Quadratic Optimization

This is the last ingredient we'll need for the SVD/PCA.

It is the simplest case of quadratic programming, which is a big subfield of optimization. (Least squares is also a kind of quadratic programming.)

Def: An optimization problem means finding extremal values (the minimum and/or maximum) of a function $f(x_1, x_2, \dots, x_n)$, subject to some constraint on the input $(x_1, x_2, \dots, x_n)$.

In quadratic optimization, we want to extremize a quadratic form $q(x_1, x_2, \dots, x_n)$. But this almost never has a maximum — recall that $q(c \cdot x) = c^2 q(x)$, so if $c = 100000$ then $q(cx)$ is very large! Hence we introduce the constraint $\|x\| = 1$.

Quadratic Optimization Problem:

Find the minimum & maximum values of a quadratic form $q(x_1, x_2, \dots, x_n)$ subject to the constraint $x_1^2 + x_2^2 + \dots + x_n^2 = 1$.

Of course, if $x = (x_1, x_2, \dots, x_n)$ then

$$\boxed{x_1^2 + x_2^2 + \dots + x_n^2 = 1 \iff \|x\| = 1.}$$

As usual, quadratic optimization is relatively easy when q is diagonal.

Eg: Extremize $q(x_1,x_2)=3x_1^2-2x_2^2$ subject to $x_1^2+x_2^2=1$.

Maximum:

$$-2x_2^2 \le 3x_2^2$$

$$q(x_1,x_2)=3x_1^2-2x_2^2 \le 3x_1^2+3x_2^2$$
$$=3(x_1^2+x_2^2)=3$$

So the maximum value is 3; it is attained at $\begin{pmatrix}x_1\\x_2\end{pmatrix}=\pm\begin{pmatrix}1\\0\end{pmatrix}$.

Minimum:

$$3x_1^2 \ge -2x_1^2$$

$$q(x_1,x_2)=3x_1^2-2x_2^2 \ge -2x_1^2-2x_2^2$$
$$=-2(x_1^2+x_2^2)=-2$$

So the minimum value is -2; it is attained at $\begin{pmatrix}x_1\\x_2\end{pmatrix}=\pm\begin{pmatrix}0\\1\end{pmatrix}$.

This trick works whenever q is diagonal:

$$q(x)=\lambda_1x_1^2+\lambda_2x_1^2+\cdots+\lambda_nx_n^2$$

Order the x_i such that $\lambda_1\ge\lambda_2\ge\cdots\ge\lambda_n$.

- The maximum value is λ_1, attained at e_1.
- The minimum value is λ_n, attained at e_n.

$\Big\}$ subject to $\|x\|=1$

(Note that the λ_i could be negative.)

So how to extremize a quadratic form in general?

Orthogonally diagonalize!

Recall: We can write $q(x) = x^T S x$ for a symmetric matrix S. If we orthogonally diagonalize S:

$$S = QDQ^T \qquad D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

and change variables $x = Qy$, then

$$q = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2.$$

Important: since Q has orthonormal columns, it preserves lengths & dot products (L13):

$$\boxed{\|Qy\| = \|y\|} \qquad \boxed{(Qy_1)\cdot(Qy_2) = y_1 \cdot y_2.}$$

In particular, $\|y\| = 1 \iff \|x\| = \|Qy\| = 1$, so this doesn't change our constraint.

Eg: Extremize $q(x_1, x_2) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 - 5x_1x_2$ subject to $x_1^2 + x_2^2 = 1$.

We have $q = x^T S x$ for $S = \begin{pmatrix} 1/2 & -5/2 \\ -5/2 & 1/2 \end{pmatrix}$.

Orthogonally diagonalize: the characteristic polynomial is

$$p(\lambda) = \lambda^2 - \lambda + 6 = (\lambda - 3)(\lambda + 2)$$

We compute an orthonormal eigenbasis:

$$\lambda = 3 \rightsquigarrow \begin{pmatrix} -b \\ a-\lambda \end{pmatrix} = \begin{pmatrix} 5/2 \\ -5/2 \end{pmatrix} \rightsquigarrow w_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$\lambda = -2 \rightsquigarrow \begin{pmatrix} -b \\ a-\lambda \end{pmatrix} = \begin{pmatrix} 5/2 \\ 5/2 \end{pmatrix} \rightsquigarrow w_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

So $S = QDQ^T$ for

$$Q = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \quad D = \begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix}.$$

Setting $x = Qy$ we have $q = 3y_1^2 - 2y_2^2$, so:

- The maximum value is $3 =$ largest eigenvalue.
  It is attained at $y = \pm\begin{pmatrix} 1 \\ 0 \end{pmatrix}$
  $$\Rightarrow x = \pm Q\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \pm w_1 \quad (= 1^{st} \text{ column of } Q)$$
  $$= \text{any unit } 3\text{-eigenvector of } S.$$

