V.3 : Classification of critical points

(Fraleigh and Beauregard, §8.3)

This section deals with multivariable versions of the second derivative test for relatve maximum and minimium values of a function that has continuous second partial derivatives. In single variable calculus, the second derivative test for such values has two basic steps:

(1) Find the points x for which the derivatitive $f'(x)$ of the function f is equal to zero (or is undefined).

(2) For each x such that $f'(x) = 0$, find the second derivative $f''(x)$. If it is positive, then f has a relative minimum at x , but if it is negative then f has a relative maximum at x. If $f''(x) = 0$, then the test yields no information on whether there is a realtive maximum, relative minimum, or neither.

Generalzations of this result to functions of two, and sometimes even three, variables, are often found in the sections of calculus texts that discuss partial differentiation. but generally the justification for these tests is not presented because it involves and understanding of the following question about matrices:

POSITIVITY QUESTION. *Let A be a real symmetric matrix, and let q_A be the quadratic form defined by $q_A(\mathbf{x}) = {}^{\mathbf{T}}\mathbf{x}\, A\, \mathbf{x}$. Under what conditions on A can we conclude that $q_A(\mathbf{x})$ is positive for every nonzero vector $\mathbf{x}$?*

A symmetric matrix satisfying the condition in the preceding question is said to be **positive definite**. The main algebraic result of this section gives an answer to this question in relatively simple terms. We shall state the main result now and prove it after comploeting the discussion of the second derivative test for functions of several variables.

PRINCIPAL MINORS THEOREM. *Suppose that A is an $n \times n$ symmetric matrix over the real numbers, and for each integer k between 1 and n let $A^{(k)}$ be the $k \times k$ matrix whose (i,j) entry is $a_{i,j}$ for $1 \leq i, j \leq k$ (visually, this is the $k \times k$ submatrix in the upper left hand part of the original array with all other entries discarded). Then A is positive definite if and only if $\det A^{(k)}$ is positive for all k between 1 and n.*

In older terminology for matrix algebra, the quantity $\det A^{(k)}$ was called the k^{th} **principal minor** of A (minors referred to determinants of matrices formed by deleting suitable numbers of rows and columns), and this is the reason behind the naming of the result. When we return to the derivation of the Principal Minors Theorem we shall also show that it is much easier to check its validity for examples than one might initially think.

Taylor's Theorem in several variables

Recall that Taylor's Theorem — or Taylor's Formula — is a result on approximating functions with $(n+1)$ continuous derivatives by polynomials of degree $\leq n$ and that it includes a formula for the error in the given "optimal" approximations. Usually this is expressed in terms of the $(n+1)^{\text{st}}$ derivative of the function, but for some purposes other descriptions of the error term are also worth knowing. Nearly every calculus text includes a treatment of Taylor's Theorem and at least one version of the error term. For our purposes it will suffice to use the version stated as Theorem 8.19 on page 611 of *Calculus* (Seventh Edition), by Larson, Hostetler and Edwards.

TAYLOR'S THEOREM WITH THE LAGRANGE REMAINDER. *Suppose that f is a real valued function defined on an interval $(c-r,\, c+r)$ where $c \in \mathbf{R}$ and $r > 0$, and suppose that f has continuous derivatives of all orders through $(n+1)$ on that interval, where $n \geq 1$. Then for each $h \neq c$ such that $|h| < r$ there is a point α between c and $c+h$ such that*

$$f(x+h) \quad = \quad f(c) \; + \; \sum_{k=1}^{n} \frac{f^{(k)}(c)}{k!}\, h^k \quad + \quad R_n(h)$$

where

$$R_n(h) \quad = \quad \frac{f^{(n+1)}(\alpha)}{(n+1)!}\, h^{\,n+1} \qquad .\blacksquare$$

We shall need a version of this result for functions of m variables. The first step in formulating this generalization is to specify where f is defined, and we replace the interval of radius r centered at $c \in \mathbf{R}$ with the open disk of radius r about a point $\mathbf{c} \in \mathbf{R}^n$:

$$N_r(\mathbf{c}) \quad = \quad \{\, \mathbf{x} \in \mathbf{R}^m \;\;|\;\; |\mathbf{x} - \mathbf{c}| \;\; < \;\; r \,\}$$

