Accelerating mathematical research with language models: A case study of an interaction with GPT-5-Pro on a convex analysis problem

Adil Salim

October 30, 2025

Abstract

Recent progress in large language models has made them increasingly capable research assistants in mathematics. Yet, as their reasoning abilities improve, evaluating their mathematical competence becomes increasingly challenging. The problems used for assessment must be neither too easy nor too difficult, their performance can no longer be summarized by a single numerical score, and meaningful evaluation requires expert oversight.

In this work, we study an interaction between the author and a large language model in proving a lemma from convex optimization. Specifically, we establish a Taylor expansion for the gradient of the biconjugation operator—that is, the operator obtained by applying the Fenchel transform twice—around a strictly convex function, with assistance from GPT-5-pro, OpenAI's latest model.

Beyond the mathematical result itself, whose novelty we do not claim with certainty, our main contribution lies in documenting the collaborative reasoning process. GPT-5-pro accelerated our progress by suggesting, relevant research directions and by proving some intermediate results. However, its reasoning still required careful supervision, particularly to correct subtle mistakes. While limited to a single mathematical problem and a single language model, this experiment illustrates both the promise and the current limitations of large language models as mathematical collaborators.

1 Introduction

1.1 Mathematical context

More than two centuries ago, Monge (1781) formulated a question that arises naturally across several areas of science: Given two sets $U, V \subset \mathbb{R}^d$ of equal volume, find the optimal volume-preserving map between them, where optimality is measured with respect to a cost function $c(x, y) \geq 0$. This problem is now known in the mathematical literature as the *optimal transport problem*.

In a breakthrough result, Brenier (1991) showed that when $c(x, y) = \|x - y\|^2$, the optimal map exists, is unique, and is the gradient of a convex potential. Later, Gangbo (1994) simplified Brenier's proof. A key ingredient of this simplification is the following simple convex analysis lemma (Lemma 2.4 therein):

$$(\phi^* + t\,h)^*(x) = \phi(x) - t\,h(\nabla\phi(x)) + o(t), \tag{1}$$

for almost every (a.e.) $x \in \mathbb{R}^d$, where $\phi : \mathbb{R}^d \to \mathbb{R}$ is convex, h is a (possibly nonconvex) test function, and * denotes the Fenchel transform[1]. This lemma was later extended to c-transforms by Gangbo and McCann (1996), to solve the optimal transport problem for general costs of the form $c(x, y) = h(\|x - y\|)$, with h strictly convex.

As part of a research project on optimal transport, we aimed to prove the following variant of 1, assuming ϕ is strictly convex:

$$\nabla(\phi + th)^{**}(x) = \nabla\phi(x) + t\nabla h(x) + o(t). \tag{2}$$

Without additional assumptions, we could not rigorously derive (2) by applying (1) twice in a straightforward way, nor did we find it stated in the convex analysis literature. However, given its apparent simplicity, we do not claim the novelty of (2) with certainty.

[1] The expansion is exact at $t = 0$ because $(\phi^*)^* = \phi$.

In this paper, we present a proof of Equation (2). In particular, we show that the remainder term $o(t)$ is equal to zero if $|t|$ is small enough (depending on x). But, beyond the simple mathematical result, our main contribution is to document the interaction with GPT-5-pro that led to the proof of (2).

1.2 Social context

Language models such as Claude, Gemini and GPT are machine learning systems trained on vast amounts of text data to answer prompts without human supervision. Their performance is typically measured by how much they can help—or potentially harm—the human user.

Although they can be tested on nearly any domain of human knowledge, they are often evaluated on mathematics, as it is viewed as a direct measure of reasoning or "intelligence".

With the recent increase in the reasoning abilities of language models, evaluating them in mathematics has become increasingly challenging. Many benchmark problems are now too easy, while most research questions remain too difficult. An alternative to automated benchmarks is to evaluate language models as research assistants. That requires setting up realistic research collaborations—a process that is inherently qualitative and non-scalable, since it depends on human experts, extended interactions, and subjective judgment.

Since the release of GPT-5-pro by OpenAI in August 2025, several mathematicians have documented their use of GPT-5-pro to solve math problems. In a post on X, Sébastien Bubeck showed how GPT-5-pro could improve a bound on the step size of an optimization algorithm. Then, Diez et al. (2025) used GPT-5-pro to compute the convergence speed of a Central Limit Theorem. Later on, Feldman and Karbasi (2025) submitted five combinatorial optimization open questions to GPT-5-pro, and the model could solve three of them. In a post on X, Paata Ivanisvili showed that GPT-5-pro could provide a counter example to a conjecture made in Fall 2013, in the context of a Simons Institute program on real analysis. Finally, Ernest Ryu recently announced on X that he could prove the pointwise convergence of the Nesterov ODE with help of GPT-5-pro.

1.3 Contribution

Beyond Diez et al. (2025), we have limited available evaluation data of frontier language models as research collaborators in mathematics. The goal of this study is to add a data point to this evaluation.

To this end, we focus on proving Equation (2) with help of GPT-5-pro, and we document our interaction with the model. We believe that proving Equation (2) is an interesting math problem that is not trivial, but simple enough to be discussed with a language model and to be understood by a large number of mathematicians.

Beyond the mathematical result (2), this study gives an example of a productive use of a language model for research. GPT-5-pro accelerated our progress by suggesting relevant research directions such as the key conjecture, and by proving some lemmas. But its reasoning still required careful supervision, particularly to correct mistakes.

1.4 Outline

This paper is organized as follows. In Section 2, we state and prove the main mathematical result (2). The text in blue indicates the ideas that mostly came from GPT-5-pro and the text in black indicates the ideas that mostly came from us. In Section 3, we summarize the interaction we had with the language model. We conclude in Section 4. The full chat can be found in Appendix, with our comments on some of the model's mistakes.

2 Main mathematical result

Before presenting our main result, we specify some notations. For every function $f : \mathbb{R}^d \to (-\infty, +\infty]$, we denote by f^* its Fenchel transform

$$f^*(y) := \sup_{x \in \mathbb{R}^d} \langle x, y \rangle - f(x). \tag{3}$$

The biconjugate $(f^*)^*$ of f is denoted f^{**}. For every $x \in \mathbb{R}^d$, the subdifferential of f at x, denoted $\partial f(x)$, is the subset of $\mathbb{R}^d$ defined by

$$p \in \partial f(x) \iff \forall y \in \mathbb{R}^d, \quad f(y) \geq f(x) + \langle p, y - x \rangle. \tag{4}$$

We now state our main result.

Proposition 1. Let $x \in \mathbb{R}^d$ and $\phi : \mathbb{R}^d \to \mathbb{R}$ be convex and twice differentiable at x. Assume $\nabla^2\phi(x)$ symmetric positive definite, and set

$$\lambda := \lambda_{\min}\big(\nabla^2\phi(x)\big) \ > \ 0.$$

Let $h \in C_c^2(\mathbb{R}^d)$ with compact support, and let $L := \|\nabla^2 h\|_\infty$ and $M := \|h\|_\infty$. Define $f := \phi + th$ and set

$$t_x := \min\left(\frac{\lambda}{4L}, \frac{\lambda\delta^2}{64M}\right) > 0,$$

where $\delta > 0$ is a constant depending on x defined in the proof. Then, for every $t \in (-t_x, t_x)$,

$$(\phi + th)^{**}(x) \ = \ (\phi + th)(x), \tag{5}$$

and

$$\nabla(\phi + th)^{**}(x) = \nabla\phi(x) + t\nabla h(x). \tag{6}$$

Remark 1. The subtlety of Proposition 1 lies in its locality. We do not know *a priori* if $(\phi + th)$ is convex for $|t|$ small enough, nor we know if $\nabla^2\phi$ (well defined a.e. because ϕ is convex) is positive definite in a neighborhood of x[2]. Hence, Equation (5), which is valid at x, might fail to hold around x, and Equation (6) cannot be obtained directly by differentiating Equation 5.

Proof. By Taylor at x, there exists a function $r : \mathbb{R}^d \to \mathbb{R}$ with $r(u) \to 0$ as $u \to 0$ such that,

$$\phi(x + u) = \phi(x) + \langle \nabla\phi(x), u\rangle + \frac{1}{2}\langle u, \nabla^2\phi(x)u\rangle + r(u)\,\|u\|^2. \tag{7}$$

Using $\langle u, \nabla^2\phi(x)u\rangle \geq \lambda\|u\|^2$,

$$\phi(x + u) \ \geq \ \phi(x) + \langle \nabla\phi(x), u\rangle + \left(\frac{\lambda}{2} + r(u)\right)\|u\|^2. \tag{8}$$

There exists $\delta > 0$ such that $|r(u)| < \lambda/4$ if $\|u\| < \delta$. Therefore, for every u such that $\|u\| < \delta$,

$$\phi(x + u) \ \geq \ \phi(x) + \langle \nabla\phi(x), u\rangle + \frac{\lambda}{4}\|u\|^2. \tag{9}$$

Besides, for every $t \in \mathbb{R}$, $\nabla(th)$ is $|t|L$–Lipschitz. Therefore,

$$(th)(x + u) \ \geq \ (th)(x) + \langle \nabla(th)(x), u\rangle - \frac{|t|L}{2}\|u\|^2. \tag{10}$$

1) On the ball of radius δ centered at x. Adding (9) and (10) gives, for $\|u\| \leq \delta$,

$$f(x + u) \ \geq \ f(x) + \langle \nabla f(x), u\rangle + \left(\tfrac{\lambda}{4} - \tfrac{|t|L}{2}\right)\|u\|^2. \tag{11}$$

Taking $|t| < \lambda/4L$, the bracket is at least $\lambda/8$,

$$f(x + u) \ \geq \ f(x) + \langle \nabla f(x), u\rangle + \frac{\lambda}{8}\|u\|^2 \qquad \text{for all } \|u\| \leq \delta. \tag{12}$$

[2]Even if we were assuming ϕ strictly convex, we would not know *a priori* that the eigenvalues of $\nabla^2\phi$ are uniformly bounded below by a positive constant in a neighborhood of x.

2) On the sphere of radius δ centered at x. If $\|u\| = \delta$, the bound (11) yields a positive margin

$$f(x+u) \;\geq\; f(x) + \langle \nabla f(x), u\rangle + \frac{\lambda}{8}\delta^2. \tag{13}$$

We write it in terms of ϕ,

$$\phi(x+u) \;\geq\; \phi(x) + \langle \nabla f(x), u\rangle + \frac{\lambda}{8}\delta^2 + t(h(x) - h(x+u)). \tag{14}$$

Since $|t(h(x) - h(x+u))| \leq 2|t|M$, if $|t| < \lambda\delta^2/(32M)$ we have $t(h(x) - h(x+u)) > -\lambda\delta^2/16$. Therefore, for all $u \in \mathbb{R}^d$ such that $\|u\| = \delta$,

$$\phi(x+u) \;\geq\; \phi(x) + \langle \nabla f(x), u\rangle + \frac{\lambda}{16}\delta^2. \tag{15}$$

The r.h.s. defines an affine map that we denote $L(u) := \phi(x) + \langle \nabla f(x), u\rangle + \frac{\lambda}{16}\delta^2$.

4) Outside the ball of radius δ centered at x. We have shown that $\phi(x+u) \geq L(u)$ if $\|u\| = \delta$. We first show that this is still true if $\|u\| > \delta$. Assume by contradiction that there exists $\bar{u}$ such that $\|\bar{u}\| > \delta$ and $\phi(x+\bar{u}) < L(\bar{u})$. Note that we obviously have $\phi(x) = \phi(x+0) < L(0)$. Let $g(u) = \phi(x+u) - L(u)$. Then g is convex, $g(0) < 0$ and $g(\bar{u}) < 0$. By convexity, $g(\alpha\bar{u}) < 0$ for every $\alpha \in [0,1]$. Noting that $\alpha = \delta/\|\bar{u}\| \in [0,1]$, we have $\phi(x+\alpha\bar{u}) < L(\alpha\bar{u})$. However, $\|\alpha\bar{u}\| = \delta$ which contradicts (15). Therefore, for all u s.t. $\|u\| \geq \delta$, $\phi(x+u) \geq L(u)$, *i.e.*,

$$\phi(x+u) \;\geq\; \phi(x) + \langle \nabla f(x), u\rangle + \frac{\lambda}{16}\delta^2. \tag{16}$$

Since $|t(h(x) - h(x+u))| \leq 2|t|M$, if $|t| < \lambda\delta^2/(64M)$ we have

$$t(h(x+u) - h(x)) > -\frac{\lambda\delta^2}{32}.$$

Adding the last two inequalities, we obtain

$$f(x+u) \;\geq\; f(x) + \langle \nabla f(x), u\rangle + \frac{\lambda}{32}\delta^2, \qquad \text{for all } \|u\| \geq \delta \tag{17}$$

Conclusion. Considering both cases $\|u\| \leq \delta$ or $\|u\| \geq \delta$, we have using (12) and (17)

$$f(x+u) \;\geq\; f(x) + \langle \nabla f(x), u\rangle \qquad \text{for all } u \in \mathbb{R}^d, \tag{18}$$

and the inequality is strict if $u \neq 0$. Equivalently,

$$f(y) \;\geq\; f(x) + \langle \nabla f(x), y - x\rangle \qquad \text{for all } y \in \mathbb{R}^d. \tag{19}$$

Consider the subset of $\mathbb{R}$ defined by

$$E := \{a(x), \quad a : \mathbb{R}^d \to \mathbb{R} \text{ affine function s.t. } \forall y \in \mathbb{R}^d, a(y) \leq f(y)\}.$$

Taking $a(y) = f(x) + \langle \nabla f(x), y - x\rangle$, we see using (19) that $f(x) \in E$. Besides, for all $\alpha \in E$, $\alpha \leq f(x)$. Therefore $f(x) = \sup E$. Using the characterization of f^{**} as the affine envelope of f (see (Rockafellar, 1970, Section 12)), $f^{**}(x) = \sup E$. We conclude that $f^{**}(x) = f(x)$, *i.e.*, that Equation (5) holds.