- The minimum value is $-2 =$ smallest eigenvalue.
  It is attained at $y = \pm\begin{pmatrix} 0 \\ 1 \end{pmatrix}$
  $$\Rightarrow x = \pm Q\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \pm w_2 \quad (= 2^{nd} \text{ column of } Q)$$
  $$= \text{any unit } (-2)\text{-eigenvector of } S.$$

NB: If x is a unit eigenvector of S with eigenvalue λ then

$$q(x) = x^T S x = x^T(Sx) = x^T(\lambda x) = \lambda x^T x = \lambda x \cdot x$$
$$= \lambda \|x\|^2 = \lambda$$

How to Extremize a Quadratic Form q subject to $\|x\|=1$:

Diagonalize $q(x)$: express it as

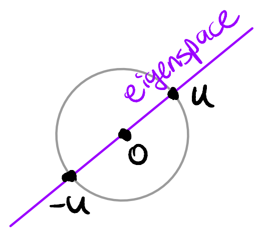
$$q(x) = x^T S x \qquad S = QDQ^T.$$

- The maximum value is the largest eigenvalue of S. It is attained at any unit eigenvector.
- The minimum value is the smallest eigenvalue of S. It is attained at any unit eigenvector.

Conventionally, we choose the largest eigenvalue to be first, and order them decreasing:

$$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \qquad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n.$$

NB: If $GM(\lambda)=1$, then there are only 2 unit eigenvectors $\pm u$: there are only 2 unit vectors on any line.



Otherwise, there are infinitely many! For instance, if the eigenspace is a plane then there's a circle of unit eigenvectors. In fact, if $\{u_1, u_2\}$ is an orthonormal basis of the λ-eigenspace, then any unit λ-eigenvector can be written $u = x_1 u_1 + x_2 u_2$ where $\|u\|^2 = x_1^2 + x_2^2 = 1$.

Additional Constraints

Sometimes the largest or smallest value of q is not very interesting — for example, in the spectral graph theory problem on the homework. We can "rule out" that value by imposing an additional constraint.

"Second-Largest" Value:

Suppose $q(x)$ is maximized (subject to $\|x\|=1$) at u_1. What is the maximum value of $q(x)$ subject to

$$\|x\|=1 \quad \text{and} \quad x \cdot u_1 = 0 \text{ ?}$$

This rules out u_1 (because $u_1 \cdot u_1 = 1$), so we get the "second-largest" value.

(Actually q attains every value in between, too — not on vectors orthogonal to u_1, though — hence the quotes.)

Eg: Find the largest and second-largest values of

$$q(x_1,x_2,x_3) = 2x_1^2 + 2x_2^2 + 5x_3^2 + 2x_1x_2 - 8x_1x_3 + 8x_2x_3$$

subject to $x_1^2 + x_2^2 + x_3^2 = 1$.

Diagonalize q: we have $q(x) = x^T S x$ for

$$S = \begin{pmatrix} 2 & 1 & -4 \\ 1 & 2 & 4 \\ -4 & 4 & 5 \end{pmatrix} = QDQ^T$$

$$Q = \begin{pmatrix} -1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & -1/\sqrt{3} \\ 2/\sqrt{6} & 0 & 1/\sqrt{3} \end{pmatrix} \quad D = \begin{pmatrix} 9 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -3 \end{pmatrix}$$

Change variables to $x = Qy$: then

$$q = 9y_1^2 + 3y_2^2 - 3y_3^2.$$

The largest (maximum) value is 9; it is achieved at $y = \pm\begin{pmatrix}1\\0\\0\end{pmatrix} \Rightarrow x = \pm Q\begin{pmatrix}1\\0\\0\end{pmatrix} = \pm u_1$, $u_1 = \frac{1}{\sqrt{6}}\begin{pmatrix}-1\\1\\2\end{pmatrix}$ (the unit 9-eigenvectors).

What about the second-largest value? This is easy in the y-coordinates: the extra constraint is

$$\begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix} \cdot \begin{pmatrix}1\\0\\0\end{pmatrix} = 0, \text{ which means } y_1 = 0.$$

Then $q(0, y_2, y_3) = 3y_2^2 - 3y_3^2$. This has maximum value 3, achieved at $\pm\begin{pmatrix}0\\1\\0\end{pmatrix}$.

Now we remember that

Q preserves dot products!

Since $u_1 = Q\begin{pmatrix}1\\0\\0\end{pmatrix}$, if $x = Qy$ then

$$u_1\cdot x = \left(Q\begin{pmatrix}1\\0\\0\end{pmatrix}\right)\cdot(Qy) = \begin{pmatrix}1\\0\\0\end{pmatrix}\cdot y$$

Hence $u_1\cdot x = 0 \iff \begin{pmatrix}1\\0\\0\end{pmatrix}\cdot y = 0$. Therefore, the second-largest value is 3, and it is achieved at $\pm Q\begin{pmatrix}0\\1\\0\end{pmatrix} = \pm u_2$ $\quad u_2 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\\0\end{pmatrix}$ (the unit 3-eigenvectors).