Given that we are starting with the single variables form of Taylor's Theorem, it should not be surprising that we want to derive the multivariable form by constructing some single variable function out of the given multivariable function. If we let $\mathbf{h} \neq \mathbf{0}$ such that $|\mathbf{h}| < r$, then the function

$$g(t) \quad = \quad f(\mathbf{c} + t\,\mathbf{h})$$

will be a function defined on an open interval of radius

$$\frac{r}{|\mathbf{h}|} \quad > \quad 1 \; .$$

In order to show that this function has the required differentiability properties, one assumes that f has continuous partial derivatives with respect to all combinations of variables through order $(n+1)$; the differentiability properties of g then can be obtained by repeated application of the Chain Rule from multivariable calculus (see Section 12.5 on pages 876–883 of the text by Larson, Hostetler and Edwards).

Strictly speaking, the statements of the Chain Rule in calculus textbooks often only give a formula such as

$$\frac{d\,w}{d\,s} \quad = \quad \sum_{j=1}^{n} \frac{\partial\, w}{\partial\, u_j}\, \frac{d\, u_j}{d\, s}$$

(this is essentially the formula at the top of page 877 of Larson, Hostetler and Edwards), but in our case it is easy to check that this expression has continuous derivatives of order k for each $k \leq n$ by induction. Specifically, if $\mathbf{h}$ is expressed in coordinates as $(h_1,\; \cdots\;, h_n)$, then we have the following explicit formula:

$$g^{(k)}(t) \quad = \quad \sum_{1 \leq i_1, \cdots,\, i_k \leq m} \quad \frac{\partial^k f}{\partial x_{i_1} \;\cdots\; \partial x_{i_k}} \big(\mathbf{c} + t\mathbf{h}\big) \cdot h_{i_1} \;\cdots\; h_{i_k}$$

We can now state a multivariable version of Taylor's Theorem that is sufficient for our purposes.

MULTIVARIABLE TAYLOR'S FORMULA FOR DEGREE 1 APPROXIMATIONS. *Let f be a function of m variables with continuous second partial derivatives with respect to all pairs of variables on the open disk $N_r(\mathbf{c})$, and let $\mathbf{h} \neq \mathbf{0}$ be such that $|\mathbf{h}| < r$. Then there is some $\alpha \in [0,1]$ such that*

$$f(\mathbf{c}+\mathbf{h}) \quad = \quad f(\mathbf{c}) \ + \ \sum_{k=1}^{m} \frac{\partial f}{\partial x_k}(\mathbf{c})\, h_k \ + \ \sum_{1 \leq i,j \leq m} \frac{\partial^2 f}{\partial x_i\, \partial x_j}\big(\mathbf{c}+\alpha\mathbf{h}\big) \cdot h_i\, h_j \ .$$

We shall describe a more concise way of writing the right hand side. One can use gradients to write the linear part of the function as $f(\mathbf{c}) + \nabla f(\mathbf{c}) \cdot \mathbf{h}$, and if one defines the *Hessian* of f at a point $\mathbf{p}$ to be the symmetric $m \times m$ matrix $\mathrm{Hess}(f;\mathbf{p})$ whose (i,j) entry is equal to

$$\frac{\partial^2 f}{\partial x_i\, \partial x_j}(\mathbf{p})$$

then we can rewrite the relevant case of Taylor's Formula as follows:

$$f(\mathbf{c}+\mathbf{h}) \quad = \quad f(\mathbf{c}) \ + \ \nabla f(\mathbf{c}) \cdot \mathbf{h} \ + \ {}^{\mathbf{T}}\mathbf{h} \left[\mathrm{Hess}(f;\mathbf{c}+\alpha\mathbf{h})\right] \mathbf{h}$$

With this terminology we are finally ready to write down the second derivative test for relative extrema in several variables:

MULTIVARIABLE SECOND DERIVATIVE TEST. *Let f be a function with continuous second partial derivatives on the open disk $N_r(\mathbf{c})$ in $\mathbf{R}^n$, and suppose that $\nabla f(\mathbf{c}) = \mathbf{0}$. Then the following conclusions hold:*

(1) *If the Hessian of f at $\mathbf{c}$ is a* **positive definite** *matrix, then f has a relative* **minimum** *at $\mathbf{c}$.*

(2) *If the* **negative** *of the Hessian of f at $\mathbf{c}$ is a positive definite matrix, then f has a relative* **maximum** *at $\mathbf{c}$.*

