Research, 7 (1982) 40–56.

8. J. F. Nash, The bargaining problem, Econometrica, 18 (1950) 155–162.

9. J. S. Ransmeier, The Tennessee Valley Authority: A Case Study in the Economics of Multiple Purpose Stream Planning, Vanderbilt Univ. Press, Nashville, Tenn., 1942.

10. L. S. Shapley, A value for n-person games, in H. W. Kuhn and A. W. Tucker (Editors), Contributions to the Theory of Games, vol. II, Princeton Univ. Press, Princeton, N.J., 1953, pp. 307–317.

11. W. W. Sharkey, Suggestions for a game-theoretic approach to public utility pricing, Bell Laboratories Economic Discussion Paper 61, 1974.

12. M. Shubik, Incentives, decentralized control, the assignment of joint costs and internal pricing, Management Science, 8 (1962) 325–343.

13. A. L. Thomas, A Behavioral Analysis of Joint-Cost Allocation and Transfer Pricing, Stipes Publishing Co., 1980.

14. H. P. Young, Cost allocation and demand revelation in public enterprises, International Institute of Applied Systems Analysis Working Paper WP-80-130, 1980.

DO SYMMETRIC PROBLEMS HAVE SYMMETRIC SOLUTIONS?

WILLIAM C. WATERHOUSE

Department of Mathematics, The Pennsylvania State University, University Park, PA 16802

There are many problems where the function under study is symmetric in several variables and its maximum or minimum occurs when the variables are equal. Here are some examples:

(1) Of all rectangles with given perimeter, the square has the largest area.

(2) The triangle of minimum area circumscribed about a given circle is equilateral.

(3) Take four positive numbers whose product is 16. Then their sum is at least 8. (That is, the sum is least when all four numbers are equal.)

(4) A box is to be made from sheet metal, with rectangular bottom and sides but no top. Then the design using least material has a square bottom and height half the width. (To reduce this to a symmetric situation, put another copy of the box upside down on top of it.)

(5) Determine the minimum value of

$$(r-1)^2+\left(\frac{s}{r}-1\right)^2+\left(\frac{t}{s}-1\right)^2+\left(\frac{4}{t}-1\right)^2$$

for all real numbers r, s, t with $1 \leqslant r \leqslant s \leqslant t \leqslant 4$. (This is a problem from the 1981 Putnam Competition; the minimum comes when $(r, s, t) = (\sqrt{2}, 2, 2\sqrt{2})$. The problem becomes symmetric if we introduce new variables r, s/r, t/s, and $4/t$.)

(6) For a given mean $\bar{x} = (\Sigma_1^n x_i)/n$, the value $\Sigma_1^n x_i^2$ is least when all x_i are equal.

(7) For any positive $x_1, \dots, x_n$ we have $(\Sigma_1^n x_i)/n \geqslant (x_1 \cdots x_n)^{1/n}$ (the inequality of arithmetic and geometric means, a generalization of (3) above).

It seems clear that some general principle must be lurking here, and I propose to call it the Purkiss Principle. This name honors one of the authors who (independently) noticed this principle in the middle of the last century. Their discussions were somewhat inconclusive and seem to have been completely forgotten; I have not come across any recent references to the idea. Probably this is because the principle is not true without qualifications. But still, there is something to it.

William C. Waterhouse studied at Harvard, receiving his Ph.D. in 1968 under the guidance of John Tate. He taught at Cornell until 1975, when he moved to Penn State. His research has been mainly in algebra, algebraic geometry, and algebraic number theory. He has also published an *Introduction to Affine Group Schemes* and several articles on the history of mathematics.

This work was supported in part by the National Science Foundation, Grant MCS 8102697.

The first person to notice the general pattern behind the specific results was a Frenchman, O. Terquem [13]; the year was 1840. Terquem, who was then nearly sixty years old, was more a man of scholarship than a creative mathematician, and he is perhaps best remembered for his Voltairean credo, "I count all honest men as true believers, and only scoundrels as heretics." (More information about him and our other authors can be found in the biographical appendix below.) Terquem's paper was just a one-page note, and it was prompted by inequalities (6) and (7) above, which occur (at different places) in Cauchy's *Cours d'Analyse*. He considered the general principle obvious, saying that if the numbers x_i "all remain positive, and a given symmetric function preserves a given value, it is evident that every other symmetric function... will attain an extreme value, maximum or minimum, when all the [numbers] become equal." His main concern then was to distinguish the cases of maximum and minimum, which he did by considering the values that occur when one or more of the numbers go to zero.

