How to Recognize a Quadratic Form

What is a *Quadratic Form* ? It is a scalar-valued functional $f(\mathbf{v}) = \mathbf{¥vv}$ obtained from a *Symmetric Bilinear Form* $\mathbf{¥uv}$, a functional that is both *symmetric*, $\mathbf{¥uv} = \mathbf{¥vu}$, and *linear* : $\mathbf{¥u}(\mu\mathbf{v} + ß\mathbf{w}) = \mu\mathbf{¥uv} + ß\mathbf{¥uw}$ for all real scalars μ and ß and all vectors **u**, **v** and **w** in a real vector space. If column vectors u and v represent vectors $\mathbf{u} = \mathbf{B}u$ and $\mathbf{v} = \mathbf{B}v$ in some basis **B** , then $\mathbf{¥uv} = u^TYv = v^TYu$ for some symmetric matrix $Y = Y^T$, and $f(\mathbf{v}) = v^TYv$.

Symmetric bilinear operator ¥ maps the space of vectors **u** and **v** linearly to its dual space:

¥u_ is a linear functional in the space dual to vectors **v** , and $\mathbf{¥uv} = \mathbf{¥vu}$ is its scalar value;

¥_v is a linear functional in the space dual to vectors **u** , and $\mathbf{¥uv} = \mathbf{¥vu}$ is its scalar value.

(See the notes on "Least Squares Approximation and Bilinear Forms".) Matrix Y represents this operator ¥ in the basis **B** , representing **¥u_** by u^TY and **Y_v** by v^TY . Changing **B** to a new basis $\mathbf{B}C^{-1}$ changes **u** 's representative u to Cu in order to keep $\mathbf{u} = \mathbf{B}C^{-1}(Cu)$, and similarly changes v to Cv ; and then, to keep $\mathbf{¥uv} = (Cu)^T(C^{T-1}YC^{-1})(Cv)$ unchanged, Y changes to $C^{T-1}YC^{-1}$. In other words, the *Congruent* matrices Y and $C^{T-1}YC^{-1}$ represent the same symmetric bilinear form ¥ and the same quadratic form $f(\ldots)$ in different coordinate systems (bases).

A quadratic form $f(\mathbf{v}) = \mathbf{Zvv}$ is obtainable from any *non-symmetric* bilinear form $\mathbf{Zuv} \neq \mathbf{Zvu}$ represented by a nonsymmetric matrix Z thus: $f(\mathbf{v}) = v^TZv$ even though $u^TZv \neq v^TZu$. No good purpose is served this way; $f(\ldots)$ depends upon only the *symmetric* part ¥ of **Z** defined thus: $\mathbf{¥uv} := (\mathbf{Zuv} + \mathbf{Zvu})/2$ and so $Y = (Z + Z^T)/2$. Moreover, only the symmetric bilinear form ¥ can be recovered from $f(\ldots)$ via the identity $\mathbf{¥uv} = (f(\mathbf{u}+\mathbf{v}) - f(\mathbf{u}-\mathbf{v}))/4$.

Exercise: Suppose Q is a matrix (perhaps neither square nor real) that satisfies $(Qx)^T(Qx) = x^Tx$ for all *real* column vectors x of the right dimension. Must $Q^TQ = I$? Why?

Another way to recover a symmetric bilinear form from its quadratic form is by differentiation: $df(\mathbf{v}) = f'(\mathbf{v})d\mathbf{v} = 2\mathbf{¥v}d\mathbf{v}$. In other words, the derivative of a quadratic form $f(\mathbf{v})$ is a linear functional $f'(\mathbf{v})$ (of course) that is also linear in the form's argument **v** and symmetric in so far as $f'(\mathbf{v})\mathbf{u} = f'(\mathbf{u})\mathbf{v}$. In matrix terms $f'(\mathbf{v}) = 2v^TY$ and $f'(\mathbf{v})\mathbf{u} = 2v^TYu = 2u^TYv = f'(\mathbf{u})\mathbf{v}$. Conversely if $y(\mathbf{v})$ satisfies $y(\mathbf{o}) = 0$ and if its derivative $y'(\mathbf{v}) = 2\mathbf{¥v}$ is linear in **v** , then $y''(\mathbf{v}) = 2\mathbf{¥}$ is a constant symmetric bilinear operator (symmetric because all continuous second derivatives are symmetric) and $y(\mathbf{v}) = \int_{\mathbf{o}}^{\mathbf{v}} y'(\mathbf{u})d\mathbf{u} = \mathbf{¥vv}$ is a quadratic form. In matrix terms, $y'(\mathbf{v})\mathbf{u} = 2v^TYu$ is linear in v , $y''(\mathbf{v})\mathbf{uw} = 2w^TYu = 2u^TYw = y''(\mathbf{v})\mathbf{wu}$ because $Y = Y^T$ must be symmetric, and $y(\mathbf{v}) = \int_o^v 2u^TYdu = v^TYv$ regardless of the path of integration. In short, a quadratic form can be recognized as such by determining whether its derivative depends linearly on the form's argument.

