Challenge Problem Set 5, Math 291 Fall 2013

Eric A. Carlen[1]
Rutgers University

November 17, 2013

The focus of this challenge problem set is *factorization* of general linear transformations into composition products of very simple transformations, so that we may understand their geometric effect, and particularly, their effect on volume.

Recall that *matrix-matrix* multiplication is defined so that the matrix-matrix product of BA is the matrix representing the composition of the linear transformation sending $\mathbf{x}$ to $\mathbf{y} = A\mathbf{x}$, and then $\mathbf{y}$ to $B\mathbf{y}$. Here is how this was done:

Let A be any $m \times n$ matrix. Let B by any $\ell \times m$ matrix. Then for any $\mathbf{x} \in \mathbb{R}^n$, $A\mathbf{x} \in \mathbb{R}^m$, so it makes sense to consider the matrix-vector product $B(A\mathbf{x})$. By definition, if $A = [\mathbf{v}_1, \dots, \mathbf{v}_n]$, and $\mathbf{x} = (x_1, \dots, x_n)$,

$$A\mathbf{x} = \sum_{j=1}^{n} x_j \mathbf{v}_j \ .$$

But then, since matrix-vector multiplication is linear,

$$B(A\mathbf{x}) = B\left(\sum_{j=1}^{n} x_j \mathbf{v}_j\right) = \sum_{j=1}^{n} x_j B\mathbf{v}_j = [B\mathbf{v}_1, \dots, B\mathbf{v}_n]\mathbf{x} \ ,$$

we define

$$BA = [B\mathbf{v}_1, \dots, B\mathbf{v}_n] \ .$$

That is, to form the matrix-matrix product of B and A the number of rows in A must match the number of columns in B, and then the the jth column of the product BA is the product of B and the jth column of A. This is how we arrived at the definition of matrix-matrix multiplication in terms of matrix-vector multiplication, that we now recall:

0.1 DEFINITION (Matrix multiplication). Given an $m \times n$ matrix A with columns $\mathbf{v}_1, \dots, \mathbf{v}_n$, so that $A = [\mathbf{v}_1, \dots, \mathbf{v}_n]$, and an $\ell \times m$ matrix B, define the matrix-matrix product BA by

$$BA = [B\mathbf{v}_1, \dots, B\mathbf{v}_n] \ .$$

The matrix product BA is only defined when the number of rows in A equals the number of columns in B.

---

1. Let $B = \begin{bmatrix} 1 & 2 \\ 3 & 0 \\ 2 & 1 \end{bmatrix}$ and $A = \begin{bmatrix} 2 & 2 & 1 \\ 3 & 1 & 2 \end{bmatrix}$. Compute BA.

2. Define matrices G, H and K by

$$G = \begin{bmatrix} 1 & a & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad H = \begin{bmatrix} 1 & 0 & b \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad K = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix}$$

for some numbers a, b and c.

Compute the matrix products $K(GH)$ $(KG)H$ and $G(HK)$.

As your results from should show, the order of matrices in a product matters; matrix multiplication does not commute. However, it is associative, as in $K(GH) = (KG)H$, since it represents the compositions of functions, and this is associative. Indeed, for any sets X, Y, Z and W, and any functions $f : X \to Y$, $g : Y \to Z$ and $h : Z \to W$,

$$((h \circ g) \circ f)(x) = (h \circ g)(f(x)) = h(g(f(x))) = h(g \circ f(x)) = (h \circ (g \circ f))(x) \ ,$$

for all $x \in X$. That is, $(h \circ g) \circ f = h \circ (g \circ f)$ The associativity of matrix multiplication is a special case of this.

Multiplying matrices is easy. Our goal in what follows is to *factor* a general $n \times n$ matrix as a product of *simple* $n \times n$ matrices, each of which represents a transformation of $\mathbb{R}^n$ that has an easy to compute effect on volumes of sets. Our main interest is in $\mathbb{R}^3$, where you know what volume means, at least for parallelepipeds, but our methods easily generalize to other values of n, and we shall be general where convenient.

1 Factorization of triangular matrices

1.1 DEFINITION (diagonal matrices and upper triangular matrices)**.** An $n \times n$ matrix A is a non-negative diagonal matrix if an only if $A_{i,j} = 0$ for all $i \neq j$. An $n \times n$ matrix A is upper triangular in case $A_{i,j} = 0$ for all $i > j$. An $n \times n$ matrix is a *unit upper triangular matrix* if it is upper triangular, and every diagonal entry is 1. A *shear matrix* is a unit upper triangular matrix that has only one non-zero off diagonal entry.

For example,

$$S = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1.1}$$

is a diagonal matrix with non-negative entries.

