# Why Local Minima Are Rare in High-Dimensional Landscapes

A note on Hessians, saddles, and Morse critical points

Grant Stenger

May 17, 2025

**Abstract**

A common slogan in high-dimensional optimization is that local minima are rare while saddle points are abundant. This slogan is not literally true for every smooth function: convex functions, for example, may have a unique global minimum and no saddles at all. A precise version requires a probabilistic model of the landscape and a distinction between degenerate and nondegenerate critical points. This note makes the slogan precise in three steps. First, for Morse functions, critical points are isolated, so they occupy zero volume in the ambient domain. Second, conditional on being at a nondegenerate critical point, a local minimum requires the Hessian to be positive definite, meaning that all $N$ independent curvature directions are positive. Third, in random landscape models whose Hessians behave like large centered random symmetric matrices, positive definiteness is an exponentially rare event, whereas an indefinite Hessian—and hence a saddle point—is overwhelmingly likely. The Kac–Rice formula supplies the rigorous counting framework for this phenomenon in Gaussian random fields, and Morse theory explains why the Hessian index is the natural topological invariant of a critical point.

## 1 Introduction

*"The greatest whitepill of all is that local minima are rare in high dimensional spaces."*

—Roon

High-dimensional optimization problems arise throughout machine learning, physics, statistics, and engineering. A neural network loss, a spin-glass Hamiltonian, or a complicated design objective may be a function

$$f : \mathbb{R}^N \to \mathbb{R},$$

where $N$ may be thousands, millions, or more. Our low-dimensional geometric intuition pictures such functions as landscapes of hills, valleys, and mountain passes. But this picture becomes misleading as $N$ grows. In high dimensions there are many independent directions in which a function can bend, and a nondegenerate local minimum must curve upward in every one of them.

The main claim of this note is therefore not that local minima are impossible. The claim is that, in broad random models of high-dimensional landscapes, nondegenerate local minima are rare among critical points. A critical point is a point where the gradient vanishes. A nondegenerate local minimum is a critical point whose Hessian has only positive eigenvalues. A saddle point is a critical point whose Hessian has both positive and negative eigenvalues. Thus, the question becomes a question about eigenvalue signs: among the Hessians encountered at critical points of a typical high-dimensional random function, how often are all eigenvalues positive?

The answer is: very rarely. In the simplest random-matrix heuristic, a Hessian at a typical critical point behaves like a large centered symmetric random matrix. Random matrix theory says that such matrices usually have many positive and many negative eigenvalues. Requiring all eigenvalues to be positive is a large-deviation event. For Gaussian Orthogonal Ensemble (GOE) matrices, the probability of positive definiteness decays like

$$\exp\{-cN^2 + o(N^2)\}$$

for a positive constant $c$ [5]. Thus an unbiased random Hessian is overwhelmingly likely to be indefinite, and an indefinite Hessian is exactly the second-order signature of a saddle.

One caveat should be stated immediately. A strict local minimum need not have a positive definite Hessian. For example, $f(x) = x^4$ has a strict local minimum at $x = 0$, but its second derivative there is zero. The clean Hessian classification applies to nondegenerate, or Morse, critical points. This note is about that setting: the probability that a random critical point has index 0.

The rest of the note develops this argument carefully. Section 2 reviews critical points, Hessians, and the second derivative test. Section 3 distinguishes the elementary statement that critical points occupy zero volume from the stronger high-dimensional statement that minima are rare among critical points. Section 4 gives the random-matrix argument. Section 5 explains how the Kac–Rice formula turns the heuristic into a counting method for Gaussian random fields. Section 6 explains the Morse-theoretic interpretation of the Hessian index. Section 7 states the limitations of the slogan.

## 2 Calculus of Critical Points

Throughout, let $f : \mathbb{R}^N \to \mathbb{R}$ be a twice continuously differentiable function, written $f \in C^2(\mathbb{R}^N)$. The $C^2$ assumption ensures that the Hessian is well-defined and symmetric.