... ...

Quadratic forms can be defined over complex vector spaces, but in two ways. There are complex quadratic forms that take complex scalar values algebraically the same as above but quite different from what follows. There are real-valued quadratic forms very much like what follows but each obtained from an *Hermitian* bilinear form **Huv** that is linear in one argument, say **u** , and *Conjugate-Linear* in the other: $\mathbf{Hu}(\mu\mathbf{v} + ß\mathbf{w}) = \bar{\mu}\mathbf{Huv} + \bar{ß}\mathbf{Huw}$ where $\bar{\mu}$ and $\bar{ß}$ are the complex conjugates of μ and ß . Bilinear operator **H** is *Hermitian* just when $\mathbf{Hvu} = \overline{\mathbf{Huv}}$. In matrix terms, $\mathbf{Huv} = v^*Hu$ with an Hermitian matrix $H = H^*$, its complex-conjugate transpose; H is congruent to $C^{*-1}HC^{-1}$. The Hermitian quadratic form $f(\mathbf{v}) = \mathbf{Hvv} = v^*Hv$ obtained from **H** is real for all complex **v** and consequently slightly more complicated than real quadratic forms on real spaces, to which this note is confined.

Every quadratic form $f(\ldots)$ satisfies an identity called the *Parallelogram Law* :

$$f(\mathbf{x}+\mathbf{y}) + f(\mathbf{x}-\mathbf{y}) = 2f(\mathbf{x}) + 2f(\mathbf{y}) .$$

This law is easy to deduce when $f(\ldots)$ is obtained from a given bilinear form **¥** ; do so! The law gets its name from the case $\sqrt{f(\mathbf{v})} = \|\mathbf{v}\|$ of ordinary length defined in an Euclidean space by the Pythagorean formula. Conversely …

Theorem: Any real continuous scalar function $f(\ldots)$ that satisfies the parallelogram law for all vectors $\mathbf{x}$ and $\mathbf{y}$ in a real linear space must be a quadratic form.

This seems plausible; setting $\mathbf{x} = \mathbf{y} = \mathbf{o}$ implies $f(\mathbf{o}) = 0$, and then setting $\mathbf{x} = \mathbf{o}$ and letting $\mathbf{y} \longrightarrow \mathbf{o}$ in $f(\mathbf{y}) - f(-\mathbf{y}) = 0$ implies $f'(\mathbf{o}) = \mathbf{o}^T$ if the derivative exists, and then doing it again with $f(\mathbf{x}+\mathbf{y}) - 2f(\mathbf{x}) + f(\mathbf{x}-\mathbf{y}) = 2f(\mathbf{y})$ suggests that $f''(\mathbf{x})$ is independent of $\mathbf{x}$. But a suggestion is not a proof. The Theorem's proof found by C. Jordan and J. von Neumann early in the 20th century is unusual enough to be worth reproducing here. First we need a …

Lemma: If a continous scalar functional $\varsigma(\mathbf{x})$ satisfies $\varsigma(\mathbf{o}) = 0$ and $\varsigma(\mathbf{x}+\mathbf{y}) + \varsigma(\mathbf{x}-\mathbf{y}) = 2\varsigma(\mathbf{x})$ for all vectors $\mathbf{x}$ and $\mathbf{y}$ in a real linear space, $\varsigma(\mathbf{x})$ must be a linear functional $\varsigma(\mathbf{x}) = \mathbf{c}^T\mathbf{x}$.