Terquem's reaction seems to be a very natural one. That is, as soon as the symmetry of a problem is brought out, people are inclined to say that "by symmetry" the extreme value must occur when the variables are equal. But a bit of thought shows that there is no simple symmetry argument to this effect. Indeed, there cannot be, because such a symmetry conclusion is sometimes false. This was the discovery of our next author, the Russian V. Bouniakovsky [2], in 1854. Unlike the other two, Bunyakovskii (to give another common transliteration of his name) was a major mathematician, and there is no problem with his results. He looked at a single symmetric polynomial, asking whether its maxima or minima (if such exist) occur at points where the variables are equal. For polynomials of degree at most three he proved this is true, but in higher degrees he gave counterexamples. His basic example is neat and simple,

$$f(x, y) = \left[x^2 + (y-1)^2\right]\left[(x-1)^2 + y^2\right].$$

Obviously this is symmetric, and obviously also it has its minimum value at $(1,0)$ and $(0,1)$ rather than at any point where $x = y$.

Bouniakovsky did not deal directly with the constrained extremum problems that were Terquem's concern (and ours), but it is not too hard to find counterexamples in that context as well. Take for instance the symmetric function

$$f(x, y) = \left(x^2 + y^2 - 5/8\right)^2,$$

and consider its values for positive x and y satisfying $x + y = 1$. At the equality point $x = y = 1/2$ we get $f(1/2, 1/2) = (1/8)^2$. But f also takes on values larger and smaller than this, for instance $f(1/8, 7/8) = (5/32)^2$ and $f(1/4, 3/4) = 0$. Thus it would seem that our principle simply isn't true.

Still, it is hard to believe that all our original examples had symmetric extrema just by chance. It is time, therefore, that we turned to our third author, the Englishman H. J. Purkiss, who published his observation in 1862 in the *Messenger of Mathematics* [9]. This was then a new journal designed for English undergraduates, and Purkiss was one of the founders; at that time he was in fact a student at Trinity College, Cambridge, and was just twenty years old. Unlike Terquem, he realized that his observation required proof. Needless to say, his argument was unsatisfactory, but it has one significant feature: it deals only with small variations away from equality of the variables. Though Purkiss himself slurred over the distinction between relative and absolute maxima, his argument is the first indication of a basic fact: whatever validity the principle has will be *local*. For example, our function $(x^2 + y^2 - 5/8)^2$ on the line $x + y = 1$ does indeed have a local maximum at $x = y = 1/2$, even though it reaches greater values far away from that point. Thus we can now give a more precise formulation:

THE PURKISS PRINCIPLE. *Let $f(x_1, \ldots, x_n)$ and $g(x_1, \ldots, x_n)$ be two symmetric functions. On the set where g stays equal to $g(r, \ldots, r)$, the function f should have a local maximum or minimum at $(r, \ldots, r)$.*

Of course we do not yet have any idea why, or indeed whether, the principle should be true. All we know (from Bouniakovsky) is that no rudimentary symmetry argument will prove it. But at least we can try testing it by the general method to test for a constrained extremum at a point P. The classical way to remember this "Lagrange multiplier" method is to think of curves $c(t)$ that have $c(0) = P$ and lie in the set where g is constant. Differentiating the condition $g(c(t)) =$ constant, we find that

$$0 = \Sigma(D_i g)(P)c_i'(0),$$

which says that the curve's tangent vector $c'(0)$ is perpendicular to the gradient vector $(\nabla g)(P)$ formed by the $D_i(g) = \partial g/\partial x_i$. Conversely, so long as $(\nabla g)(P)$ is nontrivial, the implicit function theorem shows that every vector perpendicular to $(\nabla g)(P)$ occurs as the tangent to some such curve $c(t)$. Now if f has a local (constrained) extremum at P, each $f(c(t))$ has an extremum at 0, and its derivative vanishes there. Hence we have

$$0 = \Sigma(D_i f)(P)c_i'(0).$$

This says that $(\nabla f)(P)$ is also perpendicular to all the $c'(0)$, which implies that

$$(\nabla f)(P) \text{ is a scalar multiple of } (\nabla g)(P).$$

Points where this condition holds (together with points where $\nabla g = 0$) are the critical points for the constrained extremum problem.

All this is familiar in general. Does something special happen when f and g are symmetric? Let us work out an example, say

$$f(x_1, x_2, x_3) = x_1x_2 + x_2x_3 + x_3x_1,$$
$$g(x_1, x_2, x_3) = (x_1x_2x_3)^2.$$

The gradients of these are

$$\nabla f = (x_2 + x_3, x_1 + x_3, x_1 + x_2),$$
$$\nabla g = 2x_1x_2x_3(x_2x_3, x_1x_3, x_1x_2).$$

We are interested in points P of the form (r, r, r), and there we have

$$\nabla f(r, r, r) = (2r, 2r, 2r) \quad = 2r(1, 1, 1),$$
$$\nabla g(r, r, r) = (2r^5, 2r^5, 2r^5) = 2r^5(1, 1, 1).$$

Thus ∇f is a multiple of ∇g simply because in each of them all the entries are equal. Once we see that this is what we want to prove, we have no difficulty proving it.