Gradient. Since f^{**} is convex and $f^{**}(y) \leq f(y) < \infty$ for all $y \in \mathbb{R}^d$, $\partial f^{**}(y) \neq \emptyset$ using (Rockafellar, 1970, Theorem 23.4). Let $z \in \partial f^{**}(x)$, *i.e.*,

$$f^{**}(x+u) \ \geq \ f^{**}(x) + \langle z, u\rangle \qquad \text{for all } u \in \mathbb{R}^d. \tag{20}$$

Since $f^{**}(x+u) \leq f(x+u)$ and $f^{**}(x) = f(x)$,

$$f(x+u) \ \geq \ f(x) + \langle z, u\rangle \qquad \text{for all } u \in \mathbb{R}^d. \tag{21}$$

Replacing u by εu, $\varepsilon > 0$, $\|u\| = 1$, we rewrite the last inequality as

$$\frac{f(x+\varepsilon u) - f(x)}{\varepsilon} \ \geq \ \langle z, u\rangle \qquad \text{for all } \|u\| = 1. \tag{22}$$

Letting $\varepsilon \to 0$ and using that f is differentiable at x,

$$\langle \nabla f(x), u\rangle \ \geq \ \langle z, u\rangle \qquad \text{for all } \|u\| = 1, \tag{23}$$

which classically implies $z = \nabla f(x)$. In other words, $\partial f^{**}(x) = \{\nabla f(x)\}$. Using (Rockafellar, 1970, Theorem 25.1), f^{**} is differentiable at x and $\nabla f^{**}(x) = \nabla f(x)$. □

Proposition 2. Let $\phi : \mathbb{R}^d \to \mathbb{R}$ be strictly convex and $h \in C_c^2(\mathbb{R}^d)$. Let $F(x,t) := \nabla(\phi + th)^{**}(x)$. Then, for a.e. $x \in \mathbb{R}^d$, F is derivable at $t = 0$ and

$$\partial_2 F(x, 0) = \nabla h(x).$$

Proof. The Alexandrov Hessian $\nabla^2\phi(x)$ exists and is symmetric positive definite for a.e. $x \in \mathbb{R}^d$ (see (Ambrosio et al., 2008, Theorem 5.5.4)). Using Proposition 1,

$$\frac{\nabla(\phi+th)^{**}(x) - \nabla\phi(x)}{t} = \nabla h(x),$$

if $|t| < t_x$. Recalling that $\phi = \phi^{**}$ and letting $t \to 0$ we get the conclusion. □

The last proposition is equivalent to Equation (2).

3 Summary of the discussion

We now summarize the chat with GPT-5-pro presented in the Appendix.

The broad context of the discussion was an optimal transport problem: compute a velocity field at $t = 0$ for the curve $t \mapsto \nu_t := \nabla(\phi + th)^{**}\#\mu$, where μ is a given probability measure and $\#$ the pushforward operation. To this end, we wanted to establish (2) for a.e. $x \in \mathbb{R}^d$.

When we asked GPT-5-pro to prove Proposition 1 directly, the model mostly provided incorrect proofs. It correctly established a quadratic lower bound on ϕ around x first, but then wrote that this lower bound implies strong convexity of ϕ around x. However, we made no inference about the continuity of the Alexandrov Hessian $\nabla^2\phi$, and so we do not know that around x, the eigenvalues of $\nabla^2\phi$ are uniformly bounded below by a positive constant.

After pointing out this issue, GPT-5-pro proposed a slightly different approach relying on showing that (i) the Taylor expansion of $(\phi+th)^{**}(x)$ is valid at x only, and not necessarily on a neighborhood of x (ii) the $o(t)$ remainder term in the Taylor expansion is actually zero for $|t|$ small enough. Even if GPT-5-pro did not manage to prove these two items immediately, (i) and (ii) constitute the key conjecture. Next, we suggested the language model a technique to prove (i) and (ii). More precisely, we asked GPT-5-pro to show that $(\phi + th)$ is above its tangent at x (see the "Conclusion" paragraph in the proof of Proposition 1).

GPT-5-pro correctly explained why $(\phi + th)$ is above its tangent at x on a ball of radius $r > 0$ centered at x. It correctly suggested that this result can be extended outside the ball (see in the Appendix references to the "boundary gap"), but it made mistakes when trying to prove that result, repeatedly using that $-\|y - x\| > -r$ for y outside the ball. We completed the proof of Equation (5) ourselves.

Because Equation (5) is, in general, not valid in a neighborhood of x, one cannot differentiate it to directly obtain (6). Instead, GPT-5-pro wrote in one shot a proof of (6) relying on a convex analysis argument.

In summary, GPT-5-pro proposed several interesting directions. Although it did not produce a fully correct proof, it identified the key conjecture, established several intermediate results, and provided enough insight for us to complete the proof efficiently.

4 Conclusion

We investigated the collaboration between the author and GPT-5-pro in establishing Equation (2). The language model significantly accelerated the research process, although human supervision remained necessary.

This study presents a qualitative case limited to a single task and a single language model. While it highlights the promise of language models as research assistants, developing systematic and scalable frameworks to assess such collaborations remains an important direction for future work.

5 Acknowledgment

We thank Wilfrid Gangbo and Dimitri Shlyakhtenko for insightful discussions. Wilfrid Gangbo reviewed the proof of Proposition 1 and made several comments on this study, some of which were included. Part of this work was done while the author was visiting the Institute for Pure and Applied Mathematics (IPAM) at UCLA.

References

Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.

Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

Charles-Philippe Diez, Luis da Maia, and Ivan Nourdin. Mathematical research with gpt-5: a malliavin-stein experiment. *arXiv preprint arXiv:2509.03065*, 2025.

Moran Feldman and Amin Karbasi. Gödel test: Can large language models solve easy conjectures? *arXiv preprint arXiv:2509.18383*, 2025.

Wilfrid Gangbo. An elementary proof of the polar factorization of vector-valued functions. *Archive for rational mechanics and analysis*, 128:381–399, 1994.

Wilfrid Gangbo and Robert J McCann. The geometry of optimal transportation. 1996.

Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, pages 666–704, 1781.

R Tyrrell Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

A Interaction with ChatGPT-5-pro

The following chat has been minimally edited for clarity. We comment in red on some of the mistakes of the language model. The broad topic of the chat was an optimal transport problem. While we extracted the part of the chat that is relevant to our Equation (2), the reader might see references to a broader math problem. We indicate the text that was erased using [...].

Author's prompt

OK. but I am not sure of the moreau envelope smoothing showing that the g_ϵ converges to zero.
Nevermind. I think we could greatly simplify the proof if we use $(\phi + th)^{**}$ instead. Because formally the Taylor expansion should give $\phi + th + o(t)$, and so the A matrix will not appear after taking the gradient.

1. Find a reference for that. If you dont find, try to prove it
2. Make a list of all elements of the proof that we need to adapt. I think the first item is to take the gradient in this taylor expansion
3. Prove all items in the list

Let's go!

Response of GPT-5-pro

Love the idea—switching to the path

$$\phi_t := (\phi + t\,h)^{**} \qquad (t \text{ small})$$

so that we always stay inside the class "gradient of a convex potential" and, morally, the first–order perturbation in t is just h. Below I (i) point to the right references, (ii) list precisely what must be adapted, and (iii) prove each item at the level we need. I'll be very explicit about where we rely on standard theorems (with citations) and where we give short self-contained arguments.

---

1) References for the first-order expansion of $(\phi + th)^{**}$

The right language is **epi-differentiability** and **tilt-stability** in variational analysis.

- **Rockafellar–Wets, Variational Analysis (1998)**:
  - Epi-convergence and epi-derivatives (Ch. 7, 13).
  - Stability at points with **single subgradient** (i.e., where the convex ϕ is differentiable, equivalently $\partial\phi(x) = \{\nabla\phi(x)\}$).
  - **Key fact:** biconjugation $f \mapsto f^{**}$ is epi-continuous and preserves first-order epi-derivatives; at points where the active subgradient is unique, the first epi-derivative of $t \mapsto (f + th)^{**}$ in direction h is exactly h (no convexification needed).
  - See especially the "second-order convex analysis" notes by Rockafellar for quick access: first-order expansions via epi-derivatives and the proto-derivative of the subdifferential mapping (OU lecture notes "Second-Order Convex Analysis").
- **Poliquin–Rockafellar, "A calculus of epi-derivatives applicable to nonconvex functions" (1996)**: chain rules for epi-derivatives of composites and for perturbations $f + th$ with smooth h. In particular, when the outer operation is biconjugation (closed convex hull), the first epi-derivative passes through the operation, and if $\partial\phi(x)$ is a singleton, the epi-derivative at ϕ in direction h is just h (pointwise at x).

Author's comment: Here, the model seems to be confused between Taylor expansion of $(\phi + th)^{**}(x)$, where h is a function, and Taylor expansion of $\phi^{**}(x + th)$, where h is a vector. The "takeaway" below do not seem to follow from the references cited above.

Takeaway: At μ_0-a.e. x where ϕ is twice differentiable with $\nabla^2\phi(x) \succ 0$ (true by Alexandrov, and the positive definiteness is the transport one-to-one a.e. fact you already used), the map

$$t \longmapsto (\phi + th)^{**}(x)$$

admits the first–order expansion

$$(\phi + th)^{**}(x) = \phi(x) + t\,h(x) + o(t) \quad \text{and} \quad \nabla(\phi + th)^{**}(x) = \nabla\phi(x) + t\,\nabla h(x) + o(t),$$

for every $h \in C_c^2(\mathbb{R}^d)$.

If you prefer an "elementary" proof in our setting (no heavy epi-derivative jargon), see Item (A) below: we give a local argument that near any point where $\nabla^2\phi(x) \succ 0$, the convexification has **no effect to first order**—indeed, for small $|t|$ the biconjugate coincides with $\phi + th$ on a small ball around x, so the gradient expansion becomes $\nabla\phi(x) + t\nabla h(x) + o(t)$.

Author's comment: The statement "the biconjugate coincides with $\phi + th$ on a small ball around x" is incorrect, but we will see that it can be corrected. [...]

3) Proofs of the items

I'll first give an **elementary local argument** for the first-order expansion (Item 1), then indicate how the epi-derivative reference encapsulates it. The remaining items then follow quickly.

(A) Local first-order expansion of $(\phi + th)^{**}$ near points with $\nabla^2\phi(x) \succ 0$

Fix x where ϕ is twice differentiable with $\nabla^2\phi(x) \succ 0$. Let $\lambda_{\min} > 0$ be the smallest eigenvalue of $\nabla^2\phi(x)$. By Alexandrov's second-order expansion, for small $r > 0$,

$$\phi(x + u) = \phi(x) + \langle\nabla\phi(x), u\rangle + \tfrac{1}{2}\langle\nabla^2\phi(x)u, u\rangle + o(\|u\|^2) \quad \text{for } \|u\| \le r.$$

Hence, shrinking r if needed, there is $m \in (0, \lambda_{\min})$ such that

$$\phi(x + u) \;\ge\; \phi(x) + \langle\nabla\phi(x), u\rangle + \tfrac{m}{2}\|u\|^2 \quad \text{for } \|u\| \le r. \tag{1}$$

Let $M := \sup_{B(x,r)} \|\nabla^2 h\|$ (finite since $h \in C_c^2$). Choose $|t| \le t_0 := m/(2M)$. Then on $B(x, r)$,

$$(\phi + th)(x + u) \;\ge\; \phi(x) + \langle\nabla\phi(x), u\rangle + \tfrac{m}{2}\|u\|^2 \;+\; t\big(h(x) + \langle\nabla h(x), u\rangle\big) - |t|\,\tfrac{M}{2}\|u\|^2,$$

which gives the **local strong convexity** estimate

$$(\phi + th)(x + u) \;\ge\; (\phi + th)(x) + \langle\nabla\phi(x) + t\nabla h(x), u\rangle + \tfrac{m - |t|M}{2}\|u\|^2 \;\ge\; (\cdots) + \tfrac{m}{4}\|u\|^2 \tag{2}$$

for $|t| \le t_0$. In particular, $\phi + th$ is $m/4$-strongly convex on $B(x, r)$; thus it is $C^{1,1}$ there and

$$\nabla(\phi + th)(x) = \nabla\phi(x) + t\,\nabla h(x). \tag{3}$$

Author's comment: Inequality (2) proves that $\phi + th$ is above a quadratic approximation on the ball, but this does not imply strong convexity on the ball. The end of the proof is wrong because it relies on this strong convexity on the ball. [...]

Author's prompt

You said: "General fact (finite dimension): if a scalar function $f : \mathbb{R}^d \to \mathbb{R}$ is twice Fréchet differentiable at a point y_0 with second derivative H, then its gradient is strictly (Fréchet) differentiable at y_0 with derivative H. Equivalently, (SD) holds with $B = H$. This is standard; see, e.g., Rockafellar–Wets, *Variational Analysis*, §13.E (second-order expansion ⇒ strict differentiability of the gradient)."
But it seems that Rockafellar–Wets gives strict differentiability of the function (not its gradient), correct?
Author's comment: Here, we challenge a statement made by the model in previous messages that are not reported here.