**Definition 2.1** (Gradient)**.** *The* gradient *of $f$ at $x = (x_1, \ldots, x_N) \in \mathbb{R}^N$ is the vector*

$$\nabla f(x) := \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_N}(x) \end{pmatrix} \in \mathbb{R}^N.$$

**Definition 2.2** (Critical point)**.** *A point $x_c \in \mathbb{R}^N$ is a* critical point *of $f$ if*

$$\nabla f(x_c) = 0.$$

*For a smooth function on an open domain, interior local minima and local maxima must be critical points.*

**Definition 2.3** (Hessian)**.** *The* Hessian *of $f$ at $x \in \mathbb{R}^N$ is the $N \times N$ matrix*

$$H_f(x) = \nabla^2 f(x) := \left[ \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right]_{i,j=1}^{N}.$$

*Since $f \in C^2$, Clairaut's theorem implies*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x),$$

*so $H_f(x)$ is real and symmetric.*

The Hessian matters because it is the quadratic term in the Taylor expansion. If $h = x - x_0$, then

$$f(x_0 + h) = f(x_0) + \nabla f(x_0)^T h + \frac{1}{2} h^T H_f(x_0) h + o(\|h\|^2). \tag{1}$$

At a critical point $x_c$, the linear term vanishes, so the local geometry is governed to second order by

$$f(x_c + h) - f(x_c) = \frac{1}{2} h^T H_f(x_c) h + o(\|h\|^2).$$

Because $H_f(x_c)$ is real symmetric, the spectral theorem gives an orthonormal basis of eigenvectors $v_1, \ldots, v_N$ with real eigenvalues $\lambda_1, \ldots, \lambda_N$. If

$$h = \sum_{i=1}^{N} a_i v_i,$$

then

$$h^T H_f(x_c) h = \sum_{i=1}^{N} \lambda_i a_i^2. \tag{2}$$

Thus the eigenvalues of the Hessian are the principal second-order curvatures of $f$ at the critical point.

**Definition 2.4** (Nondegenerate minimum, maximum, and saddle)**.** *Let $x_c$ be a nondegenerate critical point, meaning that $H_f(x_c)$ has no zero eigenvalues.*

*(a) $x_c$ is a* nondegenerate local minimum *if all eigenvalues of $H_f(x_c)$ are positive.*

*(b) $x_c$ is a* nondegenerate local maximum *if all eigenvalues of $H_f(x_c)$ are negative.*

*(c) $x_c$ is a* saddle point *if $H_f(x_c)$ has at least one positive and at least one negative eigenvalue.*

**Definition 2.5** (Morse index)**.** *The* Morse index *of a nondegenerate critical point $x_c$ is the number of negative eigenvalues of $H_f(x_c)$:*

$$\mathrm{index}(x_c) := \#\{i : \lambda_i < 0\}.$$

*A nondegenerate local minimum has index $0$, a nondegenerate local maximum has index $N$, and a saddle point has index between $1$ and $N - 1$.*

The high-dimensional difficulty is now visible. In dimension 1, a nondegenerate minimum requires one positive second derivative. In dimension $N$, a nondegenerate local minimum requires $N$ simultaneous positive curvature conditions:

$$\lambda_1 > 0, \lambda_2 > 0, \ldots, \lambda_N > 0.$$

A saddle, by contrast, only requires at least one positive and at least one negative eigenvalue. When the Hessian is random and has no strong bias toward positive curvature, the latter condition is overwhelmingly easier to satisfy.

## 3 Critical Points Are Volume-Zero, but That Is Not the Main Point

There is a simple sense in which critical points are already rare. For a generic smooth function, the critical points are isolated, and isolated sets occupy zero volume in $\mathbb{R}^N$. This observation is true, but it is not yet the high-dimensional phenomenon that the slogan is trying to express.