LEMMA 1. *Suppose $f(x_1,\dots, x_n)$ is a symmetric differentiable function. Then at a point $x_1 = \cdots = x_n = r$, all the $D_i f$ are equal.*

Proof. Let π be a permutation of $1, 2,\dots, n$. For any function h we can define

$$h_\pi(x_1,\dots, x_n) = h(x_{\pi(1)},\dots, x_{\pi(n)}),$$

and trivially then

$$(D_{\pi(i)}h_\pi)(x_1,\dots, x_n) = (D_i h)(x_{\pi(1)},\dots, x_{\pi(n)}).$$

In the present case all f_π equal f and all x_i are equal. ∎

With hindsight (and a little charity) we might say that Purkiss also managed to show that the points $(r,\dots, r)$ are critical points. But then he undeniably fell into an error. When there is only one free variable, a critical point is "in general" a local maximum or minimum—that is, it will be one or the other except for degenerate cases where the second derivative vanishes. But in several variables this is not true, and nondegenerate critical points can equally well be saddle points.

Purkiss simply ignored this possibility, and thus he left a major gap in his argument. Yet it must be admitted that in our original examples we did not in fact encounter any saddle points. Support for the principle can also be drawn from George Chrystal, who in 1889 included it in Part 2 of his famous textbook on algebra [**4**, II. 61–63]. Recognizing the inadequacy of Purkiss' proof, Chrystal treated only the case where f and g are symmetric polynomials; these he rewrote in terms of the elementary symmetric polynomials, and after some computation he was able to establish the principle (apart from some degenerate situations). With this encouragement, then, let us take up where Purkiss left off and try to show that for some reason we never get saddle points.

For this we need the second-order terms in the Lagrange multiplier method. These ought to be familiar, but in fact most advanced calculus books seem to skip them. Briefly, then, let us again consider one of our curves $c(t)$ lying in the set where $g = g(P)$. We have

$$\begin{aligned} 0 &= \left.\frac{d^2}{dt^2} g(c(t))\right|_{t=0} \\ &= \Sigma_i (D_i g)(P) c_i''(0) + \Sigma_{i,j}(D_i D_j g)(P) c_i'(0) c_j'(0) \end{aligned}$$

and

$$\frac{d^2}{dt^2} f(c(t))|_{t=0} = \Sigma_i (D_i f)(P) c_i''(0) + \Sigma_{i,j}(D_i D_j f)(P) c_i'(0) c_j'(0).$$

We know that at our critical point we have $\nabla f(P) = \lambda(\nabla g)(P)$ for some scalar λ. Multiplying the first equation above by λ and subtracting, we get

$$\frac{d^2}{dt^2} f(c(t))|_{t=0} = \Sigma_{i,j}\left[(D_i D_j f)(P) - \lambda(D_i D_j g)(P)\right] c_i'(0) c_j'(0).$$

For a local maximum or minimum, we want these second derivatives to have the same sign for all $c(t)$. Thus the extra condition we need is that the quadratic form

$$Q(v) = \Sigma\left[(D_i D_j f)(P) - \lambda(D_i D_j g)(P)\right] v_i v_j$$

should be positive definite or negative definite on the space of all v perpendicular to $\nabla g(P)$. Using the implicit function theorem, one can show that this condition is indeed sufficient [**5**, 154].

This tells us what we need, but why should it automatically be true for symmetric f and g? Let us look again at our example,

$$f = x_1 x_2 + x_2 x_3 + x_3 x_1, \quad g = (x_1 x_2 x_3)^2.$$

At a point (r, r, r) we find that the matrices of second partials are

$$\begin{aligned} (D_i D_j f(P)) &= \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \\ (D_i D_j g(P)) &= \begin{pmatrix} 2r^4 & 4r^4 & 4r^4 \\ 4r^4 & 2r^4 & 4r^4 \\ 4r^4 & 4r^4 & 2r^4 \end{pmatrix}. \end{aligned}$$

What strikes the eye here is that all the diagonal entries are equal and all the off-diagonal entries are equal. Again, once we notice this we can easily check it in general.