Response of GPT-5-pro

You're right to question that citation.
Author's comment: [...]

Route 1 (what we're now doing): switch to the primal path $\phi_t = (\phi + th)^{**}$

For convex ϕ that is twice differentiable at x with **positive-definite** Hessian $\nabla^2\phi(x)$, one has a **local strong convexity** around x. A simple quadratic upper/lower sandwich shows that, for $|t|$ small and z near x, the convexification is **inactive**:

$$(\phi + th)^{**} = \phi + th \quad \text{on a small ball around } x.$$

Author's comment: The last equation is close to the key conjecture, but the statement is true at x only and not in a neighborhood of x.
Consequently,

$$\nabla(\phi + th)^{**}(x) = \nabla\phi(x) + t\,\nabla h(x) \quad \text{and} \quad \frac{d}{dt}\Big|_{t=0} \nabla(\phi + th)^{**}(x) = \nabla h(x)$$

without any need for strict differentiability of $\nabla\phi^*$. *This is the elementary local argument we already wrote out in detail (local strong convexity + "convexification doesn't move the graph to first order").*
Author's comment: [...]

Bottom line for your specific question

- **Yes**, your suspicion is correct: RW §13.E does not directly give strict differentiability of the **gradient** from second-order differentiability of f. It gives the **second-order expansion of** f (equivalently, Fréchet differentiability of ∇f at a point), which is weaker than the pairwise estimate we labeled (SD).
- In our proof we no longer need that stronger (SD) fact, because with the **primal variation** $(\phi + th)^{**}$ the first-order formula

  $$\nabla(\phi + th)^{**}(x) = \nabla\phi(x) + t\,\nabla h(x) + o(t)$$

  follows from a local "convexification is inactive" lemma—an elementary argument that uses only convexity plus the positive-definite Hessian at x. This bypasses the delicate "strict differentiability of $\nabla\phi^*$" issue altogether.

Author's comment: [...]

Author's prompt

Add more details for Route 1

Response of GPT-5-pro

Great—here is **Route 1** written out in full detail, with every "why" spelled out. The core idea is to vary along

$$\phi_t \;:=\; (\phi + t\,h)^{**} \quad (t \text{ small},\; h \in C_c^2(\mathbb{R}^d)),$$

so we remain in the class of convex potentials, and to show that **to first order** the convexification is *inactive* near almost every point x where ϕ has a positive–definite Hessian. This yields

$$\nabla\phi_t(x) \;=\; \nabla\phi(x) + t\,\nabla h(x) + o(t) \quad (\mu_0\text{-a.e.}),$$

i.e. the velocity field is simply $v_0(y) = \nabla h(\nabla\phi^*(y))$. We then use **OU smoothing after the pushforward** to compute the entropy derivative classically, and conclude as before.
Below: (0) the "good" set and why $\nabla^2\phi \succ 0$ μ_0-a.e.; (1) the local "convexification is inactive" lemma; (2) the gradient expansion and velocity; (3) why we still smooth, and why **after** the pushforward; (4) the one-sided derivatives and the limiting Stein identity; (5) the completion to the **standard** Stein identity.

---

0) The "good" set and positive–definite Hessian a.e.

- $\mu_0 \ll \mathcal{L}^d$ and $\nu_0 := \nabla\phi\#\mu_0 \ll \mathcal{L}^d$ (since $\mathsf{K}L(\nu_0 \mid \pi) \le \mathsf{K}L(\mu_0 \mid \pi) < \infty$).
- By Alexandrov, ϕ and ϕ^* are twice differentiable a.e. w.r.t. Lebesgue; since μ_0, ν_0 are a.c., we may pick a set $G \subset \mathbb{R}^d$ of full μ_0-measure such that for all $x \in G$:
  - $\nabla\phi(x)$ exists; set $y_0 = \nabla\phi(x)$.
  - $\nabla\phi^*(y_0)$ exists and equals x (subgradient reciprocity + ν_0-a.e. differentiability).
  - $\nabla^2\phi(x)$ and $\nabla^2\phi^*(y_0)$ both exist. Differentiating $\nabla\phi^*(\nabla\phi(x)) = x$ at such x gives

    $$\nabla^2\phi^*(y_0)\,\nabla^2\phi(x) = I_d.$$

    Hence $\nabla^2\phi(x) \succ 0$, $\nabla^2\phi^*(y_0) \succ 0$, and they are inverses.

We work pointwise on this full μ_0-measure set.

---

1) "Convexification is inactive" near x

Lemma (local inactivity of biconjugation). Fix $x \in G$, let $H := \nabla^2\phi(x) \succ 0$ and $h \in C_c^2$. There exist $r > 0$ and $t_0 > 0$ such that for all $|t| \le t_0$,

$$(\phi + th)^{**}(z) \;=\; \phi(z) + t\,h(z) \quad \text{for all } z \in B(x, r),$$

and consequently

$$\nabla(\phi + th)^{**}(x) \;=\; \nabla\phi(x) + t\,\nabla h(x).$$

Proof (fully elementary).

1. *Local quadratic lower bound for ϕ.* By Alexandrov at x, for small $r_1 > 0$,

   $$\phi(x+u) \;\ge\; \phi(x) + \langle\nabla\phi(x), u\rangle + \tfrac{1}{2}\langle Hu, u\rangle - \varepsilon_1\|u\|^2, \qquad \|u\| \le r_1,$$

   with $\varepsilon_1 \in (0, \frac{1}{2}\lambda_{\min}(H))$. Set $m := \lambda_{\min}(H) - 2\varepsilon_1 > 0$. Shrinking r_1 if needed,

   $$\phi(x+u) \;\ge\; \phi(x) + \langle\nabla\phi(x), u\rangle + \tfrac{m}{2}\|u\|^2, \qquad \|u\| \le r_1. \tag{1}$$

2. *Control h.* Let $M := \sup_{B(x,r_1)} \|\nabla^2 h\| < \infty$. For all $|t| \le t_0 := m/(2M)$,

$$h(x+u) \;\ge\; h(x) + \langle \nabla h(x), u\rangle - \tfrac{M}{2}\|u\|^2.$$

3. *Strong convexity of $\phi + th$ on $B(x, r_1)$.* Combine (1) with the previous bound: for $|t| \le t_0$ and $\|u\| \le r_1$,

$$(\phi+th)(x+u) \;\ge\; (\phi+th)(x) + \langle \nabla\phi(x) + t\nabla h(x), u\rangle + \tfrac{m-|t|M}{2}\|u\|^2 \;\ge\; (\cdots) + \tfrac{m}{4}\|u\|^2. \tag{2}$$

   Thus $\phi + th$ is $C^{1,1}$ and $\frac{m}{4}$-strongly convex on $B(x, r_1)$, and in particular

$$\nabla(\phi+th)(x) \;=\; \nabla\phi(x) + t\,\nabla h(x). \tag{3}$$

4. *Convexification does nothing near x.*

   Author's comment: The argument below relies again on strong convexity on the ball, which is not true. [...]

   Strong convexity on a ball implies a Jensen–variance inequality: for any convex combination $z = \sum \lambda_i z_i$ with $z_i \in B(x, r_1)$,

$$\sum \lambda_i(\phi+th)(z_i) \;\ge\; (\phi+th)(z) + c\sum \lambda_i\|z_i - z\|^2, \qquad c := \tfrac{m}{8} > 0. \tag{4}$$

   Choose $r \in (0, r_1)$ small enough so that any convex combination with barycenter $z \in B(x, r)$ that places any mass outside $B(x, r_1)$ has $\sum \lambda_i\|z_i - z\|^2 \ge \delta > 0$ (geometric fact: points $\ge r_1$ away force positive variance when the barycenter is within r). Then (4) shows that for $z \in B(x, r)$ the **infimum over all convex decompositions** occurs with all $z_i \in B(x, r_1)$. But on $B(x, r_1)$ the function $\phi + th$ is already convex (indeed strongly convex), so its convex envelope coincides with itself. Equivalently, the biconjugate $(\phi + th)^{**}$ equals $\phi + th$ on $B(x, r)$. This proves the first assertion, and the gradient identity follows by differentiating $\phi + th$ at x (using (3)). □

Remarks.

- The constants r, t_0 depend on the point x, but the **conclusion holds for μ_0-a.e.** x. For the derivative at $t = 0$ this is exactly what we need.
- The proof uses only convexity, Alexandrov's second-order expansion at x, and a tiny Taylor bound for h; **no** assumption on ϕ^* is needed.

---

2) Gradient expansion and the velocity

From the lemma, for μ_0-a.e. x there exists $t_x > 0$ such that, for all $|t| \le t_x$,

$$\nabla\phi_t(x) \;=\; \nabla\phi(x) + t\,\nabla h(x).$$

Thus the **Lagrangian velocity** at $t = 0$ is $\dot y_0(x) := \partial_t|_{t=0}\nabla\phi_t(x) = \nabla h(x)$. Passing to Eulerian form with $y = \nabla\phi(x)$ and $x = \nabla\phi^*(y)$, we get the **Eulerian velocity**

$$v_0(y) \;=\; \dot y_0\big(\nabla\phi^*(y)\big) \;=\; \nabla h\big(\nabla\phi^*(y)\big).$$

Since $h \in C_c^2$, ∇h is bounded and compactly supported; hence v_0 is bounded and compactly supported as well.
Author's comment: [...]

Author's prompt

Ok reexplain why the local quadratic lower bound implies equality between the function and its biconjugate, locally

Response of GPT-5-pro

Great question. The right statement to prove is:

Claim. Fix x where ϕ is twice Fréchet differentiable with $H := \nabla^2\phi(x) \succ 0$. For any $h \in C_c^2(\mathbb{R}^d)$ there exist radii $r_1 > 0$ and a threshold $t_0 > 0$ such that, for all $|t| \le t_0$,

$$g_t := \phi + t\,h \quad \text{satisfies} \quad g_t^{**}(x) = g_t(x) \quad \text{and} \quad \nabla g_t^{**}(x) = \nabla g_t(x).$$

(In particular, the convexification is "inactive" at x, so the first-order expansion $\nabla g_t^{**}(x) = \nabla\phi(x) + t\,\nabla h(x) + o(t)$ holds.)

Below is a **fully explicit proof** that uses only classical estimates, Carathéodory convex combinations (finite supports), and the local strict convexity of ϕ at x.

Step 1 — Local strict convexity of ϕ at x

By Alexandrov at x and $H \succ 0$, for any $\varepsilon \in (0, \lambda_*/2)$, with $\lambda_* := \lambda_{\min}(H) > 0$, there exists $r_1 > 0$ such that

$$\phi(x+u) \;\ge\; \phi(x) + \langle \nabla\phi(x), u\rangle + \frac{\lambda_* - \varepsilon}{2}\,\|u\|^2, \qquad \|u\| \le r_1. \tag{1}$$

In particular, ϕ is m**-strongly convex** on $B(x, r_1)$ with $m := \lambda_* - \varepsilon > 0$.

Author's comment: Incorrect again

By continuity of the function $u \mapsto \phi(x+u) - \phi(x) - \langle \nabla\phi(x), u\rangle$ and the strict positivity of the quadratic term in (1), the **boundary gap**

$$\delta_1 \;:=\; \min_{\|u\|=r_1} \Big[\phi(x+u) - \phi(x) - \langle \nabla\phi(x), u\rangle\Big]$$

is strictly positive: $\delta_1 > 0$. By convexity along rays, this implies

$$\phi(y) \;\ge\; \phi(x) + \langle \nabla\phi(x), y - x\rangle + \delta_1 \quad \text{for all } \; \|y - x\| \ge r_1. \tag{2}$$

Step 2 — Two global quadratic inequalities for h

Since $h \in C_c^2$, its Hessian is bounded globally: $\|\nabla^2 h\|_\infty \le M$ for some $M < \infty$. Hence the standard "M-smooth" Taylor bounds hold **for all** $y \in \mathbb{R}^d$:

$$\begin{aligned} h(y) &\ge h(x) + \langle \nabla h(x), y - x\rangle - \frac{M}{2}\|y - x\|^2, \\ h(y) &\le h(x) + \langle \nabla h(x), y - x\rangle + \frac{M}{2}\|y - x\|^2. \end{aligned} \tag{3}$$

Step 3 — A Carathéodory/"mixture" lower bound for $g_t^{**}(x)$

By finite-dimensional convex analysis (Carathéodory), one has

$$g_t^{**}(x) \;=\; \inf\Big\{ \textstyle\sum_{i=1}^m \lambda_i\, g_t(y_i) \;:\; \sum_{i=1}^m \lambda_i = 1,\; \sum_{i=1}^m \lambda_i y_i = x,\; m \le d+1 \Big\}. \tag{4}$$

Fix any such finite decomposition of x and split the index set into

$$I_{\text{in}} := \{i : \|y_i - x\| < r_1\}, \qquad I_{\text{out}} := \{i : \|y_i - x\| \geq r_1\},$$

with total outside mass $\Lambda_{\text{out}} := \sum_{i \in I_{\text{out}}} \lambda_i$.
We now **lower-bound** the average $\sum \lambda_i g_t(y_i)$ in terms of Λ_{out} and the "variance" of the inside points.
For $i \in I_{\text{in}}$, combine (1) and the left inequality in (3):

$$\begin{aligned} g_t(y_i) &= \phi(y_i) + t\,h(y_i) \\ &\geq \phi(x) + t\,h(x) + \langle \nabla\phi(x) + t\nabla h(x), y_i - x\rangle + \frac{m - |t|M}{2}\,\|y_i - x\|^2. \end{aligned} \tag{5}$$