Also, the matrices G, H and K from Exercise 2 are shear transformations, and their product, as you will have computed, is unit upper triangular.

Now let us examine the effects of the corresponding transformations on volume.

The matrix S in (1.1) represents a *scaling transformation*: The corresponding linear transformation stretches distances along the x-axis by a factor of 2, stretches distances along the y-axis by a factor of 3, and leaves distances along the z axis unstretched.

---

3. Let $\mathcal{C}$ denote the cube of side length h given by

$$\mathcal{C} = \{a \le x \le a+h\ ,\ b \le y \le b+h\ ,\ c \le z \le c+h\ \}$$

for some numbers a, b, and c, and some $h > 0$.

Let $S(\mathcal{C})$ denote the *image* of $\mathcal{C}$ under the linear transformation sending $\mathbf{x}$ to $S\mathbf{x}$. That is, $S(\mathcal{C})$ denote the set of all points $\mathbf{y} \in \mathbb{R}^3$ such that $\mathbf{y} = S\mathbf{x}$ for some $\mathbf{x} \in \mathcal{C}$. What is the volume of $S(\mathcal{C})$? (Your answer should be a specific numerical multiple of h^3.)

More generally let $\mathcal{D}$ be an finite disjoint union of such cubes. What is the ration of the volume of $S(\mathcal{D})$ to the ration of the volume of $\mathcal{D}$?

4. Again, let $\mathcal{C}$ denote the cube of side length h given by

$$\mathcal{C} = \{a \le x \le a+h\ ,\ b \le y \le b+h\ ,\ c \le z \le c+h\ \}$$

for some numbers a, b, and c, and some $h > 0$. Consider any of the shear transformations G, H ,or K defined above. Show that applying any of these to $\mathcal{C}$ produces a parallelepiped whose "slices" perpendicular to one coordinate axis are still squares of side length h, and whose volume is still h^3. More generally, show that if $\mathcal{D}$ is any finite disjoint union of cubes, and A is any shear transformation, the volume of $A(\mathcal{D})$ equals the volume of $\mathcal{D}$.

5. Let $D := \begin{bmatrix} r & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & t \end{bmatrix}$ be any 3×3 diagonal matrix. Let E be an 3×3 matrix with rows $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$, and columns $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$. That is

$$E = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{r}_3 \end{bmatrix} = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]\ .$$

Show that

$$DE = \begin{bmatrix} r\mathbf{r}_1 \\ s\mathbf{r}_2 \\ t\mathbf{r}_3 \end{bmatrix}\ .$$

What is ED?

We are now ready to do some matrix factorization: We will consider a "complicated" matrix, and take it apart into a product of simple factors. By understanding what each of the factors do to volume, we can understand what the whole transformation does to volume.

---

6. Let $A := \begin{bmatrix} 2 & -2 & 4 \\ 0 & 3 & 6 \\ 0 & 0 & 2 \end{bmatrix}$. Use the results of the exercises above to write

$$A = DT$$

where D is diagonal, and T is unit upper triangular. Then find three shear transformations S_1, S_2 and S_3 so that $T = S_1S_2S_3$. Thus, you have

$$A = DS_1S_2S_3 \ . \tag{1.2}$$

Using this factorization, and the results of previous exercises, determine the volume of $A(\mathcal{C})$. finally, show that every 3×3 upper triangular matrix has a factorization of the form (1.2).

7. Show that every 4×4 upper triangle matrix can be written as the the product of a diagonal matrix and six shear transformations. What do you conclude about higher n?

We are now ready to draw an important conclusion: Let A be any 3×3 or 4×4 upper triangular matrix A, and let $\rho(A)$ denote the product of the absolute values of its diagonal entries. Then for any region $\mathcal{D}$ that is a finite disjoint union of cubes,

$$\mathrm{vol}(A(\mathcal{D})) = \rho(A)\mathrm{vol}(\mathcal{D}) \ ,$$

since the volume of each cube gets magnified (shrunk, really if $\rho(A) < 1$) by this factor. Note that $\rho(A)$ is determined entirely by A, and then you simply mutiply the vilume of $\mathcal{D}$ by this factor to compute the volume of the image of $\mathcal{D}$. The same is true for all regions $\mathcal{D}$ that can be well-approximated by a finite disjoint union of cubes, which are the only kind we will meet in this course.

2 Factorization of general square matrices

Let A be any $n \times n$ matrix, and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be its columns so that $A = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$. Our goal is to write factor A as a product

$$A = QR$$

where Q has orthonormal columns, and R is upper triangular. Matrices with orthonormal columns are called *orthogonal matrices.*

2.1 About orthogonal matrices

2.1 DEFINITION (Orthogonal matrices)**.** An $n \times n$ matrix U is an orthogonal matrix if an only if the columns of U are an orthonormal basis of $\mathbb{R}^n$.