(3) *If the Hessian of f at $\mathbf{c}$ is invertible, but neither it nor its negative is positive definite, then f has neither a relative maximum nor a relative minimum at $\mathbf{c}$ (these are essentially the* **saddle points** *one discusses in multivariable calculus).*

(4) *If the Hessian of f at $\mathbf{c}$ is not invertible, then no conclusion can be drawn.*

The standard test the positive definiteness of the Hessian is given by the Principal Minors Theorem, and in fact this test plays an important part in the derivation of the Multivariable Second Derivative Test. Therefore we shall restate the latter using the Principal Minors Theorem:

ALTERNATE FORMULATION. *In the setting above, one can restate the conclusions as follows:*

(1) *If the the principal minors for the Hessian of f at $\mathbf{c}$ are all* **positive***, then f has a relative* **minimum** *at $\mathbf{c}$.*

(2) *If the* **odd** *principal minors for the Hessian of f at $\mathbf{c}$ are all* **negative** *and the* **even** *principal minors for the Hessian of f at $\mathbf{c}$ are all* **positive***, then f has a relative* **maximum** *at $\mathbf{c}$.*

(3) *If the Hessian of f at $\mathbf{c}$ has a nonzero determinant but the principal minors* **do not satisfy either** *of the sequences of positivity and negativity conditions described above, then f has neither a relative maximum nor a relative minimum at $\mathbf{c}$ (this is the* **saddle point** *case).*

(4) *If the determinant of the Hessian of f at $\mathbf{c}$ is zero, then no conclusion can be drawn.*

Since no information is obtained if the determinant of the Hessian vanishes, we shall ignore this case henceforth and assume that the Hessian at $\mathbf{c}$ has a nonzero determinant. The proof of the second derivative test then splits into three cases; namely, the tests for relative minima, relative maxima and saddle points.

Relative minimum test. Since the gradient vanishes at $\mathbf{c}$, the concise formulation of Taylor's Theorem yields the following identity, in which $\alpha \in [0,1]$:

$$f(\mathbf{c}+\mathbf{h}) \;-\; f(\mathbf{c}) \quad = \quad {}^{\mathbf{T}}\mathbf{h}\left[\mathrm{Hess}(f;\mathbf{c}+\alpha\mathbf{h})\right]\mathbf{h}$$

Suppose that the Hessian of f and $\mathbf{c}$ is positive definite, so that all of its principal minors are positive. We need to show that the positivity of these principal minors for $\mathrm{Hess}(f;\mathbf{c})$ implies a corresponding statement for $\mathrm{Hess}(f;\mathbf{c}+\alpha\mathbf{h})$, at least if $|\mathbf{h}|$ is sufficiently small. Specifically, here is what we need:

Lemma. *If the principal minors of* $\mathrm{Hess}(f;\mathbf{c})$ *are all positive then there is a number $s \in (0,r)$ such that $|\mathbf{v}| < s$ implies that the principal minors of* $\mathrm{Hess}(f;\mathbf{c}+\mathbf{v})$ *are also all positive.*

Proof of Lemma. Since the determinant is a polynomial in the entries of a matrix, it is a continuous function of the entries of a matrix. In particular, if the principal minors of a square matrix A are all positive and the entries of a second matrix B are all sufficiently close to those of A, then the principal minors of B will also be all positive. Take A to be the Hessian at $\mathbf{c}$ and B to be the Hessian at some other point $\mathbf{c}+\mathbf{v}$. Since f has continuous second partial derivatives there is an $s > 0$ such that $|\mathbf{v}| < s$ implies the entries of B are sufficiently close to the entries of A for the principal minors condition to hold.■

Return to the relative minimum test. If we now restrict ourselves to vectors $\mathbf{h}$ of length $\leq s$, then the Lemma tells us that the principal minors for $\mathrm{Hess}(f;\mathbf{c}+\alpha\mathbf{h})$ will all be positive, and hence that

$${}^{\mathbf{T}}\mathbf{h}\left[\mathrm{Hess}(f;\mathbf{c}+\alpha\mathbf{h})\right]\mathbf{h} \quad > \quad 0\ .$$