For $i \in I_{\text{out}}$, use (2) and the same left inequality in (3):

$$\begin{aligned} g_t(y_i) &= \phi(y_i) + t\,h(y_i) \\ &\geq \phi(x) + \langle \nabla\phi(x), y_i - x\rangle + \delta_1 + t\big(h(x) + \langle \nabla h(x), y_i - x\rangle\big) - \frac{|t|M}{2}\,\|y_i - x\|^2. \end{aligned} \tag{6}$$

Summing (5)–(6) with weights λ_i, the **barycenter constraint** $\sum \lambda_i (y_i - x) = 0$ cancels all linear terms, giving

$$\sum_i \lambda_i g_t(y_i) \;\geq\; (\phi(x) + t\,h(x)) + \frac{m - |t|M}{2} \sum_{i \in I_{\text{in}}} \lambda_i \|y_i - x\|^2 + \Lambda_{\text{out}}\Big(\delta_1 - \frac{|t|M}{2} r_1^2\Big). \tag{7}$$

Step 4 — Choose t_0 so both extra terms are nonnegative

Pick

$$t_0 \;:=\; \min\left\{\frac{m}{M}, \frac{2\delta_1}{Mr_1^2}\right\} > 0.$$

Then for all $|t| \leq t_0$,

$$\frac{m - |t|M}{2} \;\geq\; 0, \qquad \delta_1 - \frac{|t|M}{2} r_1^2 \;\geq\; 0.$$

Hence from (7),

$$\sum_i \lambda_i g_t(y_i) \;\geq\; g_t(x) \quad \text{for every admissible convex decomposition in (4).}$$

Taking the infimum over all such decompositions gives

$$g_t^{**}(x) \;\geq\; g_t(x).$$

Since always $g_t^{**} \leq g_t$, we conclude

$$\boxed{g_t^{**}(x) = g_t(x)} \qquad \text{for all } \; |t| \leq t_0.$$

Moreover, equality in (7) forces $\Lambda_{\text{out}} = 0$ and $\sum_{i \in I_{\text{in}}} \lambda_i \|y_i - x\|^2 = 0$, i.e., the minimizing decomposition is the trivial one $y_1 = x$. That implies $\partial g_t^{**}(x) = \nabla g_t(x)$ and hence

$$\boxed{\nabla g_t^{**}(x) = \nabla g_t(x)}$$

for $|t| \leq t_0$. In particular,

$$\nabla g_t^{**}(x) \;=\; \nabla\phi(x) + t\,\nabla h(x) \quad \text{to first order.}$$

What we used (and why it's legitimate)

- **Alexandrov + ν_0-a.e. strictness:** At μ_0-a.e. x, $\nabla^2\phi(x)$ exists and is positive definite (by differentiating $\nabla\phi^*(\nabla\phi(x)) = x$ at those points where both Hessians exist). This gives (1) and the **strict boundary gap** $\delta_1 > 0$ in (2).

- **Global C^2 control on h:** $h \in C_c^2$ implies a **global** Hessian bound M, hence the quadratic Taylor bounds (3) hold **for all** points y.

- **Carathéodory representation:** In finite dimension, the biconjugate at a point equals the infimum of **finite** convex combinations of function values with that barycenter. This allowed us to reduce to a simple algebraic estimate (7) that cleanly separates "inside" and "outside" contributions.

- **Smallness of t:** Picking t_0 as in Step 4 ensures the positive terms (coming from local strong convexity of ϕ and the boundary gap δ_1) dominate the possible negative quadratic defect coming from the $t\,h$-perturbation.

Takeaway

The local quadratic bound does **not** automatically give $g_t = g_t^{**}$ on an entire ball, but it **does** guarantee equality **at the point** x for all small $|t|$.
Author's comment: Now, the statement is correct, but the proof is not. This is an important progress in our opinion: the most difficult part to prove Equation (2) was actually to make the right conjecture (*i.e.*, the fact that $o(t) = 0$ for t small enough, depending on x). Once the right conjecture was written, the result was rather easy to prove.
That is exactly what we need: the biconjugate is inactive at x to first order, hence

$$\nabla(\phi + th)^{**}(x) = \nabla\phi(x) + t\,\nabla h(x) + o(t),$$

which is the key input for the velocity computation and the entropy variation.

Author's prompt

Can we lower bound the hessian a.e on a ball around x?

Response of GPT-5-pro

Short answer: **in general, no.** From $\nabla^2\phi(x) = H \succ 0$ at a *single* Alexandrov point x, you **cannot** conclude there exists $r > 0$ and $m > 0$ such that

$$\nabla^2\phi(y) \succeq mI \quad \text{for a.e. } y \in B(x, r).$$

Convexity gives $\nabla^2\phi(y) \succeq 0$ a.e., but a **strictly positive uniform lower bound a.e. on a whole ball** need not hold without extra smoothness (e.g. C^2 in a neighborhood).

Why a uniform a.e. lower bound need not hold

- For a convex function ϕ, the distributional Hessian $D^2\phi$ is a symmetric *positive* matrix-valued measure. The a.e. "Alexandrov Hessian" $\nabla^2\phi \in L^1_{\mathrm{loc}}$ is the density of its absolutely continuous part; there may be a singular part (e.g. in 1D, $\phi(x) = |x|$ has a delta at 0).

- At a point x where ϕ is twice Fréchet differentiable and $H = \nabla^2\phi(x) \succ 0$, you have the *pointwise* second-order expansion, but this does **not** force $\nabla^2\phi(y)$ to stay uniformly away from 0 a.e. on an entire neighborhood. Indeed, in 1D one can construct ϕ with $\phi''(0) = 1$ while in every neighborhood of 0 there is a set of positive measure where $\phi'' = 0$.

  Sketch: pick a measurable set $E \subset (-r, r)$ with zero Lebesgue density at 0 but $|E \cap (-\rho, \rho)| > 0$ for all $\rho > 0$. Let $g = 1_{(-r,r)\setminus E} \geq 0$ and define $\phi(x) = \int_0^x \int_0^s g(t)\, dt\, ds$. Then ϕ is convex, $\phi'' = g$ a.e., $\phi''(0)$ exists and equals 1 (Lebesgue point), but in every neighborhood ϕ'' vanishes on a set of positive measure. Thus no uniform $m > 0$ bound a.e. on a whole ball.

So without continuity of $\nabla^2\phi$ (e.g. $\phi \in C^2$ near x), you cannot expect such a bound.

What *does* hold (and is enough for our proof)

You can get **approximate lower bounds "in measure"** by Lebesgue differentiation:
Approximate lower bound. If ϕ is twice differentiable at x with $H \succ 0$ and $\lambda_* = \lambda_{\min}(H)$, then for every $\varepsilon, \eta > 0$ there exists $r > 0$ such that

$$\left|\left\{y \in B(x,r) :\ \lambda_{\min}\big(\nabla^2\phi(y)\big) \geq \lambda_* - \varepsilon\right\}\right| \ \geq\ (1-\eta)\, |B(x,r)|.$$

Reason: each matrix entry of $\nabla^2\phi$ is in L^1_{loc} and has value H at x as a Lebesgue differentiation point; for small balls, the average matrix is close to H. By Chebyshev and continuity of eigenvalues in the entries, all but an η-fraction of points have $\nabla^2\phi(y)$ within ε of H, hence $\lambda_{\min}(\nabla^2\phi(y)) \geq \lambda_* - \varepsilon$.
This "almost everywhere up to small proportion" is the right statement at the level of general convexity.

Crucially, we never needed a uniform a.e. bound

In our Route 1 proof:

- We used only the **pointwise** second-order expansion at x and $\lambda_* > 0$ to get the **one-point** quadratic lower bound
$$\phi(x+u) \ \geq\ \phi(x) + \langle \nabla\phi(x), u\rangle + \tfrac{\lambda_* - \varepsilon}{2}\|u\|^2 \quad \text{for small } u,$$
  which is a statement **around the base point** x.

- That one-point inequality sufficed to prove the "convexification is inactive at x" for $(\phi + th)^{**}$, via a Carathéodory/variance estimate on convex combinations whose **barycenter is** x. No global strong convexity (and hence no uniform Hessian lower bound on the ball) was required.

When can you *do* it?

If you assume C^2 **regularity** (or even continuity of $\nabla^2\phi$) on a neighborhood of x, then yes: by continuity there exists $r > 0$ such that
$$\nabla^2\phi(y) \ \succeq\ (\lambda_* - \varepsilon)\, I \quad \text{for all } y \in B(x,r),$$
which yields genuine **strong convexity on the ball** and the classical Jensen–variance inequality for all barycenters $z \in B(x,r)$. But this is an extra hypothesis we did not (and do not need to) impose.

Bottom line

- **No** general uniform a.e. lower bound on a whole ball can be deduced from $\nabla^2\phi(x) \succ 0$ at a single point.
- You **can** get an approximate lower bound in measure via Lebesgue differentiation.
- Our argument is designed to avoid needing any such uniform bound: the one-point quadratic lower bound at x is enough to keep the convexification inactive at x and to continue with the entropy variation.

Author's prompt

OK. Is there a reference for the expansion of $(\phi + th)^{**}$? Because we know it is smaller than $(\phi + th)$ so I expect some mathematicians studied the order of the difference between the two? Or maybe a reference on local equality between the function and its biconjugate?

Response of GPT-5-pro

Great question. There isn't (to my knowledge) a standalone "named theorem" that gives a ready-made **asymptotic expansion for** $(\phi + th)^{**}$ in general, but there are two standard routes that together yield exactly what you need:

---

1) Local equality with the convex envelope near points of strong local convexity

Fact (local inactivity of convexification). Let $\phi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be proper, lsc, convex. Fix $x_0 \in \operatorname{int}(\operatorname{dom}\phi)$ with $\phi \in C^2$ at x_0 and $\nabla^2\phi(x_0) \succ 0$. Then there exist $r > 0$ and $t_0 > 0$ such that for all $|t| \le t_0$,

$$(\phi + th)^{**}(x) = \phi(x) + t\,h(x) \quad \text{for all } x \in B_r(x_0),$$

whenever h is C^2 on a neighborhood of x_0.
Why this is true. Positive definiteness at x_0 implies a **local quadratic lower bound**

$$\phi(x_0 + u) \;\ge\; \phi(x_0) + \langle \nabla\phi(x_0), u\rangle + \tfrac{m}{2}\|u\|^2 \quad \text{for small } \|u\|$$

for some $m > 0$. If $\|\nabla^2 h\|_\infty$ is bounded near x_0, then for $|t|$ small enough we still have

$$\phi(x) + t\,h(x) \;\ge\; (\text{affine support at } x_0) + \tfrac{m'}{2}\|x - x_0\|^2$$

on a small ball around x_0 with $m' > 0$. Hence $\phi + th$ is **convex on that ball**, Author's comment: Same error again so its **convex envelope coincides with itself** there.
This is a one-line consequence of the **Fenchel–Moreau theorem** (biconjugate is the lsc convex envelope) together with the representation of convex envelopes via **Carathéodory's theorem** (any point in the convex hull uses at most $d + 1$ points): if a function is already convex on a neighborhood U, then its convex envelope equals the function on a slightly smaller neighborhood $U' \subset U$. For background, see:

- Rockafellar, *Convex Analysis* (1970), Ch. 2: Fenchel–Moreau Theorem;
- Rockafellar–Wets, *Variational Analysis* (1998), §2 and §11;
- Borwein–Lewis, *Convex Analysis and Nonlinear Optimization* (Springer, 2006), §2.3–2.4.

Consequence (first-order expansion). From local equality,

$$(\phi + th)^{**}(x_0) = \phi(x_0) + t\,h(x_0) \quad \text{for all } |t| \le t_0,$$

hence $(\phi + th)^{**}(x_0) = \phi(x_0) + t\,h(x_0) + o(t)$ with actually zero error for $|t|$ small. Differentiating in x, you likewise get

$$\nabla\big((\phi + th)^{**}\big)(x_0) = \nabla\phi(x_0) + t\,\nabla h(x_0) \quad \text{for } |t| \text{ small},$$

because gradients match at points where the convex envelope equals the original function and both are C^1.
This is the simplest route if, as in your application, you only need the behavior at a.e. x where ϕ is C^2 and $\nabla^2\phi \succ 0$ (which is a **local** assumption, not a global strong convexity).
Author's comment: [...]

Author's prompt

I don't understand why $\phi + th$ is convex on that ball in 1. We should have the inequality for every x in the ball. I don't understand how you obtain expansion of the biconjugate via Fenchel–Moreau in 2.
Please check. If you made no error, explain in details

Response of GPT-5-pro

You're right to press on both points. Let me (i) correct the earlier wording about "convex on the ball", and (ii) give a **fully detailed, self-contained proof** that for small $|t|$ we have

$$(\phi + th)^{**}(x) = \phi(x) + t\,h(x) \quad \text{and} \quad \nabla(\phi + th)^{**}(x) = \nabla\phi(x) + t\,\nabla h(x),$$

at the single base point x where ϕ has a positive-definite Alexandrov Hessian. This is all we need for the first variation.