**Definition 3.1** (Morse function)**.** *A $C^2$ function $f : \mathbb{R}^N \to \mathbb{R}$ is a* Morse function *if every critical point is nondegenerate, equivalently if*

$$\det H_f(x_c) \neq 0$$

*for every critical point $x_c$.*

**Proposition 3.2.** *If $f$ is a Morse function, then its critical points are isolated. In particular, in every compact subset of $\mathbb{R}^N$, there are only finitely many critical points, and therefore the set of critical points has Lebesgue measure zero.*

*Proof.* Critical points are zeros of the gradient map

$$\nabla f : \mathbb{R}^N \to \mathbb{R}^N.$$

The derivative of this map at $x_c$ is the Hessian $H_f(x_c)$. If $x_c$ is nondegenerate, then $H_f(x_c)$ is invertible. By the inverse function theorem, $\nabla f$ is locally invertible near $x_c$, so $x_c$ is the unique zero of $\nabla f$ in some neighborhood. Hence each critical point is isolated. Since the set of critical points is closed, an infinite set of critical points inside a compact region would have an accumulation point that is also critical, contradicting isolation. Thus there are only finitely many critical points in any compact region. Finite sets have Lebesgue measure zero. □

This proposition says that if one samples a point uniformly from a box in $\mathbb{R}^N$, the probability of landing exactly on a critical point is zero. But this is not special to high dimensions; it is already true for typical one-dimensional functions. The more interesting question is conditional:

> Given that we are at a critical point of a high-dimensional random function, what is the probability that this critical point is a nondegenerate local minimum?

That conditional question is answered by the signs of the Hessian eigenvalues.

## 4 The Random-Matrix Argument

A first heuristic is to model the Hessian at a typical critical point as a random real symmetric matrix. This model is not literally correct for every random function: the Hessian may be correlated with the function value, with the gradient, or with other structure in the problem. But the model captures the core mechanism. If the Hessian has no strong positive bias, then its eigenvalue spectrum should contain both signs.

### 4.1 The oversimplified independent-sign heuristic

Suppose, only for intuition, that each Hessian eigenvalue has probability 1/2 of being positive and that these signs are independent. Then

$$\mathbb{P}(\text{nondegenerate local minimum}) = \mathbb{P}(\lambda_1 > 0, \dots, \lambda_N > 0) = 2^{-N}.$$

This already decays exponentially in $N$. The independence assumption is false for eigenvalues of symmetric random matrices: eigenvalues strongly repel one another. Nevertheless, the conclusion that positive definiteness is rare remains correct, and the true random-matrix estimate is even sharper.

### 4.2 GOE Hessians and the semicircle law

The Gaussian Orthogonal Ensemble is the standard model of a centered random real symmetric matrix. Informally, an $N \times N$ GOE matrix is a symmetric matrix whose independent entries are Gaussian, with a normalization chosen so that the eigenvalues remain $O(1)$ as $N \to \infty$.

Wigner's semicircle law says that the empirical eigenvalue distribution of a centered GOE matrix converges, as $N \to \infty$, to a deterministic semicircle distribution symmetric about zero [7]. Therefore, a typical large GOE matrix has many positive and many negative eigenvalues. Its index is usually near $N/2$, not near 0 or $N$.

For a critical point to be a nondegenerate local minimum, however, the Hessian must be positive definite. In the GOE model this is the event

$$\lambda_{\min}(H) > 0.$$

This event forces the entire eigenvalue spectrum, whose natural limiting shape is centered at zero, to shift to the positive side. That is a large-deviation event.

**Theorem 4.1** (GOE positive-definiteness is exponentially rare)**.** *Let $H_N$ be an $N \times N$ centered GOE matrix, with the usual random-matrix scaling. Then there is a constant $c > 0$ such that*

$$\mathbb{P}(H_N \text{ is positive definite}) = \mathbb{P}(\lambda_{\min}(H_N) > 0) = \exp\{-cN^2 + o(N^2)\}$$

*as $N \to \infty$. More precisely, for the Gaussian $\beta$-ensembles the exponent has the form $\beta\theta(0)N^2$ with $\theta(0) = (\log 3)/4$ under the convention used by Dean and Majumdar [5].*