For example, $Q = \dfrac{1}{3}\begin{bmatrix} 2 & 2 & 1 \\ 2 & -1 & -2 \\ 1 & -2 & 2 \end{bmatrix}$ is an orthogonal matrix.

In fact, this is the matrix representing the Householder reflection that reflects $\mathbf{e}_1$ onto the unit vector $\mathbf{u}_1 = \dfrac{1}{3}(2, 2, 1)$. As is geometrically clear, reflection has no effect on the volume of regions: If $\mathcal{D}$ be an finite disjoint union of cubes, then the volume of $Q(\mathcal{D})$ equals the volume of $\mathcal{D}$.

In fact, it is easy to see that for *any* orthogonal transformation Q, if $\mathcal{D}$ be an finite disjoint union of cubes, then the volume of $Q(\mathcal{D})$ equals the volume of $\mathcal{D}$. and thus the same is true if any set that can be "well approximated" as a finite disjoint union of (small) cubes.

To see this, observe that by the basic definition of matrix vector multiplication; i.e.,

$$Q\mathbf{x} = [\mathbf{u}_1, \ldots, \mathbf{u}_n](x_1, \ldots, x_n) = \sum_{j=1}^{n} x_j \mathbf{u}_j$$

and hence

$$\|Q\mathbf{x}\|^2 = \|\mathbf{x}\|^2$$

for all $\mathbf{x} \in \mathbb{R}^n$. That is, Q preserves the lengths of vectors.

Applying this to $\mathbf{x}+\mathbf{y}$ we see that $\|Q(\mathbf{x}+\mathbf{y})\|^2 = \|\mathbf{x}+\mathbf{y}\|^2$. But $Q(\mathbf{x}+\mathbf{y}) = Q\mathbf{x}+Q\mathbf{y}$, so

$$\|Q(\mathbf{x}+\mathbf{y})\| = \|Q\mathbf{x}\|^2 + 2(Q\mathbf{x})\cdot(Q\mathbf{y}) + \|Q\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2(Q\mathbf{x})\cdot(Q\mathbf{y}) + \|\mathbf{y}\|^2 \ ,$$

and

$$\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2\mathbf{x}\cdot\mathbf{y} + \|\mathbf{y}\|^2 \ .$$

Comparing, we see that $\|Q(\mathbf{x}+\mathbf{y})\|^2 = \|\mathbf{x}+\mathbf{y}\|^2$ for all $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ implies that

$$Q\mathbf{x}\cdot Q\mathbf{y} = \mathbf{x}\cdot\mathbf{y}$$

for all $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$. Conclude that U also preserves angles between vectors. We conclude that the image of a cube is again a cube of the same side length since the lengths and perpendicularity of the edges can be expressed in terms of dot products that are unchanged.

Next, we claim that orthogonal transformations are always invertible. To see that they transform $\mathbb{R}^n$ onto $\mathbb{R}^n$, note that *every* vector $\mathbf{b}$ in $\mathbb{R}^n$ can be written as a linear combination of the vectors in the orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$; i.e.,

$$\mathbf{b} = \sum_{j=1}^{n} (\mathbf{u}_j \cdot \mathbf{b})\mathbf{u}_j \ ,$$

and we can write this in terms of matrix-vector multiplication as

$$\mathbf{b} = [\mathbf{u}_1, \ldots, \mathbf{u}_n](\mathbf{b}\cdot\mathbf{u}_1, \ldots, \mathbf{b}\cdot\mathbf{u}_n) \ .$$

That is, with

$$\mathbf{x} := (\mathbf{b}\cdot\mathbf{u}_1, \ldots, \mathbf{b}\cdot\mathbf{u}_n) \ ,$$

$$Q\mathbf{x} = \mathbf{b} \ .$$

That is, since $\mathbf{b} \in \mathbb{R}^n$ is arbitrary, Q transforms $\mathbb{R}^n$ *onto* $\mathbb{R}^n$. Also

$$\|Q\mathbf{x} - Q\mathbf{y}\| = \|Q(\mathbf{x}-\mathbf{y})\| = \|\mathbf{x}-\mathbf{y}\|$$

so that $Q\mathbf{x} = Q\mathbf{y}$ if and only if $\mathbf{x} = \mathbf{y}$. Thus Q is *one to one.* (By the Fundamental Theorem of Linear algebra, every linear transformation of $\mathbb{R}^N$ onto itself is automatically one-to-one, but we can check this easily for orthogonal matrices without invoking this theorem.)

