) Hard decision

Estimate x_n by making the hard decision:

$$\hat{x}_n := \text{Quant}\,[\frac{\tilde{r}_n}{r_{nn}}]$$

where Quant (t) is the element $\mathcal{X}$ that is closest w.r.t. 2-norm to t.

(3) Cancellation

Backward substitutions:

$$\hat{x}_n := \text{Quant}\,[\frac{\tilde{r}_n}{r_{nn}}]$$

$$\hat{x}_k := \text{Quant}\,[\frac{\tilde{r}_k - \sum_{i=k+1}^{n} r_{km}\hat{x}_i}{r_{kk}}], \quad k = n-1, n-2, \ldots, 1$$

• The algorithm is essentially least squares solution via QR decomposition.

B. Optimally Ordered Detection

• Golden et al proposed a vertical Bell Laboratories layered space-time (V-BLAST) system with an optimal ordered detection algorithm that maximizes the signal-to-noise ratio (SNR).

G.D. Golden, G.J. Foschini, R.A. Valenzuela, and P.W. Wolniansky, Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture, Electron. Lett., vol. 35, pp. 14 – 15, Jan. 1999.

• The idea is (equivalently) to find an $n \times n$ permutation matrix $\mathbf{P}$ such that if

$$\mathbf{H}' := \mathbf{H}\mathbf{P} = \mathbf{Q}\mathbf{R}$$

then the distance between $\mathbf{h}'_n$ and the other columns $\mathbf{h}'_1, \ldots, \mathbf{h}'_{n-1}$ is maximal. Then repeat the process by stripping off the column vectors one.

• Doing so amounts to finding a subchannel whose SNR is the highest among all n possible subchannels. Geometrically it is to find the column of $\mathbf{H}$ whose distance to the span of the other columns of $\mathbf{H}$ is maximum. Then do it recursively.

• Equivalently

$$\mathbf{r} = \mathbf{H}'\mathbf{x}' + \xi$$

where

$$\mathbf{H}' := \mathbf{H}\mathbf{P}, \quad \mathbf{x} = \mathbf{P}\mathbf{x}'$$

i.e., if we precode a vector with the permutation matrix $\mathbf{P}$, and apply Algorithm A to detect $\mathbf{x}'$, we get the optimally ordered successive-cancellation detector of Golden et al.

A recent optimal QR decomposition, called equal-diagonal QR decomposition, or briefly the QRS decomposition in introduced in

J.K. Zhang, A. Kavčić and K. M. Wong, Equal-Diagonal QR Decomposition and its Application to Precoder Design for Successive-Cancellation Detection, IEEE Transactions on Information Theory, **51** (2005) 154–172.

The QRS decomposition is applied to precode successive-cancellation detection, where we assume that both the transmitter and the receiver have perfect channel knowledge.

Theorem 5.1. (Zhang, Kavčić and Wong 2005) For any channel matrix $\mathbf{H}$, there exists a unitary precoder $\mathbf{S}$, such that the nonzero diagonal entries of the upper Δ matrix $\mathbf{R}$ are all equal where

$$\mathbf{HS} = \mathbf{QR}.$$

Nice properties

The precoder $\mathbf{S}$ and the resulting successive-cancellation detector have many nice properties.

• The minimum Euclidean distance between two signal points at the channel output is equal to the minimum Euclidean distance between two constellation points at the precoder input up to a multiplicative factor that equals the diagonal entry in the $\mathbf{R}$-factor.

• The superchannel $\mathbf{HS}$ naturally exhibits an optimally ordered column permutation, i.e., the optimal detection order for the vertical Bell Labs layered space – time (V-BLAST) detector is the natural order.

• The precoder $\mathbf{S}$ minimizes the block error probability of the QR successive cancellation detector.

• A lower and an upper bound for the free distance at the channel output is expressible in terms of the diagonal entries of the $\mathbf{R}$-factor in the QR decomposition of a channel matrix.

• The precoder $\mathbf{S}$ maximizes the lower bound of the channels free distance subject to a power constraint.

• For the optimal precoder $\mathbf{S}$, the performance of the QR detector is asymptotically (at large signal-to-noise ratios (SNRs)) equivalent to that of the maximum-likelihood detector (MLD) that uses the same precoder.

- Recall the QRS decomposition $\mathbf{HS} = \mathbf{QR}$ Let

$$\mathbf{H} = \mathbf{U\Sigma V}$$

be the SVD of $\mathbf{H}$ where $\mathbf{\Sigma} = \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_n \\ & O & \end{pmatrix}$. Then set $\mathbf{S} = \mathbf{V}^*\hat{\mathbf{S}}$

$$\mathbf{HS} = \mathbf{HV}^*\hat{\mathbf{S}} = \mathbf{U\Sigma}\hat{\mathbf{S}} = \mathbf{U}\begin{pmatrix} \operatorname{diag}(s_1, \ldots, s_n)\hat{\mathbf{S}} \\ 0 \end{pmatrix}$$

where the unitary $\hat{\mathbf{S}}$ is to be determined so that $\mathbf{R}$ has equal diagonal entries. Clearly the $\mathbf{R}$ factors of $\mathbf{\Sigma}\hat{\mathbf{S}}$ and $\mathbf{U\Sigma}\hat{\mathbf{S}}$ are the same.
WLOG, assume that $m = n$, i.e., $\mathbf{\Sigma} = \operatorname{diag}(s_1, \ldots, s_n)$. The problem is reduced to finding a unitary $\mathbf{S}$ such that

$$\mathbf{\Sigma S} = \mathbf{QR}$$

and $\mathbf{R}$ has identical diagonal entries.