It follows that f has a **strict** relative minimum at $\mathbf{c}$.■

Relative maximum test. This will be derived directly from the preceding test, using the fact that the relative maxima of f are the same as the relative minima of its negative function $-f$. Since the Hessians of f and $-f$ are negatives of each other, the relative minimum test immediately implies that one has a strict relative maximum if the principal minors of the matrix $-\mathrm{Hess}(f;\mathbf{c})$ are all positive. Since the determinants of $k \times k$ matrices satisfy the relation $\det(-B) = (-1)^k \det B$, it follows that the k^{th} principal minor of the matrix

$$\mathrm{Hess}(-f;\mathbf{c}) \quad = \quad -\mathrm{Hess}(f;\mathbf{c})$$

is equal to the corresponding principal minor for $\mathrm{Hess}(f;\mathbf{c})$ if k is even and the negative of the corresponding principal minor if k is odd. Therefore the conditions for f to have a strict relative maximum, or equivalently for $-f$ to have a strict relative minimum, translate into the negativity of the odd principal minors of the Hessian and the positivity of the even principal minors of the Hessian.■

Saddle point test. Before reading this it might be useful to say a few words about saddle points. The typical example for functions of two variables is given by $f(x,y) = y^2 - x^2$, and it is illustrated on page 908 of Larson, Hostetler and Edwards. If we restrict the function to points of the form $(0, y)$ then this restricted function has an absolute minimum at $(0, 0)$, while if we restrict to points of the form $(x, 0)$ then the restricted function has an abolute minimum there. This topic is also discussed on pages 193–194 of *Basic Multivariable Calculus*, by Marsden, Tromba and Weinstein.

Once again we need some additional input.

Lemma. *If A is a symmetric matrix over the real numbers, then A is positive definite if and only if all of its eigenvalues are positive.*

Proof of Lemma. By Rayleigh's Principle the minimum value of the Rayleigh quotient is the minimum eigenvalue, and by the definition of this quotient it is positive for all nonzero vectors in $\mathbf{R}^n$ if and only if q_A is positive for all such vectors. But the latter is precisely the condition for a real symmetric matrix A to be positive definite. Since all the eigenvalues of A are positive if and only if the smallest eigenvalue is positive, this proves the Lemma.■

Return to the saddle point test. In this case we are assuming that the Hessian at $\mathbf{c}$ is invertible, so that its eigenvalues are all nonzero, but by the Lemma we are also assuming conditions which mean that its eigenvalues are neither all positive nor all negative. It follows that the maximum eigenvalue must be positive and the minimum eigenvalue must be negative. Let $\lambda_\pm$ be the maximum and minimum eigenvalues, and let $\mathbf{u}_\pm$ be unit eigenvectors for these eigenvalues. If $\mathbf{h}_\pm$ is a nonzero multiple of $\mathbf{u}_\pm$ it follows that

$$ {}^{\mathbf{T}}\mathbf{h}_\pm \left[\mathrm{Hess}(f;\mathbf{c})\right] \mathbf{h}_\pm \quad = \quad |\mathbf{h}_\pm|^2 \, \lambda_\pm $$

and therefore the expression is positive for $\mathbf{h}_+$ and negative for $\mathbf{h}_-$. By the continuity of the entries in the Hessian matrix, we can find some $s \in (0, r)$ such that if $|\mathbf{v}| < s$ then

$$ {}^{\mathbf{T}}\mathbf{u}_+ \left[\mathrm{Hess}(f;\mathbf{c}+\mathbf{v})\right] \mathbf{u}_+ \quad > \quad 0 $$

and also

$$ {}^{\mathbf{T}}\mathbf{u}_- \left[\mathrm{Hess}(f;\mathbf{c}+\mathbf{v})\right] \mathbf{u}_- \quad < \quad 0 \; . $$

Suppose now that we choose $\mathbf{h}_\pm$ to have length less than s. As in the previous tests we know that

$$ f(\mathbf{c}+\mathbf{h}_\pm) \; - \; f(\mathbf{c}) \quad = \quad {}^{\mathbf{T}}\mathbf{h}_\pm \left[\mathrm{Hess}(f;\mathbf{c}+\alpha\mathbf{h})\right] \mathbf{h}_\pm $$

and we may rewrite the right hand side in the following form:

$$ |\mathbf{h}_\pm|^2 \cdot {}^{\mathbf{T}}\mathbf{u}_\pm \left[\mathrm{Hess}(f;\mathbf{c}+\alpha\mathbf{h}_\pm)\right] \mathbf{u}_\pm $$