Proof: Start by discovering that

$$\begin{aligned}
\varsigma(\mathbf{x}+\mathbf{y}) &= \varsigma(\,(\mathbf{x}+\mathbf{y})/2 + (\mathbf{x}+\mathbf{y})/2\,) + \varsigma(\,(\mathbf{x}+\mathbf{y})/2 - (\mathbf{x}+\mathbf{y})/2\,) && \ldots \text{ since } \varsigma(\mathbf{o}) = 0\\
&= 2\varsigma(\,(\mathbf{x}+\mathbf{y})/2\,) && \ldots \text{ by hypothesis}\\
&= \varsigma(\,(\mathbf{x}+\mathbf{y})/2 + (\mathbf{x}-\mathbf{y})/2\,) + \varsigma(\,(\mathbf{x}+\mathbf{y})/2 - (\mathbf{x}-\mathbf{y})/2\,) && \ldots \text{ by hypothesis}\\
&= \varsigma(\mathbf{x}) + \varsigma(\mathbf{y}) \quad \text{for all vectors } \mathbf{x} \text{ and } \mathbf{y} .
\end{aligned}$$

Next, for positive integers n = 1, 2, 3, ... in turn, use this discovery to verify by induction that

$$\varsigma(n\mathbf{x}) = \varsigma((n-1)\mathbf{x}) + \varsigma(\mathbf{x}) = (n-1)\varsigma(\mathbf{x}) + \varsigma(\mathbf{x}) = n\varsigma(\mathbf{x}) .$$

Then from $0 = \varsigma(n\mathbf{x} - n\mathbf{x}) = \varsigma(n\mathbf{x}) + \varsigma(-n\mathbf{x})$ verify that $\varsigma(-n\mathbf{x}) = -n\varsigma(\mathbf{x})$ for all $\mathbf{x}$. And from $\varsigma(\mathbf{x}) = \varsigma(n\mathbf{x}/n) = n\varsigma(\mathbf{x}/n)$ infer that $\varsigma(\mathbf{x}/n) = \varsigma(\mathbf{x})/n$. Similarly $\varsigma((m/n)\mathbf{x}) = (m/n)\varsigma(\mathbf{x})$ for all rational m/n . Since every real number μ is a limit of rational numbers and since $\varsigma(\ldots)$ is continuous, $\varsigma(\mu\mathbf{x}) = \mu\varsigma(\mathbf{x})$ for every real μ and every vector $\mathbf{x}$. This ensures that $\varsigma(\mu\mathbf{x}+ß\mathbf{y}) = \mu\varsigma(\mathbf{x})+ß\varsigma(\mathbf{y})$, confirming that $\varsigma(\ldots)$ must be a linear functional; $\varsigma(\mathbf{x}) = \mathbf{c}^T\mathbf{x}$ for some $\mathbf{c}^T$. End of lemma's proof.

To prove the Theorem take any given scalar function $f(...)$ that satisfies the parallelogram law above and from $f(\ldots)$ construct the functional $Ç(\mathbf{x}, \mathbf{y}) := (\,f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x}-\mathbf{y})\,)/4$. Evidently $Ç(\mathbf{x}, \mathbf{o}) = 0$; and $Ç(\mathbf{x}, \mathbf{x}) = f(2\mathbf{x})/4 = f(\mathbf{x})$ and $Ç(\mathbf{y}, \mathbf{x}) = Ç(\mathbf{x}, \mathbf{y})$ via the Parallelogram Law. Moreover

$$\begin{aligned}
4Ç(\mathbf{z}, \mathbf{x}+\mathbf{y}) + 4Ç(\mathbf{z}, \mathbf{x}-\mathbf{y}) &= f(\mathbf{z}+\mathbf{x}+\mathbf{y}) - f(\mathbf{z}-\mathbf{x}-\mathbf{y}) + f(\mathbf{z}+\mathbf{x}-\mathbf{y}) - f(\mathbf{z}-\mathbf{x}+\mathbf{y})\\
&= 2f(\mathbf{z}+\mathbf{x}) + 2f(\mathbf{y}) - 2f(\mathbf{z}-\mathbf{x}) - 2f(\mathbf{y}) \quad \ldots \text{ Parallelogram Law}\\
&= 8Ç(\mathbf{z}, \mathbf{x})
\end{aligned}$$

for all vectors $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$. Now identify $Ç(\mathbf{z}, \mathbf{x})$ with the $\varsigma(\mathbf{x})$ of the lemma to deduce that $Ç(\ldots)$ is a linear functional of its second argument and, because $Ç(\mathbf{x}, \mathbf{y}) = Ç(\mathbf{y}, \mathbf{x})$, also a linear functional of its first. In short, $Ç(\ldots)$ is a *symmetric bilinear functional* **¥**… of its two arguments; in any coordinate system in which $\mathbf{x}$ and $\mathbf{y}$ are represented by column vectors x and y , we conclude $Ç(\mathbf{x}, \mathbf{y}) = y^TYx$ for some real symmetric matrix $Y = Y^T$, as claimed.

Exercise: How do **¥**… and the coordinate system (basis) determine the matrix Y ?