The exact constant is less important than the scaling. The point is that positive definiteness is not merely unlikely; for centered GOE matrices it is exponentially unlikely in $N^2$. Thus, in the random-Hessian model,

$$\mathbb{P}(\text{critical point is a saddle}) \longrightarrow 1$$

while

$$\mathbb{P}(\text{critical point is a nondegenerate local minimum}) \longrightarrow 0.$$

### 4.3 A useful geometric interpretation

At a critical point $x_c$, the second-order approximation is

$$f(x_c + h) - f(x_c) \approx \frac{1}{2} h^T H_f(x_c) h.$$

If $H_f(x_c)$ has even one negative eigenvalue, then there is a direction $v$ such that

$$v^T H_f(x_c)v < 0.$$

Moving a small distance along $v$ decreases $f$ to second order:

$$f(x_c + tv) - f(x_c) \approx \frac{1}{2}t^2 v^T H_f(x_c)v < 0.$$

So a single negative eigenvalue destroys the possibility of a nondegenerate local minimum. In high dimensions, a random Hessian has many chances to acquire such a negative direction. A nondegenerate local minimum survives only if none of these directions curve downward.

# 5 Counting Critical Points with Kac–Rice

The random-matrix argument explains why a random Hessian is unlikely to be positive definite. To connect this argument to actual random functions, one needs a way to count critical points and classify them by index. The standard tool is the Kac–Rice formula.

Let $f$ be a sufficiently regular random field on a domain $D \subset \mathbb{R}^N$. For $0 \le k \le N$, define

$$C_k(D) := \#\{x \in D : \nabla f(x) = 0,\ \mathrm{index}(H_f(x)) = k\},$$

the number of critical points of index $k$ in $D$. Under regularity and nondegeneracy assumptions, Kac–Rice gives

$$\mathbb{E}C_k(D) = \int_D \mathbb{E}[|\det H_f(x)|\mathbf{1}\{\mathrm{index}(H_f(x)) = k\} \mid \nabla f(x) = 0]\, p_{\nabla f(x)}(0)\, \mathrm{d}x. \tag{3}$$

Here $p_{\nabla f(x)}(0)$ is the density of the gradient at zero. The determinant factor appears because zeros of the gradient map are counted with the local Jacobian volume change; the indicator selects the desired index.

Equation (3) makes the problem precise. Counting nondegenerate local minima means taking $k = 0$:

$$C_0(D) = \#\{x \in D : \nabla f(x) = 0,\ H_f(x) \succ 0\}.$$

Counting saddles means summing over $1 \le k \le N - 1$.

For many Gaussian random fields, the conditional distribution of the Hessian given $\nabla f(x) = 0$ is an orthogonally invariant Gaussian matrix, often a GOE-type matrix plus a scalar shift depending on the field value [1, 4]. This is where random matrix theory enters rigorously. The index distribution of the Hessian controls the expected number of critical points of each type.

In spin-glass models, this program has been carried out in detail. Bray and Dean computed the average number of critical points of Gaussian fields on large-dimensional spaces as a function of both energy and index [3]. Auffinger, Ben Arous, and Černý gave a rigorous asymptotic analysis of the complexity of spherical $p$-spin spin-glass Hamiltonians using Kac–Rice and GOE eigenvalue asymptotics [2]. These works show a characteristic pattern: critical points may be exponentially numerous, but the dominant index is typically not 0. Rather, in much of the landscape the dominant critical points have a positive fraction of negative eigenvalues, meaning they are saddles.

Thus the rigorous probabilistic statement is not

> "Every high-dimensional function has few local minima."

It is closer to

> "In broad high-dimensional random landscape models, the expected critical-point count is dominated by saddles, while index-zero critical points form a vanishing or exponentially smaller fraction except in special low-energy regimes."

## 6 Morse Theory and the Meaning of Index

Morse theory provides a topological interpretation of the index. It does not, by itself, prove that minima are rare. Instead, it explains why the index is the correct invariant to track.

