The Large-k Side Condition in the Main Induction

2026-04-29

1 At a glance

- **Difficulty: medium.** The issue is not a missing analytic estimate, but a mismatch between the size hypothesis printed in the paper and the size hypothesis consumed by the present pasting route.
- **Estimated weight:** about **3–6 scalar or boundary lemmas** if one proves a smaller-k saturation argument, or a theorem-level correction if the present large-k route is retained.
- **Mathlib/project split:** approximately **10% Mathlib** and **90% project-local**. The arithmetic is elementary; the substance is deciding which LDT theorem statement is justified by the paper's pasting pipeline.
- **Key Mathlib inputs:** elementary order arithmetic on natural numbers and nonnegative real error parameters. The main inputs are the project-local statements of the main induction, pasting, and final scalar cascade. Tracked in GitHub issue #906.

2 Key theorem forms

The comparison involves three theorem forms.

1. **Printed source form.** The theorem in `references/ldt-paper/inductive_step.tex` assumes $k \geq md$, and its successor proof assumes $k \geq (m+1)d$.
2. **Pasting-input form.** The pasting theorem used in the successor proof requires $k \geq 400md$ in the restricted m-dimensional slice.

3. **Current formal-interface form.** The Lean theorem MIPStarRE.LDT.MainInductionStep.mainInduction and the final theorem MIPStarRE.LDT.Test.mainFormal use the corrected large-k interface, displayed in the blueprint as separate current-interface entries.

3 The assertion

This note concerns the k-hypothesis in the main induction theorem of [JNV$^+$20, Section *The main induction step*], and its use in the proof of the main formal soundness theorem. [1] Here m is the number of variables, q is the field size, d is the individual degree bound, and k is the sampling parameter in the consistency estimates. The integer k is carried unchanged through the induction on m.

In the theorem as stated, an $(\varepsilon, \delta, \gamma)$-good symmetric strategy for the (m, q, d) low individual degree test is assumed to satisfy only

$$k \geq md. \qquad (\texttt{thm:main-induction})$$

Under this hypothesis, the theorem produces a polynomial measurement $G \in \mathrm{PolyMeas}(m, q, d)$ satisfying the advertised consistency estimate. The same k-condition is then inherited by the stated main formal soundness theorem after the reduction from a general two-prover strategy to a symmetric one.

4 The use of pasting

The successor step proves the induction theorem in dimension $m + 1$ assuming the theorem in dimension m. In that step the ambient strategy is a strategy for the $(m + 1, q, d)$ test, and the proof assumes

$$k \geq (m + 1)d. \qquad (\texttt{sec:proof-of-main-induction})$$

This is sufficient to apply the induction hypothesis to each x-restricted strategy, since the restricted strategies have dimension m, and $(m + 1)d \geq md$.

After the restricted measurements have been improved to projective submeasurements, the proof pastes them into a measurement in dimension $m+1$.

[1] The original discrepancy was recorded in GitHub issue #906 and first implemented in PR #912. The documentation issue for this note is GitHub issue #928. In the TeX source consulted here, the main induction theorem is in `references/ldt-paper/inductive_step.tex`, lines 7–18; the successor step is in the same file, lines 429–568; and the pasting theorem used there is in lines 299–330.

The pasting theorem invoked at this point has its own size hypothesis. In the notation of the successor step, it requires

$$k \geq 400md. \qquad\qquad (\texttt{thm:ld-pasting-in-induction-section})$$

This condition is not a consequence of $k \geq (m+1)d$. For instance, when $m = 1$ and $d > 0$, the latter allows $k = 2d$, whereas the pasting theorem requires $k \geq 400d$. Thus the argument as written does not justify the successor step throughout the interval in which $k \geq (m+1)d$ but $k < 400md$.

Equivalently, the missing implication is

$$k \geq (m+1)d \quad \Longrightarrow \quad k \geq 400md,$$

which is false for the relevant range of parameters. The issue is therefore a missing side condition, not a notational convention or a peculiarity of the formal language.

5 The corrected formal interface

The repair made in the checked formal statements is to state explicitly the hypothesis used by the proof. The source-labelled blueprint theorem is now stated with the corrected bound $k \geq 400md$. In the successor step from m to $m+1$, one assumes $k \geq 400(m+1)d$. This immediately supplies the pasting hypothesis $k \geq 400md$ needed for the restricted dimension.