LEMMA 2. *Suppose $f(x_1, \ldots, x_n)$ is a symmetric twice-differentiable function. Then at a point $(r, \ldots, r)$ all $D_i D_i f$ are equal and all $D_i D_j f$ for $i \neq j$ are equal.*

Proof. In the proof of Lemma 1 we saw that $D_{\pi(i)} h = (D_i h)_\pi$ for any h. Applying this rule twice, we get

$$D_{\pi(i)}D_{\pi(j)}f = D_{\pi(i)}D_{\pi(j)}f_\pi = D_{\pi(i)}\left[(D_j f)_\pi\right] = \left[D_i D_j f\right]_\pi.$$

Take first $i = j$, and choose π so $\pi(i) = 1$; this gives us $(D_1D_1f)(r,\ldots,r) = (D_iD_if)(r,\ldots,r)$. Take next any $i \neq j$, and choose π so $\pi(i) = 1$ and $\pi(j) = 2$; then we get $(D_1D_2f)(r,\ldots,r) = (D_iD_jf)(r,\ldots,r)$. ■

A straightforward computation now gives us the final lemma we need:

LEMMA 3. *Suppose a quadratic form Q is given by*

$$Q(v_1,\ldots,v_n) = \Sigma_i bv_iv_i + \Sigma_{i\neq j} cv_iv_j.$$

Then for all v satisfying $\Sigma v_i = 0$, we have $Q(v) = (c - b)\Sigma v_i^2$.

THEOREM (The PURKISS PRINCIPLE). *Let f and g be symmetric functions with continuous second derivatives in the neighborhood of a point $P = (r,\ldots,r)$. On the set where g equals $g(P)$, the function f will have a local maximum or minimum at P except in degenerate cases.*

Proof. Assuming $\nabla g(P) \neq 0$, we know from Lemma 1 that $\nabla f(P)$ has the form $\lambda\nabla g(P)$. The second partial derivatives of f and g satisfy the equalities in Lemma 2, and hence the terms $(D_iD_jf)(P) - \lambda(D_iD_jg)(P)$ also satisfy those equalities. Lemma 1 shows that the vectors v perpendicular to $\nabla g(P)$ are those with $\Sigma v_i = 0$. Lemma 3 then shows that on those v our quadratic form (if not identically zero) is positive or negative definite. The result then follows from the Lagrange multiplier criterion. ■

Two types of degeneracy were excluded from the proof: we assumed that $\nabla g(P)$ was nonzero, and then we assumed that the second-order terms in $f - \lambda g$ did not all vanish in the directions perpendicular to $\nabla g(P)$. Such exclusions are indeed necessary, and the Purkiss principle is not universally true. To see why a condition on ∇g is needed, consider the extreme case where g is constant everywhere; there is then no constraint, and hardly any symmetric f will satisfy the condition. It is a little harder to find an example that fails with the second-order degeneracy, but here is one such case. Let the functions be

$$f = x_1^4x_2x_3 + x_2^4x_3x_1 + x_3^4x_1x_2,$$

$$g = x_1^3x_2^3 + x_2^3x_3^3 + x_3^3x_1^3$$

around the point $P = (1,1,1)$. Consider the curve consisting of points (st, s, s), where t is close to 1 and s is chosen to keep $g = 3$ (that is, $s^6(2t^3 + 1) = 3$). For such points

$$f(st,s,s) = s^6(t^4 + 2t) = 3(t^4 + 2t)/(2t^3 + 1).$$

The derivative of this comes out to be $6(t^3 - 1)^2/(2t^3 + 1)^2$, which is zero at $t = 1$ but positive on both sides nearby. Thus f has values less than 3 for $t = 1 - \varepsilon$ and bigger than 3 for $t = 1 + \varepsilon$, and the Purkiss principle fails. It is not at all surprising that such exceptions should exist. What is surprising is how widely the principle turns out to be correct.

Thus far we have come in direct pursuit of the Purkiss principle. By good classical methods we have analyzed it; so far as it is true, we have proved it. Here we might stop. But in fact, the scope of the principle is much wider than we have yet imagined. For an alert reader, the first sign of this should have been Lemma 3. That was a straightforward computation, but in the line of argument it seemed to be a miracle, because for no apparent reason it gave us just what we needed. Such a seeming miracle should never be accepted on its face, for nine times out of ten it marks the presence of a general property not yet recognized or understood.

Furthermore, there are examples to suggest a more general result. Think for instance of a 90° rotation around the vertical axis in three-space. This rotation preserves $x_1^2 + x_2^2 + x_3^2$, and in particular it maps the unit sphere to itself, leaving the north and south poles fixed. Let f be any other function preserved by the same quarter turn, and consider its values on the unit sphere. It is intuitively clear that the north pole and south pole must be critical points for f on the sphere. Depending on the strength of your intuition, perhaps you can also see that the poles will be extrema rather than saddle points (apart from degenerate cases). Certainly it is easy to check this in examples. But this situation is not covered by the original Purkiss principle.