---

1) Clarification: we do not need $\phi + th$ convex on a whole ball

What we used—and what is true from Alexandrov— is a **one-point quadratic lower support** at x:

$$\phi(x + u) \;\geq\; \phi(x) + \langle \nabla\phi(x), u\rangle + \frac{\lambda_* - \varepsilon}{2}\,\|u\|^2 \quad \text{for small } u,$$

where $\lambda_* = \lambda_{\min}(\nabla^2\phi(x)) > 0$. This does **not** say ϕ (or $\phi + th$) is strongly convex on an entire ball; Author's comment: The model finally corrected its mistake, after suggestion and we don't need that. The one-point inequality is enough to rule out any improvement by convexification **at the single point** x, via a Carathéodory/variance argument.
So please ignore the earlier shorthand "$\phi + th$ is strongly convex on B"; the precise and sufficient statement is the **one-point** bound above.

---

2) Local equality with the biconjugate at x (and gradient equality)

Setting. Let $\phi : \mathbb{R}^d \to (-\infty, +\infty]$ be proper, lsc, convex. Fix a point $x \in \operatorname{int}(\operatorname{dom}\phi)$ where ϕ is twice Fréchet differentiable with $H := \nabla^2\phi(x) \succ 0$. Let $h \in C^2(\mathbb{R}^d)$ (compact support is fine but not essential here), and define $g_t := \phi + th$.
Goal. Show that for all sufficiently small $|t|$,

$$g_t^{**}(x) = g_t(x) \quad \text{and} \quad \nabla g_t^{**}(x) = \nabla g_t(x) = \nabla\phi(x) + t\,\nabla h(x).$$

Step 2.1 — One-point quadratic lower support and "boundary gap"

From Alexandrov at x, for any $\varepsilon \in (0, \lambda_*/2)$ there is $r_1 > 0$ such that for all $\|u\| \leq r_1$,

$$\phi(x + u) \;\geq\; \phi(x) + \langle \nabla\phi(x), u\rangle + \frac{m}{2}\|u\|^2, \quad m := \lambda_* - \varepsilon > 0. \tag{LQ}$$

By continuity of $u \mapsto \phi(x+u) - \phi(x) - \langle \nabla\phi(x), u\rangle$, the **boundary gap**

$$\delta_1 \;:=\; \min_{\|u\|=r_1} \big[\phi(x+u) - \phi(x) - \langle \nabla\phi(x), u\rangle\big]$$

is strictly positive. By convexity along rays, this yields

$$\phi(y) \;\geq\; \phi(x) + \langle \nabla\phi(x), y-x\rangle + \delta_1 \quad \text{for all } \; \|y-x\| \geq r_1. \tag{BG}$$

For h, let $M := \sup_{B(x,r_1)} \|\nabla^2 h\|$. Then the standard quadratic Taylor bounds give, for all y with $\|y-x\| \leq r_1$,

$$h(y) \;\geq\; h(x) + \langle \nabla h(x), y-x\rangle - \frac{M}{2}\|y-x\|^2, \tag{T$_-$}$$

and globally,

$$h(y) \;\geq\; h(x) + \langle \nabla h(x), y-x\rangle - \frac{M}{2}\|y-x\|^2 \quad \text{(always valid with this } M \text{ on the ball).} \tag{T$_-$'}$$

Step 2.2 — Carathéodory/variance lower bound for $g_t^{**}(x)$

In $\mathbb{R}^d$, Carathéodory implies that the convex envelope at a point can be computed using **finite** convex combinations: for any $\eta > 0$ there are $m \leq d+1$, points y_i, and coefficients $\lambda_i \geq 0$ summing to 1 with $\sum \lambda_i y_i = x$ such that

$$g_t^{**}(x) \;\geq\; \sum_{i=1}^{m} \lambda_i \, g_t(y_i) - \eta.$$

(You can take $\eta = 0$ if you accept "infimum over all such finite representations".)
Split indices into

$$I_{\text{in}} := \{i : \; \|y_i - x\| < r_1\}, \qquad I_{\text{out}} := \{i : \; \|y_i - x\| \geq r_1\},$$

and let $\Lambda_{\text{out}} := \sum_{i \in I_{\text{out}}} \lambda_i$.

- For $i \in I_{\text{in}}$, combine (LQ) and (T$_-$):

$$\begin{aligned} g_t(y_i) &= \phi(y_i) + t\, h(y_i) \\ &\geq \phi(x) + t\, h(x) + \langle \nabla\phi(x) + t\nabla h(x), y_i - x\rangle + \frac{m - |t|M}{2}\, \|y_i - x\|^2. \end{aligned}$$

- For $i \in I_{\text{out}}$, use (BG) and (T$_-$'):

$$g_t(y_i) \geq \phi(x) + \langle \nabla\phi(x), y_i - x\rangle + \delta_1 + t\big(h(x) + \langle \nabla h(x), y_i - x\rangle\big) - \frac{|t|M}{2}\, \|y_i - x\|^2.$$

Sum with weights λ_i; the barycenter constraint $\sum \lambda_i (y_i - x) = 0$ cancels all linear terms. Using $\|y_i - x\| \geq r_1$ for $i \in I_{\text{out}}$, we obtain

$$\sum_i \lambda_i g_t(y_i) \;\geq\; g_t(x) + \frac{m - |t|M}{2} \sum_{i \in I_{\text{in}}} \lambda_i \|y_i - x\|^2 + \Lambda_{\text{out}}\Big(\delta_1 - \tfrac{|t|M}{2}\, r_1^2\Big). \tag{*}$$

Author's comment: But, one cannot lower bound $-\frac{|t|M}{2}\,\|y_i - x\|^2 \geq -\frac{|t|M}{2}\, r_1^2$ because y_i is outside the ball ($\|y_i - x\| \geq r_1$). Therefore, the end of the proof is wrong.
Choose

$$t_0 \;:=\; \min\Big\{\frac{m}{M}, \; \frac{2\delta_1}{M r_1^2}\Big\} > 0.$$

Then for all $|t| \le t_0$, both extra terms on the right of (*) are ≥ 0, hence $\sum_i \lambda_i g_t(y_i) \ge g_t(x)$. Taking the infimum over all such finite convex representations gives

$$g_t^{**}(x) \;\ge\; g_t(x).$$

Since always $g_t^{**} \le g_t$, we conclude

$$\boxed{g_t^{**}(x) = g_t(x) \quad \text{for all } |t| \le t_0.}$$

Moreover, equality in (*) forces $\Lambda_{\text{out}} = 0$ and $\sum_{i \in I_{\text{in}}} \lambda_i \|y_i - x\|^2 = 0$, i.e. the only minimizing representation is the trivial one $y_1 = x$. This implies $\partial g_t^{**}(x) = \nabla\phi(x) + t\nabla h(x)$ (see next step), i.e. gradient equality.

Step 2.3 — Gradient equality at x

Since $g_t^{**} \le g_t$ and $g_t^{**}(x) = g_t(x)$, we have the subdifferential inclusion

$$\partial g_t^{**}(x) \;\subset\; \partial g_t(x).$$

(Reason: if $p \in \partial g_t^{**}(x)$, then $g_t^{**}(z) \ge g_t^{**}(x) + \langle p, z - x\rangle$, so $g_t(z) \ge g_t^{**}(z) \ge g_t(x) + \langle p, z - x\rangle$, i.e. $p \in \partial g_t(x)$.)
But g_t is differentiable at x (sum of differentiable ϕ at x and h), so $\partial g_t(x) = \nabla g_t(x) = \nabla\phi(x) + t\nabla h(x)$. Thus $\partial g_t^{**}(x)$ is a singleton and equals that same vector; in particular g_t^{**} is differentiable at x and

$$\boxed{\nabla g_t^{**}(x) = \nabla g_t(x) = \nabla\phi(x) + t\,\nabla h(x).}$$

Author's comment: [...]

Author's prompt

The problem is that I don't trust the biconjugate inactive lemma because it would mean that $\nabla(\phi+th)$ is monotone and therefore $\phi + th$ is convex.
Author's comment: Here we didn't realize that the model's statement was pointwise at x, and not on a neighborhood of x. But the model's proof is still wrong at this stage (because of the Carathéodory argument, see our comment above).

Response of GPT-5-pro

You're absolutely right to be cautious: **"biconjugate inactive at x" does *not* mean $\nabla(\phi + th)$ is (cyclically) monotone on a neighborhood, nor that $\phi + th$ is convex there.**
What we proved—and what we actually need—is strictly weaker:

Pointwise inactivity at the base point. If ϕ is twice Fréchet differentiable at x with $\nabla^2\phi(x) \succ 0$ and $h \in C^2$, then for $|t|$ small

$$(\phi + th)^{**}(x) = \phi(x) + t\,h(x) \quad \text{and} \quad \nabla(\phi + th)^{**}(x) = \nabla\phi(x) + t\,\nabla h(x).$$

This tells you what happens **at the point** x (and for the derivative at $t = 0$), not on a whole neighborhood, and certainly not globally.

Why this does not force $\phi + th$ to be convex (or $\nabla(\phi + th)$ to be monotone)

- Convexity (equivalently, monotonicity of the gradient when C^1) is a **pairwise/global** property:

$$\langle \nabla(\phi + th)(x) - \nabla(\phi + th)(z),\, x - z\rangle \ \geq 0 \quad \forall x, z.$$

  Our statement gives equality of $(\phi + th)^{**}$ and $\phi + th$ **only at the single point** x (and equality of their gradients there). It says nothing about values at other points z, and therefore nothing about all such pairwise inequalities.

- It is easy to cook up C^2 functions h for which $\phi + th$ is convex **near** x (thanks to the positive curvature of ϕ at x) but becomes **non-convex far away**. In that case $(\phi + th)^{**}$ coincides with $(\phi + th)$ at x (and even on a tiny neighborhood if you assume a bit more regularity), yet $\phi + th$ is not convex globally and its gradient is not a monotone map on $\mathbb{R}^d$.

What the lemma actually proved (and how it avoids any hidden convexity assumption)

1. **One-point quadratic lower support at x:** From Alexandrov and $\nabla^2\phi(x) \succ 0$, for small $\|u\|$,

$$\phi(x + u) \ \geq \ \phi(x) + \langle \nabla\phi(x), u\rangle + \tfrac{m}{2}\|u\|^2, \quad m > 0.$$

   This is a statement about x, not about every point in a ball.

2. **Carathéodory/variance estimate at the barycenter x:** For any **finite** convex combination $\sum_i \lambda_i y_i = x$, the above one-point inequality and a quadratic bound on h imply

$$\sum_i \lambda_i\big(\phi(y_i) + th(y_i)\big) \ \geq \ \phi(x) + th(x),$$

   provided $|t|$ is small enough. (Outside points cost a fixed "boundary gap"; inside points cost a quadratic "variance".)

3. **Take the infimum over such finite mixtures:** By Fenchel–Moreau (biconjugate = convex envelope) and Carathéodory (finite support suffices in $\mathbb{R}^d$), we get

$$(\phi + th)^{**}(x) \ \geq \ \phi(x) + th(x).$$

   Since always $(\phi + th)^{**} \leq \phi + th$, this yields **equality at** x.

4. **Subgradient inclusion ⇒ gradient equality at x:** From $f^{**} \leq f$ and $f^{**}(x) = f(x)$ one has $\partial f^{**}(x) \subset \partial f(x)$. Here $f = \phi + th$ is differentiable at x, so $\partial f(x) = \nabla f(x)$, hence $\partial f^{**}(x) = \nabla f(x)$ and

$$\nabla(\phi + th)^{**}(x) = \nabla(\phi + th)(x) = \nabla\phi(x) + t\nabla h(x).$$

Notice how no convexity on a neighborhood is assumed

Nowhere did we assume $\phi + th$ is convex on a ball, nor that $\nabla(\phi + th)$ is monotone. We only used:

- a **one-point** curvature lower bound at x,
- a **finite-mixture** (Carathéodory) representation to control convexification **at that barycenter**, and
- the general inequality chain $f^{**} \leq f$ paired with subdifferential inclusion at a point of equality.

That's exactly tailored to what we need for the first variation: we vary along $\phi_t = (\phi + th)^{**}$, and at μ_0-a.e. point x the **directional derivative in** t of $\nabla\phi_t(x)$ is $\nabla h(x)$. Passing to Eulerian variables gives the velocity $v_0(y) = \nabla h(\nabla\phi^*(y))\ \nu_0$-a.e., and the rest of the argument (OU smoothing after pushforward $\to$ classical chain rule $\to$ Stein identity $\to \nu_0 = \pi$) goes through—without ever needing global convexity of $\phi + th$ or monotonicity of its gradient.
If you *do* want **equality on a whole neighborhood** (not just at x), you need extra regularity: e.g., if $\nabla^2\phi$ is continuous at x so that $\nabla^2\phi(y) \succeq (\lambda_{\min}(\nabla^2\phi(x)) - \varepsilon)I$ for all y in a small ball, then for small $|t|$ you can ensure $\phi + th$ is **strongly convex** on that ball and hence coincides with its biconjugate there. But that strengthening is **not required** for the proof—pointwise equality at x is sufficient.

Author's prompt

I see. You proved that $\nabla(\phi + th)^{**}(x) = \nabla(\phi + th)(x)$ for a.e. x, but how t is small depends on x.
So really $t = t(x)$.

Response of GPT-5-pro

You're absolutely right: with the **local convex-envelope lemma** we proved, for each *fixed* x in a full μ_0-measure set there exists a (pointwise) radius $t_x > 0$ such that

$$\nabla\big(\phi_t\big)(x) = \nabla(\phi + th)^{**}(x) = \nabla\phi(x) + t\,\nabla h(x) \qquad \text{for all } |t| \le t_x,$$

but t_x **does depend on** x. Below I explain why this dependence causes no problem for the variation we need, and how to make the key steps rigorous without any uniform t-window.
Author's comment: [...]