Theorem 5.2. (Kostant 1973) Let $\mathbf{\Sigma} = \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_n \end{pmatrix} \in \mathbb{C}_{n\times n}$ $s_1 \geq \cdots \geq s_n > 0$. For any unitary $\mathbf{S} \in \mathbb{C}_{n\times n}$, if $\mathbf{\Sigma S} = \mathbf{QR}$, where $\mathbf{Q} \in \mathbb{C}_{m\times n}$ has orthonormal columns, $\mathbf{R} \in \mathbb{C}_{n\times n}$ is upper Δ, then

$$\prod_{i=1}^{k} r'_{ii} \leq \prod_{i=1}^{k} s_i, \quad k = 1, \ldots, n-1 \tag{1}$$

$$\prod_{i=1}^{n} r'_{ii} = \prod_{i=1}^{n} s_i, \tag{2}$$

where $r'_{11} \geq \cdots \geq r'_{nn}$ are the rearrangement of $r_{11}, \ldots, r_{nn}$. Conversely if (1) and (2) are satisfied, then there is an $n \times n$ unitary $\mathbf{S}$ such that the diagonal entries of $\mathbf{R}$ in $\mathbf{\Sigma S} = \mathbf{QR}$ are those r_{ii}, $i = 1, \ldots, n$.

B. Kostant, On convexity, the Weyl group and Iwasawa decomposition. Ann. Sci. Ecole Norm. Sup. (4), **6** (1973) 413–460.

• Clearly Kostant (1973) $\Rightarrow$ Zhang, Kavčić and Wong (2005).

• The proof of Kostant is not constructive. Indeed his result is true for semisimple Lie groups.

• The construction of the $\mathbf{S}$ precoder in [ZKW] is not very cost effective and involves a number of steps. Determinant is also involved.

Shuangchi He and Tin-Yau Tam, On equal-diagonal QR decomposition, manuscript.

Use induction. WLOG assume that $m = n$.

2×2 case:

$$\mathbf{\Sigma S} = \begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} s_1\cos\theta & -s_1\sin\theta \\ s_2\sin\theta & s_2\cos\theta \end{pmatrix}$$

Then for any $s_1 \geq \mu \geq s_2$,

$$\|(s_1\cos\theta, s_2\sin\theta)^T\|_2 = s_1\cos^2\theta + s_2\sin^2\theta = \mu$$

for some $\theta \in \mathbb{R}$. In particular $\mu = \sqrt{s_1 s_2}$

Suppose that the statement holds true for $n-1$.

For $r = n > 2$, suppose that k is the smallest index such that $s_{k-1} \geq \lambda \geq s_k$, where

$$\lambda := (\prod_{i=1}^{n} s_i)^{1/n}.$$

There is a unitary $\mathbf{S_1} \in \mathbb{C}_{2\times 2}$ such that

$$\text{diag}\,(s_1, s_k)\mathbf{S_1} = \mathbf{A_1} := \begin{pmatrix} \lambda & * \\ 0 & \frac{s_1 s_k}{\lambda} \end{pmatrix}.$$

Set $\mathbf{S_2} = \mathbf{S_1} \oplus \mathbf{I_{n-2}}$. Then

$$\begin{aligned} \mathbf{A_2} &:= \text{diag}\,(s_1, s_k, s_2, \cdots, s_{k-1}, s_{k+1}, \cdots, s_n)\,\mathbf{S_2} \\ &= \mathbf{A_1} \oplus \text{diag}\,(s_2, \cdots, s_{k-1}, s_{k+1}, \cdots, s_n)\,. \end{aligned}$$

Note that

$$\lambda^{n-1} = \frac{s_1 s_k}{\lambda} \prod_{i=2, i\neq k}^{n} s_i.$$

By the inductive hypothesis, there exists a unitary $\mathbf{S_3} \in \mathbb{C}_{(n-1)\times(n-1)}$ such that

$$\mathbf{A_3} = \text{diag}\left(\frac{s_1 s_k}{\lambda}, s_2, \cdots, s_{k-1}, s_{k+1}, \cdots, s_n\right)\mathbf{S_3}$$

is upper Δ with equal-diagonal entries λ. Then $\mathbf{S_4} = 1 \oplus \mathbf{S_3}$.

THANK YOU FOR YOUR ATTENTION