The geometric language in this example can serve to remind us that symmetry was a geometric concept to begin with. The symmetry that comes from interchangeability of the variables is only one of many types that a function might possess. What elementary geometry suggests is that we should allow other linear changes of variable. (We leave out translations because we are interested in fixed points.) As in all symmetry situations, two changes of variable that preserve a function f can be composed to get another one, and we will be dealing with a group of transformations. There is no reason to bother with continuous families of maps preserving f, since usually then we could reduce f to a function of fewer variables. Thus we are led to consider what happens to the Purkiss principle when we replace interchanges of variables by some other finite group of linear transformations.

Unavoidably now we must raise the level of discussion to assume some familiarity with linear algebra. Our basic structure will be a finite group $\mathbf{G}$ of linear transformations on a finite-dimensional real vector space V. The "symmetric" functions will be those f that satisfy $f(Tv) \equiv f(v)$ for all T in $\mathbf{G}$; the usual name for these is *invariant functions*. The points $(r,\dots,r)$ that we considered before are precisely the ones unaffected by interchanges of coordinates; in our general situation we should correspondingly consider the *fixed points* P in V, those sent to themselves by every T in $\mathbf{G}$. We must also make explicit an invariance property that was hidden in the earlier treatment: interchanging coordinates simultaneously in two vectors $(x_1,\dots,x_n)$ and $(y_i,\dots,y_n)$ does not change the dot product $x \cdot y = \Sigma x_i y_i$. The corresponding tool in general is given to us by the following standard result.

LEMMA 4. *There exists a positive definite inner product* $(\ ,\ )$ *on* V *that is* $\mathbf{G}$*-invariant; in other words,* $(Sv, Sw) = (v, w)$ *for every* S *in* $\mathbf{G}$ *and* v, w *in* V.

Proof. Let $v \cdot w$ be the dot product in some coordinate system. Define

$$(v, w) = \Sigma_{T \text{ in } \mathbf{G}} Tv \cdot Tw.$$

Clearly this is symmetric and bilinear and positive definite. For S in $\mathbf{G}$ we have $(Sv, Sw) = \Sigma TSv \cdot TSw$; and since $\mathbf{G}$ is a group, the terms of this sum are just the terms of (v, w) in different order. ■

Consider now a Taylor expansion of a function f around a point P: it can be written in the form

$$f(P+v) = f(P) + (\nabla f(P), v) + Q_P^f(v) + R(v),$$

where Q_P^f is a quadratic form and $R(v)$ is a remainder term satisfying

$$\lim_{v \to 0} R(v)/(v, v) = 0.$$

In such an expression the vector $\nabla f(P)$ and the quadratic form Q_P^f are uniquely determined. Suppose therefore that we take a fixed point P and an invariant function f. For each T in $\mathbf{G}$ we have then

$$\begin{aligned} f(P+v) &= f(T(P+v)) = f(P+Tv) \\ &= f(P) + (\nabla f(P), Tv) + Q_P^f(Tv) + R(Tv). \end{aligned}$$

By uniqueness we see that $Q_P^f(Tv) = Q_P^f(v)$ and $(\nabla f(P), Tv) = (\nabla f(P), v)$. Invariance of the inner product tells us that $(\nabla f(P), Tv) = (T^{-1}\nabla f(P), v)$; and since this equals $(\nabla f(P), v)$ for all v, we have $T^{-1}\nabla f(P) = \nabla f(P)$. Thus we get the following result, which is as close as we can come to Lemmas 1 and 2 in this generality.

LEMMA 5. *Let f be invariant and P fixed. Then $\nabla f(P)$ is fixed, and the quadratic form Q_P^f is invariant (that is, $Q_P^f(Tv) \equiv Q_P^f(v)$ for all T in* **G**).

Nothing so strong as the Purkiss principle will be true unless we make some further assumptions about our group of transformations. To get the analogue of Lemma 1, for instance, we need to know that the fixed vector $\nabla f(P)$ is forced to be a multiple of the fixed vector $\nabla g(P)$. Thus we should assume that the fixed vectors in V form a one-dimensional subspace. After that, we must separate two types of terms, as in Lemma 2; for this we need an analogue of the space where $\Sigma v_i = 0$. The general concept needed here is that of an *invariant subspace*, one sent to itself by all maps in **G**. We then have the following standard result.

LEMMA 6. *Let U be an invariant subspace. Let W be its orthogonal complement with respect to $(\, , \,)$. Then W is also an invariant subspace.*

Proof. Take any w in W and T in **G**. Let u be any element of U. Then (Tw, u) equals $(w, T^{-1}u)$, and this is zero because $T^{-1}u$ is in U. Thus Tw is orthogonal to U. ■

In particular, our one-dimensional space U_0 of fixed vectors has an orthogonal complement W_0 which is an invariant subspace. Now if we look back at the proof of Lemma 2, we see that it depended on our having a great many permutations, enough to match most of the matrix entries with each other. The general analogue of this is *irreducibility*, which means that there should be no nontrivial invariant subspaces inside W_0. Using this, we can finally establish the general property behind Lemma 3.