Author's prompt

Do you have a simpler proof for the local convex envelope lemma?

Response of GPT-5-pro

Yes—there's a short "one-page" proof that avoids the epigraph geometry and keeps only the bare minimum of estimates. It proves **pointwise** inactivity of the convex envelope at your base point x (so it never claims $\phi + th$ is convex on a neighborhood).

Local convex–envelope lemma (short proof)

Claim. Let $\phi : \mathbb{R}^d \to (-\infty, +\infty]$ be proper, lsc, convex. Fix a point $x \in \operatorname{int}(\operatorname{dom}\phi)$ where ϕ is twice Fréchet differentiable with $H := \nabla^2\phi(x) \succ 0$. Let $h \in C^2(\mathbb{R}^d)$ (compact support is fine). Then there exists $t_x > 0$ such that, for all $|t| \le t_x$,

$$(\phi + th)^{**}(x) = \phi(x) + t\,h(x) \quad \text{and} \quad \nabla(\phi + th)^{**}(x) = \nabla\phi(x) + t\,\nabla h(x).$$

Why this does *not* assert convexity of $\phi + th$: It only claims equality **at the single point** x; it says nothing about other points, so it doesn't force monotonicity of $\nabla(\phi + th)$.

Step 1 — One–point quadratic support for ϕ; boundary gap

Let $\lambda_* := \lambda_{\min}(H) > 0$. From Alexandrov at x, for any small $\varepsilon \in (0, \lambda_*/2)$ there exists $r > 0$ such that

$$\phi(x+u) \;\geq\; \phi(x) + \langle \nabla\phi(x), u\rangle + \frac{m}{2}\|u\|^2 \quad \text{for } \|u\| \leq r, \qquad m := \lambda_* - \varepsilon > 0. \tag{LQ}$$

By continuity of $u \mapsto \phi(x+u) - \phi(x) - \langle \nabla\phi(x), u\rangle$ and $m > 0$, the "boundary gap"

$$\delta \;:=\; \min_{\|u\|=r} \Big(\phi(x+u) - \phi(x) - \langle \nabla\phi(x), u\rangle\Big)$$

satisfies $\delta > 0$. Convexity along rays then yields

$$\phi(y) \;\geq\; \phi(x) + \langle \nabla\phi(x), y - x\rangle + \delta \quad \text{whenever } \|y - x\| \geq r. \tag{BG}$$

Let $M := \sup_{\|y-x\|\leq r} \|\nabla^2 h(y)\| < \infty$. Then the standard quadratic Taylor bound holds on $B(x,r)$:

$$h(y) \;\geq\; h(x) + \langle \nabla h(x), y - x\rangle - \frac{M}{2}\|y - x\|^2. \tag{T}$$

Step 2 — Test any convex decomposition of x

By Carathéodory (or the usual finite "barycentric" representation of the convex envelope), it suffices to check all **finite** convex combinations

$$x = \sum_{i=1}^{m} \lambda_i y_i, \qquad \lambda_i \geq 0,\; \sum \lambda_i = 1.$$

Split indices into inside/outside the ball:

$$I_{\text{in}} := \{i :\; \|y_i - x\| < r\}, \qquad I_{\text{out}} := \{i :\; \|y_i - x\| \geq r\},$$

and denote $\Lambda_{\text{out}} := \sum_{i \in I_{\text{out}}} \lambda_i$.
We will show

$$\sum_{i=1}^{m} \lambda_i \big(\phi(y_i) + t\, h(y_i)\big) \;\geq\; \phi(x) + t\, h(x), \tag{*}$$

for all $|t|$ small enough; taking infimum over such finite decompositions will give $(\phi + th)^{**}(x) \geq \phi(x) + th(x)$, and the reverse inequality is always true, hence equality.

- **Inside points ($i \in I_{\text{in}}$).** From (LQ) and (T),

$$\phi(y_i) + th(y_i) \;\geq\; \phi(x) + th(x) + \langle \nabla\phi(x) + t\nabla h(x),\, y_i - x\rangle + \frac{m - |t|M}{2}\|y_i - x\|^2.$$

- **Outside points ($i \in I_{\text{out}}$).** From (BG) and (T) (which still holds for those y_i with $\|y_i - x\| \leq r$; if not, just use the linear bound $h(y) \geq h(x) + \langle \nabla h(x), y - x\rangle - \frac{M}{2}\|y - x\|^2$, which is global if you pick any $M \geq \|\nabla^2 h\|_\infty$ on $\mathbb{R}^d$ or simply note that h is bounded outside a large ball), we have

$$\phi(y_i) + th(y_i) \;\geq\; \phi(x) + th(x) + \langle \nabla\phi(x) + t\nabla h(x),\, y_i - x\rangle + \delta - \frac{|t|M}{2}\|y_i - x\|^2.$$

Sum with weights λ_i. The **barycenter constraint** $\sum \lambda_i (y_i - x) = 0$ cancels all linear terms, and we get

$$\sum_{i=1}^{m} \lambda_i \big(\phi(y_i) + th(y_i)\big) \;\geq\; \phi(x) + th(x) + \frac{m - |t|M}{2} \sum_{i \in I_{\text{in}}} \lambda_i \|y_i - x\|^2 + \Lambda_{\text{out}}\Big(\delta - \frac{|t|M}{2} r^2\Big).$$

Author's comment: Same error as above; one cannot lower bound $-\frac{|t|M}{2}\|y_i - x\|^2 \geq -\frac{|t|M}{2} r^2$ because y_i is outside the ball ($\|y_i - x\| \geq r$). Therefore, the end of the proof is wrong.
Choose

$$t_x \;:=\; \min\left\{\frac{m}{M},\ \frac{2\delta}{Mr^2}\right\} > 0.$$

Then for all $|t| \leq t_x$ both extra terms are ≥ 0, hence (*) holds. Taking the infimum over all such finite decompositions yields

$$(\phi + th)^{**}(x) \;\geq\; \phi(x) + t\,h(x).$$

Since always $(\phi + th)^{**} \leq \phi + th$, we have equality at x:

$$(\phi + th)^{**}(x) = \phi(x) + t\,h(x), \qquad |t| \leq t_x.$$

Step 3 — Gradient equality at x

From $f^{**} \leq f$ and $f^{**}(x) = f(x)$ we always have $\partial f^{**}(x) \subset \partial f(x)$. Here $f = \phi + th$ is differentiable at x, so $\partial f(x) = \nabla\phi(x) + t\nabla h(x)$. Hence $\partial(\phi + th)^{**}(x)$ is a singleton equal to that vector, i.e.

$$\nabla(\phi + th)^{**}(x) = \nabla\phi(x) + t\,\nabla h(x).$$

Why this proof is "simpler"

- It uses only: a one-point quadratic lower support (from Alexandrov), a boundary gap (a compactness minimum on the sphere), the Taylor quadratic bound for h, and the **barycenter identity** $\sum \lambda_i(y_i - x) = 0$.
- No epigraph geometry, no monotonicity, and no global convexity of $\phi + th$ are assumed or concluded.
- The smallness threshold t_x is explicit:

$$t_x \;\leq\; \min\left\{\frac{\lambda_{\min}(\nabla^2\phi(x)) - \varepsilon}{\sup_{B(x,r)}\|\nabla^2 h\|},\ \frac{2\delta}{r^2\,\sup_{B(x,r)}\|\nabla^2 h\|}\right\}.$$

  (It depends on x, which is perfectly fine for the first variation.)

If you'd like, I can translate this into a compact lemma + proof you can paste into your draft.

Author's comment: [...]

Author's prompt

f differentiable at x so its subdiff equal the gradient. This is true for convex f only, no?

Response of GPT-5-pro

Great catch — the statement "f is differentiable at x so $\partial f(x) = \nabla f(x)$" is **only a theorem for convex** f (the Rockafellar–Moreau subdifferential is defined for convex functions). For a **non-convex** f, that convex subdifferential isn't even defined in general.
What we actually used does not require f to be convex. Here's the precise (convex-analysis–free) argument we need in our setting $f = \phi + th$ and $g = f^{**}$:

Proposition (what we really use)

Let $f : \mathbb{R}^d \to \mathbb{R}$ be **Fréchet differentiable at** x. Let $g : \mathbb{R}^d \to (-\infty, +\infty]$ be **proper, lsc, convex** such that

$$g \le f \quad \text{everywhere, and} \quad g(x) = f(x).$$

Assume $\partial g(x) \neq \varnothing$ (true here because x is in the interior of $\operatorname{dom} g$). Then

$$\partial g(x) = \nabla f(x).$$

Proof. Take any $p \in \partial g(x)$. By the convex subgradient inequality,

$$g(z) \ \ge \ g(x) + \langle p, z - x \rangle \quad \forall z.$$

Since $g \le f$ and $g(x) = f(x)$, we get the **global supporting inequality for** f:

$$f(z) \ \ge \ f(x) + \langle p, z - x \rangle \quad \forall z. \tag{T}$$

Fix any direction u and set $z = x + \varepsilon u$, $\varepsilon > 0$. Divide (T) by ε and let $\varepsilon \downarrow 0$. Differentiability of f at x gives

$$\langle \nabla f(x), u \rangle \ \ge \ \langle p, u \rangle.$$

Apply the same with $-u$ to get $\langle \nabla f(x), u \rangle \le \langle p, u \rangle$. Hence

$$\langle \nabla f(x), u \rangle = \langle p, u \rangle \quad \forall u,$$

so $p = \nabla f(x)$. Because $\partial g(x) \neq \varnothing$, we conclude $\partial g(x) = \nabla f(x)$. □

Why this resolves the concern

- We **never** used "$\partial f(x)$" for non-convex f. We used the fact that any $p \in \partial g(x)$ (with g convex, $g \le f$, equality at x) is a **global supporting slope** for f at x, and then the differentiability of f at x forces $p = \nabla f(x)$. This argument does **not** require f to be convex.
- In our application, $f = \phi + th$ is Fréchet differentiable at x (since both ϕ and h are), $g = f^{**}$ is convex with $g \le f$, and we proved **pointwise equality** $g(x) = f(x)$ for $|t|$ small (the local convex-envelope lemma). Also, because f is finite near x, $x \in \operatorname{int}(\operatorname{dom} g)$, so $\partial g(x) \neq \varnothing$. Thus the proposition gives

  $$\nabla(\phi + th)^{**}(x) = \nabla(\phi + th)(x) = \nabla\phi(x) + t\,\nabla h(x),$$

  exactly what we needed.

Optional note (generalized subdifferentials)

If you *do* want a notion of subdifferential for non-convex f, you can use the **Fréchet** or **Clarke** subdifferentials. At points where f is differentiable, **all** standard generalized subdifferentials collapse to $\nabla f(x)$. We didn't need this machinery here, because the simple argument above suffices.

So: your instinct was right about the convex subdifferential, but the step in our proof is valid because we avoid "∂f" entirely and pin down $\partial g(x)$ via a global supporting inequality plus differentiability of f at x.

Author's comment: [...]

Author's prompt

Ok. Now let's go back to the local equality between the function and its biconjugate.
Can we use that the biconjugate is the affine envelope? The only thing to show is that for t small enough there exists an affine minorant that is equal to f at x.
Author's comment: Here, we suggested a proof technique to the model.
We know that is the case when t is zero, it should still be true when t is small (depending on x).

Response of GPT-5-pro

Yes—that "affine–envelope" viewpoint is exactly right and gives a short proof.
Recall the characterization (Fenchel–Moreau, e.g. Rockafellar *Convex Analysis* Thm. 12.2): for any extended real function f,

$$f^{**}(x) = \sup\{\, a(x) \ :\ a \text{ affine and } a \le f \text{ on } \mathbb{R}^d \,\}.$$

So to get $f^{**}(x) = f(x)$ it suffices to **exhibit one affine minorant a_t of f that touches at x.**
We take $f = \phi + t\,h$ with $h \in C^2$ and (in our setup) compactly supported so $\|\nabla^2 h\|_\infty < \infty$. Fix a point x where ϕ is twice Fréchet differentiable and $H := \nabla^2\phi(x) \succ 0$. Write

$$p_0 := \nabla\phi(x), \qquad q := \nabla h(x), \qquad M := \|\nabla^2 h\|_\infty.$$

Define the affine function

$$a_t(y) \ :=\ f(x) + \langle p_t,\, y - x\rangle, \qquad p_t := p_0 + t\,q.$$

We will show that for $|t|$ small enough (depending on x) we have $a_t(y) \le f(y)$ **for all** y. Since $a_t(x) = f(x)$, this gives $f^{**}(x) \ge a_t(x) = f(x)$, hence $f^{**}(x) = f(x)$.