Suppose $f : M \to \mathbb{R}$ is a Morse function on a compact smooth manifold $M$. As one raises a threshold $a$ and studies the sublevel set

$$M_a := \{x \in M : f(x) \leq a\},$$

the topology of $M_a$ changes only when $a$ passes a critical value. If the critical point has index $k$, then, roughly, passing that level attaches a $k$-dimensional handle. A local minimum has index 0, so it creates a new connected component of the sublevel set. An index-1 saddle can merge components. Higher-index saddles create higher-dimensional tunnels and voids.

One consequence is the Morse relation for the Euler characteristic:

$$\sum_{k=0}^{N} (-1)^k C_k = \chi(M), \tag{4}$$

where $C_k$ is the number of index-$k$ critical points and $\chi(M)$ is the Euler characteristic of $M$ [6]. Since $\chi(M)$ is usually small compared with the total number of critical points in a complicated random landscape, many critical points must cancel in this alternating sum. Saddles of adjacent indices are naturally suited to such cancellation.

This topological picture agrees with the probabilistic one: random high-dimensional landscapes often contain a large population of saddle points spread across many indices. But it is important not to overstate the topological argument. A function on a sphere can have just one minimum and one maximum. Morse theory constrains possible critical-point counts; random matrix theory and Kac–Rice explain why, in random high-dimensional models, the bulk of the count is usually saddle-like.

## 7 Limitations and Correct Form of the Slogan

The phrase "local minima are rare in high dimensions" needs qualifications.

(1) **The statement is not universal over all smooth functions.** A strongly convex function on $\mathbb{R}^N$, such as

$$f(x) = \|x\|^2,$$

has exactly one critical point, and it is a strict global minimum. There are no saddles. Thus high dimension alone does not imply scarcity of minima.

(2) **Positive definite Hessian is a sufficient condition for a strict minimum, not a necessary one.** Degenerate strict minima exist. The one-dimensional function $f(x) = x^4$ has a strict local minimum at 0, but its Hessian there is zero. The Hessian-signature argument is therefore about nondegenerate, or Morse, critical points.

(3) **The statement depends on the random-landscape ensemble.** Centered GOE-like Hessians lead to saddle dominance. A model whose Hessian is strongly shifted in the positive direction can have many minima. In Gaussian field models, conditioning on very low function value can shift the Hessian spectrum upward, making minima more likely near the bottom of the landscape.

(4) **Rare does not mean nonexistent.** Even if the fraction of minima among all critical points tends to zero, the absolute number of minima may still grow with $N$ in some models. What becomes rare is the probability that a randomly selected critical point is a minimum.

With these qualifications, the slogan can be stated more accurately:

> In many high-dimensional random landscapes with approximately unbiased Hessian spectra, a typical nondegenerate critical point is overwhelmingly likely to be a saddle. Local minima among Morse critical points require all Hessian eigenvalues to be positive, and that positive-definiteness event becomes exponentially rare as dimension grows.

## 8 Optimization Interpretation

This helps explain why high-dimensional optimization is often described as a saddle-escape problem rather than simply a bad-local-minimum problem. The statement should not be overread: it does not prove that gradient descent always works, nor does it say that all practical objectives behave like GOE random matrices. Real losses have architecture, data, symmetries, degeneracies, and constraints.

Still, the geometric lesson is useful. In high dimensions, a critical point has many possible directions of escape. A nondegenerate local minimum must curve upward in all of them. A saddle only needs one way down. Random-like curvature therefore makes saddles the default obstruction, while true index-zero traps are special.

## 9 Conclusion

A nondegenerate local minimum in $N$ dimensions must pass $N$ simultaneous curvature tests. At a critical point, every Hessian eigenvalue must be positive. A saddle point needs only a mixture of signs. Random matrix theory says that large centered symmetric matrices naturally have such mixtures: their spectra are spread around zero, with many positive and many negative eigenvalues. The event that all eigenvalues lie on the positive side is a large deviation.

This explains the high-dimensional intuition. Critical points themselves are volume-zero features of smooth landscapes, but among critical points the decisive issue is Hessian signature. In random models, the typical signature has index near the middle, not at the extremes. Nondegenerate local minima have index 0; nondegenerate local maxima have index $N$; saddles occupy all the indices in between. As $N$ grows, the middle overwhelms the extremes.