This same procedure works for any q.

How to Find the "Second-Largest" Value of q:

Diagonalize $q(x)$: express it as

$$q(x) = x^T S x \qquad S = QDQ^T.$$

Put the eigenvalues in decreasing order:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n.$$

Let u_1 be a unit λ_1-eigenvector (so q is maximized at u_1).

The largest value of $q(x)$ subject to

$$\|x\| = 1 \quad \text{and} \quad x\cdot u_1 = 0$$

is λ_2; it is achieved at

any unit λ_2-eigenvector $\perp u_1$.

NB: Let u_2 be a unit λ_2-eigenvector. Then $u_1 \cdot u_2 = 0$ automatically, unless $\lambda_1 = \lambda_2$ (ie, $\lambda_1 = \lambda_2$ has multiplicity ≥ 2). More on this on the HW.

Third-Largest Value, Etc:

Assume the eigenvalues are in decreasing order:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

Let u_i be a unit λ_i-eigenvalue.

- The largest value subject to

  $$\|x\|=1 \quad x \cdot u_1 = 0 \quad x \cdot u_2 = 0$$

  is λ_3; it is achieved at any unit λ_3-eigenvector. ("third-largest value")

- The smallest value subject to

  $$\|x\|=1 \quad x \cdot u_n = 0$$

  is λ_{n-1}; it is achieved at any unit λ_{n-1}-eigenvector. ("second-smallest value")

etc.

Quadratic Optimization for $q(x) = \|Ax\|^2$

This is what we'll use in the PCA.

Let A be any matrix, $S = A^TA$, $q(x) = x^TSx$.
Then

$$q(x) = x^TSx = x^T(A^TA)x = (x^TA^T)(Ax)$$
$$= (Ax)^T(Ax) = (Ax)\cdot(Ax) = \|Ax\|^2.$$

$$\boxed{q(x) = \|Ax\|^2 \text{ is the quadratic form for } S = A^TA}$$

This means we can extremize $\|Ax\|^2$ subject to $\|x\| = 1$.

Recall: The matrix $S = A^TA$ is positive-semidefinite.
It's even positive-definite if A has FCR.
Indeed, if $Sx = \lambda x$ and $\|x\| = 1$ then

$$\|Ax\|^2 = x^TA^TAx = x^TSx = x^T\lambda x$$
$$= \lambda x^Tx = \lambda\|x\|^2 = \lambda.$$

So $\lambda = \|Ax\|^2 \geq 0$. (See L21.)

How to Extremize $q(x) = \|Ax\|^2$:

Orthogonally diagonalize $S = A^TA$. Put the eigenvalues of S in decreasing order:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

Let u_i be a unit λ_i-eigenvalue.

- The largest value of $\|Ax\|^2$ subject to $\|x\|=1$ is λ_1; it is achieved at any unit λ_1-eigenvector
- The smallest value of $\|Ax\|^2$ subject to $\|x\|=1$ is λ_n; it is achieved at any unit λ_n-eigenvector
- The largest value of $\|Ax\|^2$ subject to $\|x\|=1$ and $x \cdot u_1 = 0$ is λ_2; it is achieved at any unit λ_2-eigenvector ("second-largest value")

Likewise for third-largest, second-smallest, etc.

NB: These are eigenvalues and eigenvectors of S, not of A. Indeed, A need not be a square matrix!

The largest value of $\|Ax\|$ subject to $\|x\|=1$ has a name.

Def: The matrix norm of A is the maximum value of $\|Ax\|$ subject to $\|x\|=1$.

So $\|Ax\| = \sqrt{\lambda_1}$, where $\lambda_1 \geq 0$ is the largest eigenvalue of $S = A^T A$. It is achieved at any unit λ_1-eigenvector (of S).

Eg: Compute $\|A\|$ for $A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}$.

In this case, $S = A^T A = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}$.

The characteristic polynomial is

$$p(\lambda) = \lambda^2 - 6\lambda + 5 = (\lambda - 5)(\lambda - 1)$$

So $\|A\| = \sqrt{5}$ because 5 is the largest eigenvalue. A unit eigenvector is

$$\begin{pmatrix} -b \\ a-\lambda \end{pmatrix} = \begin{pmatrix} -2 \\ -2 \end{pmatrix} \longrightarrow u_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Check:

$$Au_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 2 \\ 2 \\ 1 \end{pmatrix}$$

$$\|Au_1\| = \left\| \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 2 \\ 2 \\ 1 \end{pmatrix} \right\| = \frac{1}{\sqrt{2}} \sqrt{1+4+4+1} = \frac{\sqrt{10}}{\sqrt{2}} = \sqrt{5} \checkmark$$