LEMMA 7. *Let* **G** *be a finite group of linear transformations acting irreducibly on a finite-dimensional real vector space W. Let $Q(w)$ be a* **G***-invariant quadratic form on W. Then $Q(w)$ is a scalar multiple of (w, w). In particular, Q is positive or negative definite whenever it is not identically zero.*

Proof. We can express the quadratic form Q in terms of the invariant inner product, getting $Q(w) = (Bw, w)$ for a unique self-adjoint linear map $B: W \to W$. For each T in **G** we have

$$(Bw, w) = Q(w) = Q(Tw) = (BTw, Tw) = (T^{-1}BTw, w).$$

The uniqueness of B tells us then that $T^{-1}BT = B$. In other words, $BT = TB$. But now the real self-adjoint map B has some real eigenvalue b. Let X be $\{w \text{ in } W | Bw = bw\}$, the corresponding eigenspace. For w in X and T in **G** we have $BTw = TBw = Tbw = bTw$, so Tw is in X. Thus X is an invariant subspace of W. Since it is nontrivial, it must be all of W. Thus $B = b \cdot \mathrm{Id}$, and $Q(w) = (Bw, w) = b(w, w)$. ■

We have now established the analogues of our three initial lemmas. As before, we must rule out degenerate cases: that is, we assume that $\nabla g(P)$ is nonzero and that the second-order terms in $f - \lambda g$ around P do not vanish on the space orthogonal to $\nabla g(P)$ under $(\, , \,)$. The proof of the original theorem then gives us our wider result.

THEOREM (The EXTENDED PURKISS PRINCIPLE). *Let* **G** *be a finite group of linear transformations of a finite-dimensional real vector space V. Assume that the fixed vectors form a one-dimensional subspace, and that* **G** *acts irreducibly on the complementary subspace. Let f and g be twice-differentiable* **G***-invariant functions on a neighborhood of a fixed vector P. Then on the set where $g(v) = g(P)$, the function f has a local maximum or minimum at P except in degenerate cases.*

Is this now the ultimate generalization of the Purkiss principle? Not at all. Indeed, Purkiss himself mentioned a similar result where the variables were subject to several independent constraints by different functions $g_1 = g_1(P), \ldots, g_r = g_r(P)$. Other examples will easily come to mind. If we are to do more than just accumulate further cases, we must now reach for more general geometric concepts, those from the rudimentary theory of differentiable manifolds. Assuming familiarity with those, we can examine the theorem with a critical eye, distinguishing the soul of the principle from its mere outward limbs and flourishes.

First of all, we observe that the actual values of the function g are never mentioned. All that matters is the set M of points where g is equal to $g(P)$. Instead of having g symmetric, then, all we need to require is that our group $\mathbf{G}$ should map M to itself. The condition $\nabla g(P) \neq 0$ now is the familiar condition that allows us to introduce smooth coordinates on M around P. Thus we can absorb that condition by saying that our set M should be a differentiable manifold. (There is of course no need for M to be complete; it may be nothing but a smooth piece around P.) The values of f at points off M are also irrelevant, and all we need for the statement is that f should be defined as a smooth $\mathbf{G}$-invariant function on M.

Furthermore, there is no need for the maps in $\mathbf{G}$ to be induced by linear maps on some ambient space. Linearity will enter automatically, because any smooth map of M to M that fixes P will induce a linear map on the tangent space to M at P. This tangent space in our earlier situation was precisely the orthogonal complement to the span of $\nabla g(P)$, and so this is the space where we should assume irreducibility. Thus we are led to formulate a modernized version of our theorem:

THEOREM (The MODERN PURKISS PRINCIPLE). *Let M be a differentiable manifold. Let $\mathbf{G}$ be a finite group of smooth maps from M to M. Let P be a point in M fixed by $\mathbf{G}$, and let f be a differentiable function on M invariant under $\mathbf{G}$. Assume that the action induced by $\mathbf{G}$ on the tangent space at P is (nontrivial and) irreducible. Then P is a critical point of f; and if this critical point is nondegenerate, it is a local maximum or minimum of f.*