Therefore, the image of a disjoint union of cubes under an orthogonal transformation Q is again a disjoint union of the same finite number of cubes, each with the same side length as the cube that is its pre-image. Thus, *orthogonal transformations preserve volume.*

8. As a byproduct of the above discussion, one can deduce a formula for the inverse of an orthogonal transformation. Show that if $Q = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$ is orthogonal,

$$Q^{-1} = \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_n \end{bmatrix} .$$

That is Q^{-1} is the matrix whose ith row is the ith column of Q. In other words, the inverse of every orthogonal matrix Q is its transpose matrix.

3 The QR factorization

We now show how to use the Gram-Schmidt orthogonalization procedure to factor any $n \times n$ matrix A in the form $A = QR$ where Q is orthogonal, and R is upper triangular. Let $\mathcal{D}$ be a nice region in $\mathbb{R}^n$. Then

$$\text{vol}(A(\mathcal{D})) = \text{vol}(Q(R(\mathcal{D}))) = \text{vol}(R(\mathcal{D})) \ ,$$

since orthogonal transformations do not affect volume. But we know how to factor R as a product of a diagonal matrix D and a product of shear transformations, and that the shear transformations do not affect volume, so that

$$\text{vol}(R(\mathcal{D})) = \text{vol}(D(\mathcal{D}))$$

and the latter is simply the product on the diagonal entries in D and the volume of the initial region $\mathcal{D}$.

So, by carrying out a QR factorization of a matrix A, and then reading of the diagonal elements of R, and multiplying then together, we can determine the *volume magnification factor* $\rho(A)$ of A: We have that

$$\rho(A) = \rho(R)$$

where $A = QR$ is a QR factorization of A, and we recall that $\rho(R)$ is simply the absolute value of the product of the diagonal elements of R.

It remains to see how to compute a QR factorization. As we noted above, this is essentially the Gramm-Schmidt procedure applied to the columns of A. Let us explain this when A is 3×3:

Let $A = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_2]$. We may assume that all of the vectors are non-zero, or else the image of A lies in a plane and its volume magnification factor is zero.

Define

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{v}_1\|} \ .$$

Note that this can be written as

$$\mathbf{v}_1 = (\mathbf{v}_1 \cdot \mathbf{u}_1)\mathbf{u}_1 \ . \tag{3.1}$$

Next, define,

$$\mathbf{w}_2 = \mathbf{v}_2 - (\mathbf{v}_2 \cdot \mathbf{u}_1)\mathbf{u}_1 \ .$$

By construction $\mathbf{w}_2$ is orthogonal to $\mathbf{u}_1$. If $\mathbf{w}_2 = 0$, then $\mathbf{v}_2$ is a multiple of $\mathbf{u}_1$, and hence of $\mathbf{v}_2$. In this case, the image of A lies in a plane, and the volume magnification factor is zero.

If not, we may divide by $\|\mathbf{w}_2\|$ to define

$$\mathbf{u}_2 = \frac{1}{\|\mathbf{w}_2\|}\mathbf{w}_2 \ ,$$

so that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is orthonormal.

Note that $\|\mathbf{w}_2\|\mathbf{u}_2 = \mathbf{w}_2 = \mathbf{v}_2 - (\mathbf{v}_2 \cdot \mathbf{u}_1)\mathbf{u}_1$, and then since $\|\mathbf{w}_2\| = \mathbf{v}_2 \cdot \mathbf{u}_2$,

$$\mathbf{v}_2 = (\mathbf{v}_2 \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{v}_2 \cdot \mathbf{u}_2)\mathbf{u}_2 \ . \tag{3.2}$$

Proceeding in the same way we determine $\mathbf{u}_3$, and then have (as we would for any orthonormal basis)

$$\mathbf{v}_3 = (\mathbf{v}_3 \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{v}_3 \cdot \mathbf{u}_2)\mathbf{u}_2 + (\mathbf{v}_3 \cdot \mathbf{u}_3)\mathbf{u}_3 \ . \tag{3.3}$$

Now compare (3.1), (3.2) and (3.3). Because of the way the orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is constructed out of $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, when we expand $\mathbf{v}_j$ in terms of $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$, we have non-zero coefficients only for $i < j$, or put differently

$$\mathbf{v}_j \cdot \mathbf{u}_i = 0 \qquad \text{for all } i > j \ . \tag{3.4}$$

We now define

$$Q = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3] \ , \tag{3.5}$$

so that Q is orthogonal.

