# MINIMIZATION PROBLEMS WITH LINEAR AND QUADRATIC FORMS TOGETHER

Let $Q$ be a symmetric matrix. We say that $Q$ is ***positive definite*** if $\vec{v}^T Q \vec{v} > 0$ for all nonzero vectors $\vec{v}$. Recall that a symmetric matrix $Q$ is positive definite if and only if all of its eigenvalues are nonnegative.

One of the extremely convenient things about a positive definite matrix is that we can choose a basis in which it is just the standard length-squared form you are used to on $\mathbb{R}^n$. Specifically, let $Q$ be a positive definite matrix. Let $\vec{v}_1, \ldots, \vec{v}_n$ be an orthonormal eigenbasis for $Q$, with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$.

Let $S$ be the matrix whose rows are $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$. As we have discussed several times, we have

$$Q = S^T \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} S.$$

Note that we can rewrite this as

$$Q = R^T R \quad \text{where } R = \begin{pmatrix} \sqrt{\lambda_1} & & & \\ & \sqrt{\lambda_2} & & \\ & & \ddots & \\ & & & \sqrt{\lambda_n} \end{pmatrix} S = \begin{pmatrix} - & \sqrt{\lambda_1}\vec{v}_1 & - \\ - & \sqrt{\lambda_2}\vec{v}_2 & - \\ \vdots & \vdots & \vdots \\ - & \sqrt{\lambda_n}\vec{v}_n & - \end{pmatrix}$$

So, $\vec{x}^T Q \vec{x} = \vec{x}^T R^T R \vec{x} = (R\vec{x})^T (R\vec{x})$. In other words, if $R\vec{x} = \vec{y}$, then $\vec{x}^T Q \vec{x} = \vec{y}^T \vec{y} = |y|^2$. The matrix $R$ turns the funny quadratic form $Q$ into the length-squared which we have geometric intuitions for.

In this note, I'll discuss some cases where I find this a useful trick. All of the problems I discuss can also be solved in other ways. In particular, those of you who know multivariable calculus will find all of these problems vulnerable to it.

## 1. Minimizing a quadratic form plus lower order terms

In high school algebra, you learn how to compute the minimum value of $ax^2 + bx + c$ by completing the square: Writing $ax^2 + bx + c = a(x + b/(2a))^2 + c - b^2/(4a)$, we see that the minimum is $c - b^2/(4a)$, and is achieved at $x = -b/(2a)$. All of this is for $a$ positive; if $a$ is negative then the parabola curves down and there is no minimum value.

The corresponding problem in linear algebra is to minimize $\vec{x}^T Q \vec{x} + \vec{b}^T \vec{x} + c$ as a function of the length $n$ vector $\vec{x}$. Here $Q$ is an $n \times n$ positive definite matrix; $\vec{b}$ is a vector of length $n$ and $c$ is a scalar.

We make the change of variables $\vec{y} = R\vec{x}$. So

$$\vec{x}^T Q \vec{x} + \vec{b}^T \vec{x} + c = \vec{y}^T \vec{y} + \vec{b}^T R^{-1} \vec{y} + c$$

We can complete the square.

$$= (\vec{y}^T + \frac{1}{2}\vec{b}^T R^{-1})(\vec{y} + \frac{1}{2}(R^{-1})^T\vec{b}) + c - \frac{1}{4}\vec{b}^T R^{-1}(R^T)^{-1}\vec{b}$$
$$= |\vec{y} + \frac{1}{2}(R^{-1})^T\vec{b}|^2 + c - \frac{1}{4}\vec{b}^T R^{-1}(R^T)^{-1}\vec{b}$$

We clean up a little by noticing that $R^{-1}(R^T)^{-1} = (R^T R)^{-1} = Q^{-1}$.

$$= |\vec{y} + \frac{1}{2}(R^{-1})^T\vec{b}|^2 + c - \frac{1}{4}\vec{b}^T Q^{-1}\vec{b}$$

So the minimal value is $c - \frac{1}{4}\vec{b}^T Q^{-1}\vec{b}$, generalizing the single variable formula $c - \frac{1}{4}b^2/a$. The minimum value occurs when $\vec{y} = -\frac{1}{2}(R^{-1})^T\vec{b}$. In the original coordinates, the minimum is at

$$\vec{x} = R^{-1}\vec{y} = -\frac{1}{2}R^{-1}(R^T)^{-1}\vec{b} = -\frac{1}{2}(R^T R)^{-1}\vec{b} = -\frac{1}{2}Q^{-1}\vec{b}.$$

This is the generalization of the single variable formula $-\frac{1}{2}b/a$.