Proof. This theorem shows nicely how much can be done with the machinery of manifolds; for though it is a much more general result, it requires no new ideas in its proof. To begin with, f induces a linear function df_P on the tangent space at P. The kernel of df_P will be $\mathbf{G}$-invariant, and irreducibility thus requires df_P to be zero. (We assume that the action is nontrivial in order to rule out a counterexample in dimension one.) Hence P is a critical point for f. Consequently, the second-order terms of f induce a quadratic form on the tangent space, and clearly this form will be $\mathbf{G}$-invariant. By definition one says that the critical point is nondegenerate when this quadratic form is nondegenerate. In that case Lemma 7 again shows that the form is positive definite or negative definite, and P accordingly gives a local minimum or maximum. ■

This theorem has a quite up-to-date sound, and so with it we can end our pursuit of the Purkiss principle. I do not happen to have seen this final version anywhere before, but I make no great claims for it. More important is that we now see how such a result may stand at the end of a search that began with nothing more than some interchangeable variables in calculus problems.

BIOGRAPHICAL APPENDIX

Olry Terquem ([3], [7], [8]) was born in Metz, France, in 1782. His native language was a form of Yiddish, and his early studies were limited to Hebrew and the Talmud. But the upheaval of the French Revolution happened to bring his older brother in contact with the Jewish community at Coblentz, and a tutor engaged there led Olry into more general studies. His weakness in French made him fail in his first application to the École Polytechnique, but he was admitted on his second attempt (1801). In 1804 he began to teach higher mathematics at the lycée in Mayence, moving in 1811 to a professorship in the artillery school there. In 1814 he was called to Paris as professor attached to the Comité de l'Artillerie; in this post he served as librarian for the Dépôt Central d'Artillerie and also as a general scientific consultant. Exempted from mandatory retirement, he continued active until his death at age 80.

Terquem translated technical works related to artillery and also wrote textbooks on algebra, geometry, and mechanics. Of his discoveries in pure mathematics, we might single out his computation of the number of normals from a given point to an algebraic surface of given degree. His mastery of languages helped make him an authority on the history of mathematics. He also wrote many articles on Jewish concerns, advocating for instance prayers in the vernacular languages rather than Hebrew. Perhaps his most important contribution to mathematics began in 1842, when he was sixty: he was cofounder of the journal *Nouvelles Annales de Mathématiques*, which he continued to edit until his death. Characteristic of the man are these lines from one of his very last letters: "I believe that human intelligence approaches the divine intelligence *asymptotically*. Let us hope!"

Viktor Yakovlevich Bunyakovskii (1804–1889) is a much better known mathematician than the others, and hence less needs to be said about him here [**6**]. After basic studies in Russia, he received his doctorate in mathematics at Paris in 1825. For most of his career he was a professor at the University of St. Petersburg (now Leningrad) and a member of the St. Petersburg Academy of Sciences; from 1864 to his death he was vice-president of the Academy. He wrote about 150 published works in mathematics and mechanics. His most famous discovery, of course, was the inequality

$$\left(\int_a^b fg\right)^2 \leqslant \left(\int_a^b f^2\right)\left(\int_a^b g^2\right),$$

which he published in 1859. Unfortunately it appeared only in a separate pamphlet written in Russian, and the inequality remained unknown in western Europe until it was rediscovered by Schwarz in 1884.

Henry John Purkiss ([10], [11], [14]) was born in London in 1842. In 1859 he was admitted to Trinity College, Cambridge, as a scholarship student (sizar). In 1862 he became one of the founders and editors of *The [Oxford, Cambridge, and Dublin] Messenger of Mathematics.* He received his B.A. degree in 1864 as the top mathematics student in his class (Senior Wrangler, 1st Smith's Prize). From 1864 to 1865 he was Vice-Principal of the College of Naval Architecture, South Kensington, and in 1865 he was named Principal of the Royal School of Naval Architecture. He drowned on September 17, 1865, in the river Cam. Little more is recorded of his sadly short life, but we can list his published papers:

1. Theorem in maxima and minima, Messenger of Math., 1 (1862) 180–183.
2. Dynamical note, Messenger of Math., 2 (1864) 228.
3. The cardioid, Messenger of Math., 2 (1864) 241–249.
4. Note on the comparative value of Simpson's two rules, and on Dr. Woolley's rule, Naval Architects' Trans., 6 (1865) 48–50.
5. The equation of the tangent, Messenger of Math., 3 (1866) 19–22.
6. Notes on pedal coordinates, Messenger of Math., 3 (1866) 83–88.
7. Notions of infinity derived from gnomonic projection, Messenger of Math., 3 (1866) 171–172.
8. On certain formulas of mensuration, Quart. J. Math., 7 (1866) 235–241.