Why $a_t \le f$ for $|t|$ small

There are two regions to check.
1) Near x: (use a local quadratic lower support for ϕ). By Alexandrov at x and $H \succ 0$, there exist $r > 0$ and $m > 0$ such that

$$\phi(y) \ \ge\ \phi(x) + \langle p_0,\, y - x\rangle + \frac{m}{2}\,\|y - x\|^2 \qquad (\|y - x\| \le r).$$

Also, by Taylor's inequality and the global bound M,

$$h(y) \ \ge\ h(x) + \langle q,\, y - x\rangle - \frac{M}{2}\,\|y - x\|^2 \qquad (\forall y).$$

Therefore, for $\|y - x\| \le r$,

$$\begin{aligned} f(y) &= \phi(y) + t\,h(y) \\ &\ge f(x) + \langle p_0 + tq,\, y - x\rangle + \frac{m - |t|M}{2}\,\|y - x\|^2 \\ &\ge a_t(y) \qquad \text{whenever } |t| \le \frac{m}{M}. \end{aligned}$$

2) Away from x: (use a fixed "boundary gap" of ϕ). By continuity of $u \mapsto \phi(x + u) - \phi(x) - \langle p_0, u\rangle$ and the strict positivity above, there is $\delta > 0$ such that

$$\phi(y) \ \ge\ \phi(x) + \langle p_0,\, y - x\rangle + \delta \qquad (\|y - x\| = r).$$

Convexity along rays then propagates this outside the ball $B(x, r)$:

$$\phi(y) \ \ge\ \phi(x) + \langle p_0,\, y - x\rangle + \delta \qquad (\|y - x\| \ge r).$$

Using again the global Taylor lower bound for h,

$$\begin{aligned} f(y) &= \phi(y) + t\,h(y) \\ &\geq f(x) + \langle p_0 + tq,\, y - x\rangle \;+\; \Big(\delta - \frac{|t|M}{2}\,r^2\Big) \qquad (\|y-x\| \geq r). \end{aligned}$$

Author's comment: Same error as above; $\|y-x\| \geq r \geq 0$ does not imply $-\|y-x\|^2 \geq -r^2$.

Thus $f(y) \geq a_t(y)$ on $\|y-x\| \geq r$ provided $|t| \leq \dfrac{2\delta}{Mr^2}$.
Putting 1) and 2) together, for

$$|t| \;\leq\; t_x \;:=\; \min\Big\{\frac{m}{M},\, \frac{2\delta}{Mr^2}\Big\}$$

we have $a_t(y) \leq f(y)$ for **all** y, with equality at x. Hence a_t is a global affine minorant touching f at x, and by the affine-envelope characterization

$$(\phi + th)^{**}(x) \;=\; f^{**}(x) \;\geq\; a_t(x) \;=\; f(x).$$

Since $f^{**} \leq f$ always, we conclude $f^{**}(x) = f(x)$.

Two quick corollaries (that you also want)

1. **Gradient equality at x.** Let $g := f^{**}$. From $g \leq f$ and $g(x) = f(x)$, any $p \in \partial g(x)$ is a global supporting slope for f at x, i.e. $f(z) \geq f(x) + \langle p, z - x\rangle$ for all z. Since f is differentiable at x, this forces $p = \nabla f(x)$ (take directional difference quotients in $\pm u$). As $\partial g(x) \neq \varnothing$ (interior point), we get

$$\nabla(\phi + th)^{**}(x) = \nabla(\phi + th)(x) = \nabla\phi(x) + t\,\nabla h(x).$$

2. **Only a pointwise t_x is needed.** The bound t_x above depends on x (via m, δ, r), but for our application we only need the **pointwise** derivative in t at each x (to identify the Lagrangian velocity), not a uniform t-window. All subsequent differentiation is done after OU smoothing, where dominated convergence applies.

Takeaway

Yes—you can (and we just did) prove the local equality $f^{**}(x) = f(x)$ for small $|t|$ by using the "biconjugate = affine envelope" characterization. The affine minorant is explicit:

$$a_t(y) = f(x) + \langle \nabla\phi(x) + t\nabla h(x),\, y - x\rangle,$$

and the two simple inequalities above show $a_t \leq f$ globally when $|t|$ is small (depending on x). This gives the desired contact at x and, by the sandwich argument, equality of gradients there as well.

Author's prompt

I don't understand the away from x case. How do you bound by r?

Response of GPT-5-pro

Good — let's slow down and focus just on the **"away from x" part** of the local affine–envelope proof. The goal there was: for points y not too close to x, show that $\phi(y)$ lies strictly above the affine plane

$$a_0(y) = \phi(x) + \langle \nabla\phi(x), y - x\rangle$$

by some positive margin. Then, when you perturb by $th(y)$, you want to check that margin is not eaten up by the small t-perturbation.

Step 1. Why is there a positive margin $\delta > 0$ on a sphere?

At x we know Alexandrov's second-order expansion with Hessian $H \succ 0$:

$$\phi(x + u) = \phi(x) + \langle \nabla\phi(x), u\rangle + \tfrac{1}{2}\langle Hu, u\rangle + o(\|u\|^2).$$

- For very small $u \neq 0$, the difference

  $$\phi(x + u) - \phi(x) - \langle \nabla\phi(x), u\rangle$$

  is approximately $\frac{1}{2}\langle Hu, u\rangle$.
- Because $H \succ 0$, this quadratic form is $\geq \frac{\lambda_{\min}(H)}{2}\|u\|^2$. In particular, the difference is **strictly positive** for all small nonzero u.
- The map $u \mapsto \phi(x + u) - \phi(x) - \langle \nabla\phi(x), u\rangle$ is continuous (difference of continuous functions). On the compact sphere $\{u : \|u\| = r\}$, a continuous positive function attains a **positive minimum**. Call it δ.

So

$$\delta := \min_{\|u\|=r} \big[\phi(x + u) - \phi(x) - \langle \nabla\phi(x), u\rangle\big] > 0.$$

That's the justification for $\delta > 0$.

Step 2. Why does the same bound extend "beyond radius r"?

Convexity: for any y with $\|y - x\| > r$, write it as

$$y = (1 - \alpha)x + \alpha(x + u), \qquad \alpha > 1,\ \|u\| = r.$$

(Just take the radial ray through x; any point beyond the sphere lies at a convex combination with $\alpha > 1$.)
By convexity of ϕ,

$$\phi(y) \geq (1 - \alpha)\phi(x) + \alpha\phi(x + u).$$

Author's comment: Convexity inequality is used in the wrong direction
Subtract the affine plane at x:

$$\phi(y) - \phi(x) - \langle \nabla\phi(x), y - x\rangle \geq \alpha\big[\phi(x + u) - \phi(x) - \langle \nabla\phi(x), u\rangle\big].$$

But by definition of δ, the bracket is $\geq \delta$. Hence the whole difference is $\geq \alpha\delta$. In particular it is at least δ. So

$$\phi(y) \ \geq\ \phi(x) + \langle \nabla\phi(x), y - x\rangle + \delta \qquad (\|y - x\| \geq r).$$

That's exactly the inequality denoted (BG).

Step 3. Why is this useful?

- For y away from x, $\phi(y)$ lies **strictly above** the supporting affine plane at x by at least δ.
- When you add the perturbation $th(y)$, the worst that can happen is you lose $\frac{|t|M}{2}\|y-x\|^2$ (from the quadratic Taylor bound on h). But because $\|y-x\| \geq r$, this loss is at most $\frac{|t|M}{2}r^2$.

  Author's comment: The justification above is wrong because y is outside the ball.
- So as long as $|t| \leq \frac{2\delta}{Mr^2}$, the positive margin δ dominates the possible negative perturbation from th. Then $\phi(y) + th(y)$ still stays above the affine plane through $(x, f(x))$ with slope $\nabla\phi(x) + t\nabla h(x)$. That's what ensures $a_t(y) \leq f(y)$ for all y outside $B(x, r)$.

Summary.

- $\delta > 0$ comes from the fact that the quadratic correction in Alexandrov's expansion is strictly positive definite, and hence the difference with the tangent plane is continuous and strictly positive on the sphere.
- The extension beyond radius r is just convexity along rays, which scales up that positive difference.

Together this ensures that the affine minorant with slope $\nabla\phi(x) + t\nabla h(x)$ is globally below $\phi + th$ when $|t|$ is small (depending on x), giving the local equality between f and f^{**} at x.

Author's comment: Somehow we received a second answer to our last prompt.

Response of GPT-5-pro

Good — let me slow down and make the "away from x" part completely transparent.
We are trying to construct, for $f = \phi + th$, an **affine minorant**

$$a_t(y) = f(x) + \langle p_t,\, y - x\rangle, \qquad p_t = \nabla\phi(x) + t\nabla h(x),$$

valid for all y, that touches f at x.

Step 1. Why there is a positive "gap" at distance r

At $t = 0$, by Alexandrov differentiability with Hessian $H = \nabla^2\phi(x) \succ 0$, we have a **local quadratic support**:

$$\phi(x+u) \geq \phi(x) + \langle\nabla\phi(x), u\rangle + \tfrac{m}{2}\|u\|^2, \quad \|u\| \leq r$$

for some $m > 0, r > 0$. In particular, if you restrict to the *sphere* $\|u\| = r$, this lower bound says

$$\phi(x+u) - \phi(x) - \langle\nabla\phi(x), u\rangle \geq \tfrac{m}{2}\, r^2.$$

So if you define

$$\delta := \min_{\|u\|=r} \big(\phi(x+u) - \phi(x) - \langle\nabla\phi(x), u\rangle\big),$$

then by the inequality above $\delta \geq \frac{m}{2}r^2 > 0$. That is where the positive number δ comes from.

Step 2. Extend the gap outside the ball

Now why does this inequality propagate to all $\|y - x\| \geq r$? This is simply convexity of ϕ.
Fix any direction $u \in \mathbb{S}^{d-1}$. Consider the function along the ray

$$g(s) := \phi(x+su) - \phi(x) - \langle\nabla\phi(x), su\rangle, \quad s \geq 0.$$

- From the quadratic support near 0 we know $g(s) \geq \frac{m}{2}s^2$ for small s.
- Convexity of ϕ implies convexity of $g(s)$ as a function of s.
- In particular, $g(s)/s$ is nondecreasing in $s > 0$.

So if at radius r we have $g(r) \geq \delta$, then for every $s \geq r$,

$$\frac{g(s)}{s} \geq \frac{g(r)}{r} \geq \frac{\delta}{r}.$$

Thus for every $y = x + su$ with $\|y - x\| = s \geq r$,

$$\phi(y) \geq \phi(x) + \langle \nabla\phi(x), y - x\rangle + \tfrac{\delta}{r}\,\|y - x\|.$$

This is the **"boundary gap" inequality**:
Author's comment: The proof of the boundary gap is now correct.
Convexity ensures that once there is a gap δ on the sphere of radius r, that gap (rescaled linearly with distance) persists for all larger radii.

Step 3. Why this helps with the affine minorant

For y with $\|y - x\| \geq r$, we thus have

$$\phi(y) \geq \phi(x) + \langle \nabla\phi(x), y - x\rangle + \tfrac{\delta}{r}\,\|y - x\|.$$

Adding the Taylor lower bound for h,

$$h(y) \geq h(x) + \langle \nabla h(x), y - x\rangle - \tfrac{M}{2}\|y - x\|^2,$$

we get for $f(y) = \phi(y) + th(y)$:

$$f(y) \geq f(x) + \langle \nabla\phi(x) + t\nabla h(x), y - x\rangle + \frac{\delta}{r}\|y - x\| - \tfrac{|t|M}{2}\|y - x\|^2.$$

So indeed,

$$f(y) \geq a_t(y) + \left(\frac{\delta}{r}\|y - x\| - \tfrac{|t|M}{2}\|y - x\|^2\right).$$

The term in parentheses is nonnegative whenever

$$|t| \leq \frac{2\delta}{Mr^2}.$$

Author's comment: The model is again bounding $-\frac{|t|M}{2}\|y-x\|^2 \geq -\frac{|t|Mr}{2}\|y-x\|$ which is true when $\|y-x\| \leq r$, whereas here y is outside the ball: $\|y - x\| \geq r$.
That gives the "away from x" condition.

Summary.

- The constant $\delta > 0$ is guaranteed because the local quadratic lower bound forces a strict gap on the sphere $\|y - x\| = r$.
- Convexity propagates this gap outwards (linearly in $\|y - x\|$), so you can control all y outside the ball.
- That's how you obtain a **global affine minorant** touching f at x whenever $|t|$ is small enough.

So yes: you can phrase the whole proof in the "affine envelope" language: for each x there is an affine function a_t with slope $p_t = \nabla\phi(x) + t\nabla h(x)$ that underestimates $f = \phi + th$ globally and touches it at x. That is exactly what shows $f^{**}(x) = f(x)$ and then $\nabla f^{**}(x) = \nabla f(x)$ at x.

Author's comment: [...]

Author's prompt

Outside the ball the quadratic penalty coming from h becomes larger. Not r^2.

Response of GPT-5-pro

You're right to be wary of the "r^2" that showed up before. There is a simpler, cleaner way to do the "away from x" part that doesn't rely on a crude quadratic bound at large radii, and it makes explicit why the choice of t can (and in general will) depend on x.

I'll give a short proof that produces an affine minorant a_t touching $f = \phi + th$ at x and staying below f everywhere when $|t|$ is small (depending on x). The only inputs are:

- Alexandrov at x: ϕ has a positive-definite Hessian $H = \nabla^2\phi(x) \succ 0$.
- $h \in C^2$ (no global convexity of h needed; compact support makes life even easier).

Throughout, write

$$p_0 := \nabla\phi(x), \qquad q := \nabla h(x).$$

We will use the **affine-envelope characterization** of the biconjugate:

$$f^{**}(x) = \sup\{a(x) : a \text{ affine, } a \le f \text{ on } \mathbb{R}^d\}.$$

So it suffices to build **one affine minorant** a_t with $a_t \le f$ everywhere and $a_t(x) = f(x)$. We'll choose

$$a_t(y) := f(x) + \langle p_0,\, y - x\rangle = \phi(x) + t\,h(x) + \langle \nabla\phi(x), y - x\rangle.$$

Note: I am **not** including the $t\,q$ slope; using p_0 (the slope of ϕ at x) is the safe choice globally.
The task is to show:

$$\phi(y) + t\,h(y) \ge \phi(x) + t\,h(x) + \langle p_0,\, y - x\rangle \qquad \forall y \in \mathbb{R}^d,$$

for $|t|$ small (depending on x). We split the space into two regions.

