Holger Broman, August 19, 1999

Quadratic forms

It is part of human nature that we sometimes want to do something the best way possible. In a scientific context, this requires that we know exactly what 'the best way possible' means. For instance, consider estimating the parameters of a mathematical model for some real-life system. Then we need to formulate a criterion of optimality, which results in a single real number. Now, we are in a position to state explicitly that best means maximizing the value of the criterion function with respect to all possible choices of the numerical values of the parameters of the model.

The generic formulation is as follows. The criterion function V is a mapping

$$V: \qquad \mathbb{R}^N \ni \mathbf{x} \overset{V}{\curvearrowright} y \in \mathbb{R}^1$$

Note 1: In system identification and signal processing we often use $\mathbf{\Theta}$ for the parameter vector.

Note 2: If the criterion function denotes a cost, we want to find the global minimum.

Now, a necessary condition for the existence of a maximum or minimum of a differentiable V is that *the gradient* is zero. The gradient of V is a vector where we have stacked all the first order derivatives of V with respect to the components of $\mathbf{x}$:

$$\boldsymbol{\psi} = \frac{\partial V}{\partial \mathbf{x}} = \left[\frac{\partial V}{\partial x_1} \quad \cdots \quad \frac{\partial V}{\partial x_N}\right]^T$$

Note 3: It is easily shown that $\boldsymbol{\psi} = 0$ is not sufficient to state that we have found a maximum or minimum. Take as a counterexample

$$V(x) = x^3,$$

where $x \in \mathbb{R}^1$. The problem is that the second derivative vanishes at the origin as well.

To form a sufficient condition for the existence of a maximum or minimum of V, we have to study the second order derivatives. These can be placed in a matrix H as follows:

$$H = \frac{\partial^2 V}{\partial \mathbf{x}^2} = \begin{bmatrix} \frac{\partial^2 V}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 V}{\partial x_1 \partial x_N} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ \frac{\partial^2 V}{\partial x_N \partial x_1} & \cdots & \frac{\partial^2 V}{\partial x_N \partial x_N} \end{bmatrix}.$$

This matrix is called the Hessian of the criterion function V.

Now we have set the first stage of introducing quadratic forms. The second part is to notice that for 'nice' mappings V, we can expand V in a Taylor series

$$\begin{aligned} V(\mathbf{x}_0+\Delta\mathbf{x}) &= V(\mathbf{x}_0) + \boldsymbol{\psi}^T(\mathbf{x}_0)\Delta\mathbf{x} + \\ &+\frac{1}{2}(\Delta\mathbf{x})^T H(\mathbf{x}_0)\Delta\mathbf{x} + \text{ higher order terms.} \end{aligned}$$

Noting that a stationary point $\mathbf{x}_0$ requires $\boldsymbol{\psi}(\mathbf{x}_0) = 0$ and rearranging terms we arrive at

$$2\Big[V(\mathbf{x}_0 + \Delta\mathbf{x}) - V(\mathbf{x}_0)\Big] \approx (\Delta\mathbf{x})^T \; H(\mathbf{x}_0) \; \Delta\mathbf{x}\,.$$

The right hand side is a quadratic form and contains all information needed to determine the character of the behavior of V close to $\mathbf{x}_0$ (unless H is the null matrix). In particular, we can tell whether V has a maximum, a minimum or neither.

Definition: A (purely) quadratic form is the multivariate purely second degree polynomial $\mathbf{x}^T A\,\mathbf{x}$, where $\mathbf{x} \in \mathbb{R}^N$ and A is $N \times N$.

Note 4: As

$$\mathbf{x}^T A\,\mathbf{x} = \left(\mathbf{x}^T A\,\mathbf{x}\right)^T = \mathbf{x}^T A^T\,\mathbf{x}$$

we find

$$\mathbf{x}^T A\,\mathbf{x} = \mathbf{x}^T \left(\frac{A + A^T}{2}\right)\,\mathbf{x},$$

and in the sequel we will always assume that A is symmetric. This is no restriction!