The current formal interface to the main formal theorem is adjusted in the same way: its hypothesis on k is $k \geq 400md$, rather than $k \geq md$. It also requires $k > 0$. This positivity is not a bridge or measurement-construction hypothesis; it is a scalar boundary needed for the Step 8 cascade estimates, which use $1 \leq k$ to absorb lower-order terms into the final envelope. The printed theorem should therefore be read as missing both the factor-400 strengthening and the nonzero sampling-parameter boundary for the present proof route. The factor-400 strengthening is now treated as a confirmed statement correction, not as a proof obligation for the interval $md \leq k < 400md$. The degenerate case $k = 0$ when $d = 0$ is handled as the separate statement correction documented in `docs/paper-gaps/issue-422-main-formal-zero-k-boundary.tex`. A trivial construction is used only when the final error parameter is already at least 1, in which case the consistency conclusion is saturated. [2]

[2] After the 2026-05-22 statement correction, the source-labelled blueprint theorem for main induction links to the corrected Lean statement MIP-

6 Non-vacuity of the discarded printed interval

The interval $md \le k < 400md$ is not eliminated by the "large-error" branch of the formal proof. The complementary case

$$m^2\left(\nu + \exp\left(-\frac{k}{80000m^2}\right)\right) < 1.$$

This inequality can hold inside $md \le k < 400md$. For example, take $m = 2$, $k = 2d$, and choose d large enough that

$$4\exp\left(-\frac{d}{160000}\right) < \frac{1}{2}.$$

With $\varepsilon = \delta = \gamma = 0$, the remaining contribution to ν is controlled by $(d/q)^{1/1024}$. Taking the field size q sufficiently large makes

$$4 \cdot 1000 \cdot (2d)^2 \cdot 2^2 \cdot (d/q)^{1/1024} < \frac{1}{2}.$$

Thus the formal small-error branch is not vacuous; it requires a mathematical argument for the source interval, or a corrected statement restricting the theorem to the large-k range used by the pasting route. It should not be replaced by a hidden assumption on the paper-facing theorem. The project now takes the second option: the theorem statement is corrected to the large-k range.

The same phenomenon occurs in the final theorem. The source-boundary reduction MIPStarRE.LDT.Test.mainFormalConclusion proves the branch $\nu \ge$

StarRE.LDT.MainInductionStep.mainInduction. The main formal source theorem links to MIPStarRE.LDT.Test.mainFormal. The final source theorem routes through the named source-boundary reduction MIPStarRE.LDT.Test.mainFormalConclusion; its saturated-error branch is proved, and its non-vacuous branch is MIPStarRE.LDT.Test.mainFormal_smallErrorConclusion. The small-error branch is reduced to the checked role-register scalar absorption theorem under the corrected nonzero sampling hypothesis MIPStarRE.LDT.Test.mainFormalConclusion_ofRoleRegisterScalarBoundary. The scalar lemma MIPStarRE.LDT.Test.mainFormalError_zero_k records that the displayed final-theorem error is exactly zero at $k = 0$. The Lean theorem MIPStarRE.LDT.Test.mainFormal assumes $400\,md \le k$ and $0 < k$; the pasting theorem is MIPStarRE.LDT.Pasting.Bernoulli.ldPastingNontrivial, with the corresponding large-k hypothesis. The final assembly in `MIPStarRE/LDT/Test/MainTheorem/MainFormal.lean` no longer depends on an unfinished corrected large-k Section 6 successor construction. The zero-sampling boundary has likewise been removed from the source statement by the explicit hypothesis $0 < k$.

1 by the trivial two-space consistency bound, but the source theorem also has a complementary branch in which

$$100000k^2m^4\left(\varepsilon^{1/40000} + (d/q)^{1/40000} + \exp\left(-\frac{k}{2560000m^2}\right)\right) < 1.$$

Taking again $m = 2$, $k = 2d$, d sufficiently large, and then q sufficiently large, with $\varepsilon = 0$, gives examples in the printed interval $md \leq k < 400md$ where this final error is below 1. Hence the factor-400 correction is also necessary for the final theorem route; after this correction, the zero-sampling boundary is handled by the separate nonzero-k correction.

7 Conclusion

The verdict is that issue #906 records a genuine theorem-level side-condition gap in the stated proof. The hypothesis $k \geq md$, or in the successor step $k \geq (m+1)d$, does not imply the large-k hypothesis required by the pasting theorem. The current blueprint therefore records the confirmed correction directly in the source-labelled main-induction statement. The final theorem route records the same large-k correction; the zero-sampling obstruction is handled by the separate nonzero sampling correction recorded in `docs/paper-gaps/issue-422-main-formal-zero-k-boundary.tex`.

This correction does not assert that the mathematical conclusion is false for $md \leq k < 400md$. It says that the cited proof route through the stated pasting theorem does not establish that range. The formalization therefore uses the large-k theorem that the proof actually supports. The separate $k = 0$ boundary allowed when $d = 0$ is excluded from the corrected final theorem statement by the hypothesis $0 < k$.

References

[JNV+20] Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. Quantum soundness of the classical low individual degree test, 2020. arXiv:2009.12982.