Now the first factor is always positive, and the conditions on the vectors $\mathbf{h}_\pm$ ensure that the remaining factor is positive for $\mathbf{h}_+$ and negative for $\mathbf{h}_-$. Therefore $f(\mathbf{c}+\mathbf{h}_+) \; - \; f(\mathbf{c})$ is positive and $f(\mathbf{c}+\mathbf{h}_-) \; - \; f(\mathbf{c})$ is negative, since we can choose the lengths of $\mathbf{h}_\pm$ to be arbitrarily small positive numbers, it follows that f neither has a relative maximimum nor a relative minimum at $\mathbf{c}$.■

Proof of the Principal Minors Theorem

The following general result on the structure of positive definite matrices will be important for our purposes.

THEOREM. *If A is a symmetric $n \times n$ matrix, then the following conditions are equivalent:*

(1) *The matrix A is positive definite.*

(2) *The function $\varphi_A(\mathbf{x}, \mathbf{y}) = {}^{\mathbf{T}}\mathbf{y}\, A\, \mathbf{x}$ defines an inner product on $\mathbf{R}^n$.*

(3) *The matrix A can be written as a product ${}^{\mathbf{T}}P\, P$ for some invertible matrix P.*

Proof. [(1) $\Longrightarrow$ (2)] The basic conditions for an inner product, for example as stated in Theorem 1.3 on pages 24–25 of the text, are consequences of the rules for matrix multiplication and the assumption that ${}^{\mathbf{T}}\mathbf{x}\, A\, \mathbf{x}$ is positive if $\mathbf{x} \neq \mathbf{0}$.

[(2) $\Longrightarrow$ (3)] The Gram-Schmidt Process shows that there is a basis for $\mathbf{R}^n$ that is orthonormal with respect to φ_A. If we let Q be the matrix whose columns are this orthonormal basis, then direct calculation shows that the φ_A innner product of the i^{th} and j^{th} columns of Q is equal to the (i,j) entry of ${}^{\mathbf{T}}Q\, A\, Q$, and therefore we know that the latter matrix is the identity. If we let $P = Q^{-1}$, then it follows that $A = {}^{\mathbf{T}}P\, P$.

[(3) $\Longrightarrow$ (1)] If $\mathbf{x}$ is nonzero then so is $P(\mathbf{x})$, and therefore we have

$$ {}^{\mathbf{T}}\mathbf{x}\, A\, \mathbf{x} \quad = \quad {}^{\mathbf{T}}\mathbf{x}\, {}^{\mathbf{T}}P\, P\, \mathbf{x} \quad = \quad |P(\mathbf{x})|^2 \quad > \quad 0\ .\blacksquare $$

Application to the Principal Minors Theorem. We can use the preceding result to prove that the principal minors of a positive definite matrix are all positive as follows. If $A^{(k)}$ is the $k \times k$ submatrix in the upper left hand corner of A, then by the theorem we know there is an invertible matrix P_k such that $A^{(k)} = {}^{\mathbf{T}}P_k\, P_k$. Since the determinant of a matrix and its transpose are equal, it follows that

$$ \det A^{(k)} \quad = \quad \det\left({}^{\mathbf{T}}P_k\, P_k\right) \quad = \quad \det\left({}^{\mathbf{T}}P_k\right)\, \det P_k \quad = \quad \left(\det P_k\right)^2 \quad > \quad 0\ . $$

The proof of the reverse implication proceeds by induction on the size of the matrix, and it will be convenient to isolate the crucial part of the inductive step.

Lemma. *Let A be a symmetric $n \times n$ matrix such that the $(n-1)\times(n-1)$ matrix in the upper left hand corner is an identity matrix and the determinant of A is positive. Then A is positive definite.*

Proof. Let $d = a_{n,n}$, and let $w_i = a_{i,n} = a_{n,i}$ for $i < n$. We can reduce A to triangular form by subtracting w_i times the i^{th} row from the last row for each $i < n$, and if we do so we obtain a matrix whose diagonal entries are equal to 1 except in the final position, where one has $d = \sum_i\, w_i^2$. It follows that the latter must be positive because it is equal to the determinant of A.