## 2. Minimizing a quadratic form restricted to linear conditions

Consider a subset of $\mathbb{R}^n$ that looks like $\vec{p} + V$ for some subspace $d$-dimensional $V$ of $\mathbb{R}^n$. We might want to minimize the function $Q$ on the space $\vec{p} + V$. Writing $V$ as the image of some $n \times d$ matrix $A$. Then we want to minimize

$$(\vec{p} + A\vec{z})^T Q(\vec{p} + A\vec{z}) = \vec{z}^T A^T QA\vec{z} + 2(Q\vec{p})^T A\vec{z} + \vec{p}^T Q\vec{p}$$

as a function of $\vec{z}$. This is exactly a quadratic function of the sort we discussed above, with $(Q, \vec{b}, c)$ replaced by $(A^T QA, 2Q\vec{p}, \vec{p}^T Q\vec{p})$. But it's nicer not to go through this.

As before, write $Q = R^T R$. We want to minimize $\vec{x}^T Q\vec{x}$ subject to $\vec{x} = \vec{p} + A\vec{z}$ for some $\vec{z}$. As before, we make the change of variables $\vec{y} = R\vec{x}$. So we want to minimize $|y|^2$ subject to $\vec{y} = R\vec{p} + RA\vec{z}$. In other words, we want $\vec{y}$ to be the smallest vector on $R\vec{p}+\mathrm{Image}(RA)$. Alternatively, we want $\vec{y} - \vec{p}$ to be the vector on $\mathrm{Image}(RA)$ closest to $-\vec{p}$. Writing $\vec{r}$ for the orthogonal projection of $\vec{r}$ onto $\mathrm{Image}(RA)$, we want $\vec{y} = \vec{p} + \vec{r}$ and $\vec{x} = R^{-1}(\vec{p} + \vec{r})$.

Often, instead of being given our linear constraint in the form $\vec{x} = \vec{p} + A\vec{y}$, we are given a collection of inhomogenous linear equations $B\vec{x} = \vec{q}$ for some matrix $B$ and vector of constants $\vec{q}$. Of course, by row reduction, we can convert equations of the form $B\vec{x} = \vec{q}$ into equations of the form $\vec{p} + A\vec{y}$. But there is a slicker way.

One more time, put $Q = R^T R$ and $\vec{y} = R\vec{x}$. We want to minimize $\vec{y}^2$ subject to $BR^{-1}\vec{y} = \vec{q}$. The space of $\vec{y}$ such that $BR^{-1}\vec{y} = \vec{q}$ is a translate of $\mathrm{Ker}(BR^{-1})$. The smallest vector on the linear constraint $BR^{-1}\vec{y} = \vec{q}$. will therefore be in a direction perpendicular to $\mathrm{Ker}(BR^{-1})$. In equations, $\vec{y}$ is in $\mathrm{Ker}(BR^{-1})^\perp$. But we know that $\mathrm{Ker}(BR^{-1})^\perp = \mathrm{Image}((^{-1})^T) = \mathrm{Image}((R^T)^{-1}B^T)$. So $\vec{y} = (R^T)^{-1}B^T\vec{z}$ for some $\vec{z}$. We can find $\vec{z}$ by solving the equation

$$BR^{-1}(R^T)^{-1}B^T\vec{z} = \vec{q} \text{ or, in other words } BQ^{-1}B^T\vec{z} = \vec{q}.$$

Putting it all together,

$$\vec{x} = R^{-1}\vec{y} = R^{-1}(R^T)^{-1}B^T\vec{z} = Q^{-1}B^T\vec{z} = Q^{-1}B^T(BQ^{-1}B^T)^{-1}\vec{q}.$$