George Chrystal (1851–1911) was a Scotsman ([**1**], [**12**]). After studies in Aberdeen, he was in residence at Peterhouse, Cambridge, from 1872 to 1875. As a student there, he spent much time with Maxwell in the newly opened Cavendish Laboratory. Many of his friends considered this a "waste of time," since Maxwell's work was not covered in the Tripos examination; even so, Chrystal came out second in his class. He was always inclined toward physics: his first major work was an experimental verification of Ohm's Law, and shortly before his death he did significant work on the analysis of seiches (long-lasting waves in lakes). From 1879 until his death he was Professor of Mathematics at the University of Edinburgh, where he played a major role both in reforming the university curriculum and in raising the level of primary and secondary education in Scotland.

Despite all this, Chrystal will always be best known for his classic algebra textbook, which is still in print. Officially he is present in this Appendix because his book gave the first proof of any significant form of the Purkiss principle, but in fact he is included because the temptation to quote him is irresistible. Here is a bit from an address he gave a year before Part I of his book appeared:

> The whole teaching consists in example grinding. What should be merely the help to attain the end has become the end itself. The result is that algebra, as we teach it, is neither an art nor a science, but an ill-digested farrago of rules whose object is the solution of examination problems.... The end of all education nowadays is to fit the student to be examined; the end of every examination not to be an educational instrument, but to be an examination which a creditable number of men (however badly taught) shall pass. We reap, but we omit to sow.... The cure for all this evil is to give effect to a higher ideal of education in general, and of scientific education in particular.... It takes the hand of God to make a great mind, but contact with a great mind will make a little mind greater.

Here is part of the Preface to Part II of his book, which is devoted mainly to a careful treatment of power series:

> A practice has sprung up of late (encouraged by demands for premature knowledge in certain examinations) of hurrying young students into the manipulation of the machinery of the Differential and Integral Calculus before they have grasped the preliminary notion of a *Limit*... on which all the meaning and all the uses of the Infinitesimal Calculus are based. Besides being to a large extent an educational sham, this course is a sin against the spirit of mathematical progress.

And here, finally, is part of the famous Preface to Part I:

> The first object I have set before me is to develop Algebra as a science, and thereby to increase its usefulness as an educational discipline. I have also endeavoured so to lay the foundations that nothing shall have to be unlearned.... It becomes necessary, if algebra is to be anything more than a mere bundle of unconnected rules, to lay down generally the fundamental laws of the subject, and to proceed deductively.

Amen.

References

1. J. S. Black and C. G. Knott, Professor George Chrystal, M.A., LL.D., Proc. Roy. Soc. Edinburgh, 32 (1913) 477–503.

2. V. Bouniakovsky, Note sur les maxima et les minima d'une fonction symétrique entière de plusieurs variables, Bull. Classe Phys.-Math. Acad. Imper. Sci. St. Petersburg, 12 (1854) 353–361. [This academy is the predecessor of the current Akademija Nauk SSSR.]

3. M. Chasles, Rapport sur les travaux mathématiques de M. O. Terquem, Nouvelles Ann. de Math., (2), 2 (1863) 241–251.

4. G. Chrystal, Algebra: An Elementary Textbook, Parts I and II, A. and C. Black, Edinburgh, 1886 and 1889.

5. C. H. Edwards, Jr., Advanced Calculus of Several Variables, Academic Press, New York, 1973.

6. A. T. Grigorian, Bunyakovskii, Viktor Yakovlevich; in C. C. Gillespie (Editor), Dictionary of Scientific Biography, vol. 15., Scribner, New York, 1978, pp. 66–67.

7. E. Prouhet, Notice sur la vie et les travaux d'Olry Terquem, Bull. Bibliog. Hist. Biog. Math., 8 (1862) 81–90. [This was issued as part of the Nouvelles Ann. de Math., Series 2, volume 1.]

8. E. Prouhet, Terquem (Olry); in J. Michaud et al., Biographie Universelle, Nouvelle Edition 1854 ff., reprint Akademische Druck- und Verlagsanstalt, Graz, 1970, vol. 41, p. 168.

9. H. J. Purkiss, Theorem in maxima and minima, Messenger of Math., 1 (1862) 180–183.

10. Royal Society of London, Catalogue of Scientific Papers (1800–1863), vol. 5, H. M. Stationery Office, London, 1871.

11. Royal Society of London, Catalogue of Scientific Papers (1864–1873), vol. 8, John Murray, London, 1879.

12. R. Schlapp, George Chrystal; in C. C. Gillespie (Editor), Dictionary of Scientific Biography, vol. 3, Scribner, New York, 1971, pp. 264–265.

13. O. Terquem, Démonstration de deux propositions de M. Cauchy, J. Math. Pures Appl. (Liouville), 5 (1840) 37.

14. J. A. Venn (Editor), Alumni Cantabrigienses, Part II, vol. 5, Cambridge Univ. Press, Cambridge, 1953.