For every $n \times 1$ column vector $\mathbf{x}$ with entries x_i we have

$$ {}^{\mathbf{T}}\mathbf{x}\, A\, \mathbf{x} \quad = \quad \sum_{i=1}^{n-1}\, x_i^2 \ + \ \sum_{i=1}^{n-1}\, 2\, w_i\, x_i\, x_n \ + \ d\, x_n^2 $$

and by completing squares and using the previous formula for $\det A$ we may rewrite the right hand side as

$$ \sum_{i=1}^{n-1}\, (x_i + 2\, w_i\, x_n)^2 \ + \ (\det A)\, x_n^2\ . $$

If this expression is zero, then the positivity of the determinant implies that $x_n = 0$, and since we also have $0 = x_i + w_i\, x_n$ in this case it follows that $0 = x_i + w_i\, x_n = x_i$ for all $i < n$.$\blacksquare$

Completion of the proof of the Principal Minors Theorem. If A is a 1×1 matrix there is not much to prove because the quadratic form reduces to an second degree polynomial $a\,x^2$, and the latter is positive for all nonzero x if and only if $a > 0$. Suppose now that the result is known for $(n-1) \times (n-1)$ matrices, and let A_0 denote the $(n-1) \times (n-1)$ submatrix in the upper left hand corner of A. By induction we know that A_0 is positive definite, and therefore we may write $A_0 = {}^{\mathbf{T}}P_0\,P_0$ for some invertible matrix P_0. Let Q_0 be the inverse to P_0, and let Q be the block sum of Q_0 with a 1×1 identity matrix. Direct calculation then shows that ${}^{\mathbf{T}}Q\,A\,Q$ is a matrix which satisfies the conditions of the previous lemma (for example, its determinant is positive because it is equal to $\det A \cdot (\det Q)^2$). It follows that ${}^{\mathbf{T}}Q\,A\,Q$ can be factored as a product ${}^{\mathbf{T}}S\,S$ for some invertible matrix S, and therefore if we set B equal to $S\,Q^{-1}$ it will follow that B is invertible and $A = {}^{\mathbf{T}}B\,B$.■

Computational Procedure. Finally, here is a simple algorithmic process for determining whether the principal minors of an arbitrary square matrix are all positive:

For each i such that $1 \leq i \leq n$ attempt to carry out the following steps on a previously computed matrix A_{i-1}; we take $A_0 = A$, and part of the recursive assumption is that the (k, j) entries of A_{i-1} are zero if $k \leq i - 1$ and $j > k$. First check whether the (i, i) entry P_i of the matrix is positive. If not, stop the process and conclude that the matrix is does not have a positive i^{th} principal minor and therefore is not positive definite. If the entry is positive, then conclude that the i^{th} principal minor is positive. There are now two cases depending upon whether $i < n$ or $i = n$. In the second case, the procedure is finished, but in the first case one next performs row operations to subtract multiples of the i^{th} row from each subsequent row so that the resulting matrix A_i has all zero entries in the i^{th} column below the i^{th} row; by the recursive assumption this matrix will also have zero entries in the places where A_{i-1} was assumed to have zero entries. Furthermore, the first i diagonal entries of A_i will be the same as the first $(i-1)$ diagonal entries of A_{i-1}, and since the process has continued all these diagonal entries must be positive. The principal minors of the original matrix will be positive if and only if they are so determined by this process.

JUSTIFICATION. If $i = 1$ this process determines whether the first principal minor is positive and terminates if this is not the case. Suppose that the procedure is known to determine whether the first $(i-1)$ principal minors are positive and continues until reaching the i^{th} step. **Note that these operations do not change the principal minors of the matrix**. The determinant of A_{i-1} will be equal to the corresponding principal minor of A, so that one is positive if and only if the other is. Since the $(i-1) \times (i-1)$ matrix in the upper left corner of A_{i-1} is upper triangular it follows that its diagonal entries are the positive numbers P_k for $k < i$, and therefore the corresponding principal minor of A is positive. In fact, we also know that the $i \times i$ submatrix in the upper left hand corner of A_{i-1} is also upper triangular, and its determinant is the product of the previously computed principal minor with P_i. It follows that the i^{th} principal minor of A is positive if and only if $P_i > 0$. Assuming that it is and that $i < n$, the process for finding the next matrix A_i does not change the $i \times i$ submatrix in the upper left hand corner (hence the principal minor), nor does it change any columns before the i^{th} one. However, in the new matrix all entries in the i^{th} column below the i^{th} row are zero.■