Thus the slogan is best understood not as a theorem about every high-dimensional function, but as a robust probabilistic principle: in high-dimensional random landscapes, index-zero Morse critical points are exceptional, while saddle points are the default.

## A Worked Example: The Saddle $f(x, y) = xy$

The function

$$f(x, y) = xy$$

has a critical point at $(0, 0)$ because

$$\nabla f(x, y) = \begin{pmatrix} y \\ x \end{pmatrix},$$

so $\nabla f(0,0) = 0$. Its Hessian is

$$H_f(0,0) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The eigenvalues are 1 and $-1$, so the origin is a saddle point.

This example also shows why mixed partial terms can obscure the geometry. In the $(x, y)$ coordinates, the second derivatives $f_{xx}$ and $f_{yy}$ are both zero, while the mixed derivative $f_{xy}$ is nonzero. The principal curvature directions are not the coordinate axes. Rotate coordinates by defining

$$u = \frac{x+y}{\sqrt{2}}, \qquad v = \frac{x-y}{\sqrt{2}}.$$

Equivalently,

$$x = \frac{u+v}{\sqrt{2}}, \qquad y = \frac{u-v}{\sqrt{2}}.$$

Then

$$xy = \frac{u+v}{\sqrt{2}} \cdot \frac{u-v}{\sqrt{2}} = \frac{u^2 - v^2}{2}.$$

In the rotated coordinates, the Hessian is diagonal:

$$H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Now the saddle structure is obvious: the function curves upward in the $u$ direction and downward in the $v$ direction.

The invariant fact is not the particular matrix entries of the Hessian in a chosen coordinate system. The invariant fact is the signature: one positive eigenvalue and one negative eigenvalue. This is a special case of Sylvester's law of inertia, which says that the numbers of positive, negative, and zero directions of a real quadratic form are preserved under invertible linear changes of coordinates.

# B Coordinate Invariance and the Morse Lemma

Near a nondegenerate critical point, every smooth function looks like its Hessian normal form after a smooth change of coordinates. This is the content of the Morse lemma.

**Theorem B.1** (Morse lemma)**.** *Let $f : M \to \mathbb{R}$ be smooth, and let $p \in M$ be a nondegenerate critical point of index $k$. Then there are local coordinates $(u_1, \dots, u_N)$ centered at $p$ such that*

$$f(u) = f(p) - u_1^2 - \cdots - u_k^2 + u_{k+1}^2 + \cdots + u_N^2.$$

Thus, up to smooth local coordinate change, a nondegenerate critical point is completely classified by its index. A nondegenerate local minimum has normal form

$$f(p) + u_1^2 + \cdots + u_N^2,$$

a nondegenerate local maximum has normal form

$$f(p) - u_1^2 - \cdots - u_N^2,$$

and an index-$k$ saddle has $k$ downward directions and $N - k$ upward directions. This is why the index is the natural language for high-dimensional critical points.

## References

[1] R. J. Adler and J. E. Taylor. *Random Fields and Geometry*. Springer Monographs in Mathematics. Springer, 2007.

[2] A. Auffinger, G. Ben Arous, and J. Černý. Random matrices and complexity of spin glasses. *Communications on Pure and Applied Mathematics*, 66(2):165–201, 2013.

[3] A. J. Bray and D. S. Dean. Statistics of critical points of Gaussian fields on large-dimensional spaces. *Physical Review Letters*, 98:150201, 2007.

[4] D. Cheng and A. Schwartzman. Expected number and height distribution of critical points of smooth isotropic Gaussian random fields. *Bernoulli*, 24(4B):3422–3446, 2018.

[5] D. S. Dean and S. N. Majumdar. Large deviations of extreme eigenvalues of random matrices. *Physical Review Letters*, 97:160201, 2006.

[6] J. Milnor. *Morse Theory*. Princeton University Press, 1963.

[7] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Annals of Mathematics*, 67(2):325–327, 1958.