We now claim that the i, jth entry of R is exactly $\mathbf{u}_i \cdot \mathbf{v}_j$; i.e.,

$$R_{i,j} = \mathbf{u}_i \cdot \mathbf{v}_j \quad \text{for all } i, h \ . \tag{3.6}$$

and this, together with (3.4) shows that R is upper triangular.

To see this, note that if by the definition of A as $A = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]$, for each j,

$$\mathbf{v}_j = A\mathbf{e}_j \ .$$

But, since $\{bu_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal basis, and since by (3.5), $Q\mathbf{e}_i = \mathbf{u}_i$ for $i = 1, 2, 3$.

$$\begin{aligned} \mathbf{v}_j &= (\mathbf{v}_j \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{v}_j \cdot \mathbf{u}_2)\mathbf{u}_2 + (\mathbf{v}_j \cdot \mathbf{u}_3)\mathbf{u}_3 \\ &= (\mathbf{v}_j \cdot \mathbf{u}_1)Q\mathbf{e}_1 + (\mathbf{v}_j \cdot \mathbf{u}_2)Q\mathbf{e}_2 + (\mathbf{v}_j \cdot \mathbf{u}_3)Q\mathbf{e}_3 \\ &= Q\left(\sum_{i=1}^{3}(\mathbf{v}_j \cdot \mathbf{u}_i)\mathbf{e}_i\right) \end{aligned} \tag{3.7}$$

Now, if we *define* R by (3.6), then

$$R\mathbf{e}_j = \sum_{i=1}^{3}(\mathbf{e}_i \cdot R\mathbf{e}_j)\mathbf{e}_i = \sum_{i=1}^{3} R_{i,j}\mathbf{e}_i = \sum_{i=1}^{3}(\mathbf{v}_j \cdot \mathbf{u}_i)\mathbf{e}_i \ ,$$

where the first equality is nothing other than then expansion of $R\mathbf{e}_j$ in terms of the standard basis vectors, and the second is that for any matrix B, $B_{i,j} = \mathbf{e}_i \cdot B\mathbf{e}_j$.

Combining the last two equalities, we finally have that

$$A\mathbf{e}_j = \mathbf{v}_j = QR\mathbf{v}_j \ ,$$

and since this is true for each j, $A = QR$. We have already observed the Q is orthogonal, and by(3.4) and (3.6), R is upper triangular.

• *In summary, to compute a QR factorization of A, apply the Gram-Schmidt procedure to the columns of A, and then use (3.6) to define R.*

If the Gram-Schmidt procedure breaks down because some $\mathbf{v}_j$ is in the span of $\mathbf{u}_1, \ldots, \mathbf{u}_{j-1}$, then simply go on to the next vectors until you run out or find one that is not in the span of $\mathbf{u}_1, \ldots, \mathbf{u}_{j-1}$, and use it to define $\mathbf{u}_j$. In this case, you will get a set of fewer than n orthonormal vectors (working in $\mathbb{R}^n$), but then "flesh it out" by adding in more orthonormal vectors to get a basis in any way you like. However you do it, this gives Q, and then R, given by the same formula, is still upper triangular.

---

9. Let $B := \begin{bmatrix} 2 & 1 & -3 \\ -2 & 0 & -3 \\ 1 & 1 & -1 \end{bmatrix}$.

Compute a QR factorization of B, and determine $\rho(B)$, the volume magnification factor of B.

Now for the real challenge:

10. Let

$$C = \frac{1}{9}\begin{bmatrix} 14 & -5 & 3 & 11 \\ -8 & -1 & -3 & 13 \\ 0 & 0 & 9 & 9 \\ 8 & 28 & 12 & 14 \end{bmatrix} .$$

Compute a QR factorization of C, and determine $\rho(C)$, the volume magnification factor of C.

3.1 Remark. Later, when we have studied determinants, and have learned that $\det(AB) = \det(A)\det(B)$ for all $n \times n$ matrices. we will see that $|\det(A)| = |\det(Q)||\det(R)$,and we will also see that $\det(Q) = \pm 1$ for any orthogonal transformation, and $\det(R)$ is the the product of the diagonal entries for any $n \times n$ upper triangular matrix. This will give us the formula $\rho(A) = |\det(A)|$. However, for $n \geq 3$ especially, the method of computing $\rho(A)$ based on the QR factorization is a very effective way to compute $\rho(R)$. Indeed, the QR factorization provides a good way to compute the determinant of any $n \times n$ matrix.

The QR factorization provides good way to solve linear systems $A\mathbf{x} = \mathbf{b}$: Since the inverse of Q is Q^t, this system is equivalent to

$$R\mathbf{x} = Q^t\mathbf{b} \ .$$

Since R is upper-triangular, this is easily solved by back-substitution.