---

1) Inside a small ball $B(x, r)$

From Alexandrov and $H \succ 0$, for any small $\varepsilon \in (0, \lambda_{\min}(H)/2)$ we have (for some $r > 0$)

$$\phi(x+u) \ge \phi(x) + \langle p_0, u\rangle + \frac{m}{2}\|u\|^2, \qquad m := \lambda_{\min}(H) - \varepsilon > 0,\ \ \|u\| \le r.$$

Also, by a Taylor bound for h on $B(x, r)$,

$$h(x+u) \ge h(x) + \langle q, u\rangle - \frac{M}{2}\|u\|^2, \qquad M := \sup_{B(x,r)} \|\nabla^2 h\|.$$

Therefore, for $\|u\| \le r$,

$$\begin{aligned} f(x+u) - a_t(x+u) &= (\phi(x+u) - \phi(x) - \langle p_0, u\rangle) + t(h(x+u) - h(x)) \\ &\ge \frac{m}{2}\|u\|^2 + t\langle q, u\rangle - \frac{|t|M}{2}\|u\|^2. \end{aligned}$$

To handle the linear term $t\langle q, u\rangle$, complete the square (or use Young's inequality):

$$t\langle q, u\rangle \ge -\frac{m}{4}\|u\|^2 - \frac{t^2\|q\|^2}{m}.$$

Thus, for $\|u\| \le r$,

$$f(x+u) - a_t(x+u) \ge \Big(\frac{m}{2} - \frac{|t|M}{2} - \frac{m}{4}\Big)\|u\|^2 - \frac{t^2\|q\|^2}{m}.$$

If you choose $|t| \le m/M$, the coefficient in front of $\|u\|^2$ is $\ge m/4 > 0$, hence the RHS is $\ge -\frac{t^2\|q\|^2}{m}$. In particular,

$$\inf_{\|u\|\le r} \big(f(x+u) - a_t(x+u)\big) \ge -\frac{t^2\|q\|^2}{m}.$$

Since $t \to 0$, this term is $O(t^2)$. If you want **nonnegativity** (not just a tiny negative), simply shrink r a bit **and** shrink $|t|$ further so that the positive quadratic part dominates the linear perturbation everywhere on $B(x,r)$; equivalently, fix $|t| \le \min\{m/M,\, mr/(2\|q\|)\}$. Then

$$f(x+u) \ge a_t(x+u) \qquad (\|u\| \le r).$$

This is the "inside the ball" part.
(If you prefer to be even more explicit: picking $|t| \le m/M$ ensures the quadratic coefficient is $\ge m/4$. Then pick $|t| \le \frac{mr}{4\|q\|}$ so that $\frac{m}{4}\|u\|^2 \ge |t|\|q\|\,\|u\|$ for all $\|u\| \le r$.)
Author's comment: The condition $|t| \le \frac{mr}{4\|q\|}$ does not imply $\frac{m}{4}\|u\|^2 \ge |t|\|q\|\,\|u\|$. Same type of mistake as above.

---

2) Outside the ball $B(x,r)$

Here we **don't** use a quadratic bound. We use a **fixed gap** for ϕ, and the **compactness of the support of** h (or, more generally, that h is bounded on the annulus we consider).
First, because ϕ sits strictly above its tangent plane on the sphere $\partial B(x,r)$, continuity yields

$$\delta := \min_{\|u\|=r} \big(\phi(x+u) - \phi(x) - \langle p_0, u\rangle\big) > 0.$$

Convexity along rays implies this same gap persists for any point farther out:

$$\phi(y) \ge \phi(x) + \langle p_0, y-x\rangle + \delta, \qquad \|y-x\| \ge r.$$

Now split the outside region into two parts:

- **Far outside, where** $h(y) = 0$ (this is where compact support helps): Choose $R_x > r$ so that $\mathrm{supp}(h) \subset B(x, R_x)$ (this is always possible for a fixed x, since $\mathrm{supp}(h)$ is compact). For $\|y-x\| \ge R_x$, $h(y) = 0$, and hence

  $$f(y) = \phi(y) \ge \phi(x) + \langle p_0, y-x\rangle = a_t(y) - t\,h(x),$$

  which is **even stronger** than needed; in fact $f(y) \ge a_t(y)$ because the extra $t\,h(x)$ appears on the other side of a_t's definition (we included $th(x)$ in a_t already).

  Author's comment: The displayed inequality is not necessarily stronger.

- **The annulus** $r \le \|y-x\| \le R_x$: On this compact set, h is bounded, so

  $$c_x := \sup_{r \le \|y-x\| \le R_x} |h(y) - h(x)| < \infty.$$

  Using the gap for ϕ,

  $$f(y) - a_t(y) = \underbrace{\phi(y) - \phi(x) - \langle p_0, y-x\rangle}_{\ge \delta} + t\big(h(y) - h(x)\big) \ge \delta - |t|\,c_x.$$

Author's comment: But, x is not on the annulus to bound $t\big(h(y)-h(x)\big) \geq -|t|c_x$.

Hence on the annulus we ensure $f(y) \geq a_t(y)$ by choosing

$$|t| \leq \frac{\delta}{c_x}.$$

Putting the two parts together, for

$$|t| \leq t_x := \min\left(\frac{m}{M}, \frac{mr}{2\|q\|}, \frac{\delta}{c_x}\right)$$

we have $f(y) \geq a_t(y)$ for **every** $y \in \mathbb{R}^d$, with equality at x. Thus a_t is a **global affine minorant touching at** x, and by the affine-envelope characterization,

$$(\phi + th)^{**}(x) \geq a_t(x) = f(x).$$

Since $(\phi + th)^{**} \leq f$ always, we conclude

$$(\phi + th)^{**}(x) = f(x) = \phi(x) + t\,h(x).$$

Finally, the **gradient equality at** x follows from the "sandwich" subgradient argument that does *not* use convexity of f: if $g = f^{**} \leq f$ and $g(x) = f(x)$, then any $p \in \partial g(x)$ satisfies the global support inequality $f(z) \geq f(x) + \langle p, z - x\rangle$ for all z; differentiability of f at x forces $p = \nabla f(x)$. Hence

$$\nabla(\phi + th)^{**}(x) = \nabla(\phi + th)(x) = \nabla\phi(x) + t\,\nabla h(x).$$

Why this avoids your concern

- We **do not** try to control a negative quadratic outside the ball. Instead, we use a **fixed linear gap** δ for ϕ outside $B(x, r)$, and the fact that on the **bounded annulus** where h can be nonzero we have a **uniform bound** $|h(y) - h(x)| \leq c_x$. So we require $|t| \leq \delta/c_x$ there. For the region **beyond the annulus**, $h(y) = 0$ and convexity of ϕ alone gives $f(y) \geq a_t(y)$.
- The constants (r, m, δ, R_x, c_x) depend on x, so t_x does too. That's expected and harmless: we only need the **pointwise** first-order information at each x (to identify the Lagrangian velocity); no uniform t-window is required for the rest of the proof, which is carried out after OU smoothing.

If you'd like, I can compact this into a lemma you can paste—"Affine minorant touching at x for $\phi + th$"—with the constants spelled out exactly as above.

Author's prompt

Inside the ball, you have to take t depending on u. Not allowed.

Response of GPT-5-pro

You're right — the way I wrote the "inside the ball" part looked like the choice of t might depend on $\|u\|$. Here's the clean fix with a **single** t (depending on x, but **not** on u) that works **uniformly for all** $\|u\| \leq r$.

Goal (inside $B(x, r)$)

We want

$$f(x+u) - a_t(x+u) \geq 0 \qquad \text{for all } \|u\| \leq r,$$

with

$$f = \phi + t\,h, \quad a_t(y) = f(x) + \langle p_0,\, y - x\rangle, \quad p_0 := \nabla\phi(x), \quad q := \nabla h(x).$$

From the one-point quadratic lower support for ϕ and the quadratic Taylor bound for h, we have (for some $m > 0$, all $\|u\| \leq r$)

$$\begin{aligned} f(x+u) - a_t(x+u) &= (\phi(x+u) - \phi(x) - \langle p_0, u\rangle) + t(h(x+u) - h(x)) \\ &\geq \frac{m}{2}\|u\|^2 + t\langle q, u\rangle - \frac{|t|M}{2}\|u\|^2 \\ &=: \alpha\,\|u\|^2 - \beta\,\|u\|, \end{aligned}$$

where

$$\alpha = \frac{m - |t|M}{2}, \qquad \beta = |t|\,\|q\|.$$

So the problem reduces to ensuring

$$\alpha s^2 - \beta s \geq 0 \quad \text{for all } s \in [0, r]. \tag{Q}$$

Author's comment: This approach does not sound right from the beginning, because the slope of a_t is p_0 and not $p_0 + tq$. We see in the next section that the model is making a number of mistakes trying to justify its approach based on (Q).

A uniform choice of t that guarantees (Q)

- First, we need $\alpha > 0$, i.e.

$$|t| < \frac{m}{M}. \tag{C1}$$

- The convex quadratic $s \mapsto \alpha s^2 - \beta s$ attains its minimum on $[0, r]$ either at the endpoint $s = r$ or at the (unconstrained) minimizer $s^\star = \beta/(2\alpha)$ if $s^\star \leq r$. To avoid the interior minimum (which would give a negative value unless $\beta = 0$), we **force** the minimizer to lie **beyond** r:

$$\frac{\beta}{2\alpha} \geq r \iff \alpha r \geq \beta. \tag{C2}$$

  Under (C2), the minimum on $[0, r]$ occurs at $s = r$, and then

$$\min_{s\in[0,r]}(\alpha s^2 - \beta s) = \alpha r^2 - \beta r \geq 0 \quad \text{iff } \alpha r \geq \beta,$$

  i.e. the same condition.

Plugging α, β in (C2) gives the **uniform** bound

$$\frac{m - |t|M}{2} r \geq |t|\,\|q\| \iff |t| \leq \frac{mr}{Mr + 2\|q\|}. \tag{C2'}$$

Thus, a single choice

$$\boxed{|t| \leq t_x^{\text{in}} := \min\left\{\frac{m}{M},\ \frac{mr}{Mr + 2\|q\|}\right\}}$$

ensures $f(x+u) - a_t(x+u) \geq 0$ **for every** $\|u\| \leq r$.
(If $q = \nabla h(x) = 0$, condition (C2') is trivial; you only need $|t| < m/M$ to keep $\alpha > 0$.)

Outside the ball (quick recap without any r^2 issue)

- On $\partial B(x,r)$, by continuity and the strict positivity coming from the Alexandrov bound, the **gap**

$$\delta := \min_{\|u\|=r} \big(\phi(x+u) - \phi(x) - \langle p_0, u\rangle\big) \ > \ 0.$$

  Convexity along rays propagates this: for all $\|y-x\| \geq r$, $\phi(y) \geq \phi(x) + \langle p_0, y-x\rangle + \delta$.

- Since h is continuous and the annulus $r \leq \|y-x\| \leq R_x$ is compact (choose R_x so $\mathrm{supp}(h) \subset B(x, R_x)$), set

$$c_x := \sup_{r \leq \|y-x\| \leq R_x} |h(y) - h(x)| < \infty.$$

  Then

$$f(y) - a_t(y) = \underbrace{\phi(y) - \phi(x) - \langle p_0, y-x\rangle}_{\geq \delta} + t\,(h(y) - h(x)) \geq \delta - |t|\, c_x.$$

  Author's comment: The model is again confused on who is x. It is not on the annulus. However, our final proof uses a similar bound to control $h(y) - h(x)$.

  So for $|t| \leq \delta / c_x$, we have $f \geq a_t$ on the annulus; and beyond R_x we have $h = 0$, hence $f = \phi \geq a_t$ by convexity of ϕ.

Final uniform t (depending on x, not on u)

Take

$$\boxed{t_x := \min\left\{ t_x^{\mathrm{in}},\ \frac{\delta}{c_x} \right\} = \min\left\{ \frac{m}{M},\ \frac{mr}{Mr + 2\|q\|},\ \frac{\delta}{c_x} \right\}.}$$

Then for **every** $y \in \mathbb{R}^d$, $f(y) \geq a_t(y)$ when $|t| \leq t_x$, with equality at x. Hence a_t is a **global affine minorant touching at** x, so by the affine-envelope characterization

$$(\phi + th)^{**}(x) = f^{**}(x) = f(x) = \phi(x) + t\,h(x).$$

Finally, the "sandwich" argument with differentiable f at x gives

$$\nabla(\phi + th)^{**}(x) = \nabla(\phi + th)(x) = \nabla\phi(x) + t\nabla h(x).$$

Takeaway. The inside-ball nonnegativity can be enforced with a **single, uniform** inequality (C2'), so t does **not** depend on $\|u\|$. The outside-ball part uses a fixed linear gap for ϕ and boundedness of h on a compact annulus, giving another uniform bound $|t| \leq \delta / c_x$. Combining the two yields a **pointwise** $t_x > 0$ (depending on x, as expected) which makes the affine minorant work globally.

Author's comment: At this stage of the chat, we had enough ingredients to write a full and correct proof, see Section 2.