Note 5: To simplify the notation we have translated the original problem by $-\mathbf{x}_0$ so that

$$\boldsymbol{\psi}(0) = 0,$$

and by $-V(0)$ so that

$$V(0) = 0.$$

Examples: Here are some quadratic forms:

$$V(x) = x^2$$

$$V(x,y) = x^2 + 2y^2 - 3xy$$

$$V(\mathbf{x}) = \sum_{i=1}^{N}\sum_{j=1}^{i} a_{ij}\; x_i\; x_j$$

Let us thus study

$$V(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x}$$

in some neighborhood, D, to the origin. $V(\mathbf{x})$ has a local minimum at the origin iff

$$V(\mathbf{x}) > 0 \;\; \forall \;\; \mathbf{x} \in D, \;\; \mathbf{x} \neq 0.$$

Recall the eigenvalue decomposition of symmetric matrices

$$A = U \; \Lambda \; U^T,$$

where U is orthogonal, and the eigenvalues of A are found in the diagonal matrix Λ. With

$$\mathbf{y} = U^T \, \mathbf{x}$$

we find

$$V(\mathbf{y}) = \mathbf{y}^T \; \Lambda \; \mathbf{y},$$

and the neighborhood D transforms into some other neighborhood $D_{\mathbf{y}}$. We conclude

$$V(\mathbf{x}) > 0 \;\; \forall \;\; \mathbf{x} \in D, \;\; \mathbf{x} \neq 0$$

$$\iff$$

$$V(\mathbf{y}) > 0 \;\; \forall \;\; \mathbf{y} \in D_{\mathbf{y}}, \;\; \mathbf{y} \neq 0$$

$$\iff$$

All eigenvalues of A are positive.

Symmetric matrices with only positive eigenvalues are called positive definite.

To summarize:

$V(\mathbf{x})$ has a local minimum at $\mathbf{x} = \mathbf{x}_0$ iff • the gradient of V is zero at $\mathbf{x}_0$ • all eigenvalues of the Hessian of V are positive at $\mathbf{x}_0$

Exercise: Formulate the theorem that concerns local maxima.

Note 6: A positive semidefinite matrix is one with non-negative eigenvalues, also called non-negative definite. If you find such a Hessian, you cannot guarantee a local minimum. Rather V has a "valley" which is "horizontal" and runs in the direction of the eigenvector that has zero as its eigenvalue.

Note 7: A matrix with both positive and negative eigenvalues is called indefinite. Such a Hessian would correspond to a saddle point in two dimensions ($\mathbf{x}$ is two-dimensional).

Note 8: Optimization can be quite tricky, and we have only touched upon one principle. To find the global maximum of V, you must typically find all local maxima, and there can be many, and choose the largest. To find a local maximum you must typically start a numerical search within the "radius of attraction" of that maximum. That radius might be quite small. The difficulties are reflected in the number of algorithms designed:

gradient search
Gauss-Newton search
genetic algorithms
neural networks
purely numerical search
etcetera

To reiterate, if you want to find a global minimum, it is in general not enough to search for stationary points. This is the reason for the many different algorithms.

Please run the m-file quadform in Matlab.

```
% quadform.m
% illustration of quadratic forms
clear

% Generate the data
x1=-1:.02:1;
x2=x1;
x=[x1' x2'];
A=[1 0 ; 0 1];
N=length(x1);
for n1=1:N
   for n2=1:N
data(n1,n2)=[x1(n1) x2(n2)]*A*[x1(n1) x2(n2)]';
   end
end

% plot it
mesh(x1,x2,data)

% use in the above various matrices.
% [1 0 ; 0 1] is pos def
% [1 0 ; 0 0] is pos semidef
% [1 0 ; 0 -1] is indef
% [-1 0 ; 0 0 ] is neg semidef
% [-1 0 ; 0 -1] is neg def
% Compare the behaviour of the graphs with the eigenvalues of the matrix A.
% use the matlab command eig(A)
```