Pretty-Good Measurement Bounds

Formal statement

Let

$$\mathcal{E} = \{p_x, \rho_x\}_{x \in \mathcal{X}}$$

be a finite ensemble of quantum states on a finite-dimensional Hilbert space mathcal H. The label x is chosen with probability p_x, and the receiver is given the state ρ_x. A measurement for guessing x is a POVM

$$\{M_x\}_{x \in \mathcal{X}}, \qquad M_x \geq 0, \qquad \sum_x M_x = I.$$

Its success probability is

$$P_{\mathrm{succ}}(M) = \sum_x p_x \operatorname{Tr}(M_x \rho_x).$$

The optimal success probability is

$$P_{\mathrm{opt}} = \max_{\{M_x\}} \sum_x p_x \operatorname{Tr}(M_x \rho_x).$$

The pretty-good measurement, also called the square-root measurement, is defined as follows. Put

$$A_x = p_x \rho_x$$

and let

$$\Gamma = \sum_x A_x = \sum_x p_x \rho_x$$

be the average state of the ensemble. On the support of Gamma, define

$$M_x^{\mathrm{PGM}} = \Gamma^{-1/2} A_x \Gamma^{-1/2}.$$

If Gamma is not invertible, Gamma(-1/2) means the inverse square root on supp(Gamma), equivalently the square root of the Moore-Penrose pseudoinverse. The operators Mx(PGM) sum to the projector onto supp(Gamma). One may add any POVM completion on ker(Gamma), because none of the ensemble states has support there, and this completion does not affect the success probability.

The central universal bound is the Barnum-Knill pretty-good measurement bound:

$$P_{\mathrm{PGM}} \geq P_{\mathrm{opt}}^2.$$

Equivalently,

$$P_{\mathrm{opt}} \leq \sqrt{P_{\mathrm{PGM}}}.$$

In terms of error probabilities,

$$P_{\mathrm{err}}^{\mathrm{PGM}} = 1 - P_{\mathrm{PGM}}, \qquad P_{\mathrm{err}}^{\mathrm{opt}} = 1 - P_{\mathrm{opt}},$$

this implies

$$P_{\mathrm{err}}^{\mathrm{PGM}} \leq 2P_{\mathrm{err}}^{\mathrm{opt}}.$$

Indeed,

$$1 - P_{\mathrm{PGM}} \leq 1 - P_{\mathrm{opt}}^2 = (1 - P_{\mathrm{opt}})(1 + P_{\mathrm{opt}}) \leq 2(1 - P_{\mathrm{opt}}).$$

Thus, whenever optimal discrimination is already highly reliable, the pretty-good measurement is also highly reliable. Hausladen and Wootters introduced the measurement as a simple general prescription for distinguishing several possible states, especially when the states are equally likely and nearly orthogonal. Watrous presents the finite-dimensional PGM definition and states the Barnum-Knill guarantee in the form above.

Why the name “pretty good” is mathematically honest

For two hypotheses, the Helstrom theorem gives a closed-form optimal measurement. For three or more states, no comparably simple closed-form expression exists in general. The optimal measurement can be formulated as a semidefinite program, but the solution need not have a transparent analytic form.

The pretty-good measurement is a canonical closed-form substitute. It takes each weighted state

$$A_x = p_x \rho_x$$

and normalizes it by the square root of the total average state

$$\Gamma = \sum_x A_x.$$

The operation

$$A_x \mapsto \Gamma^{-1/2} A_x \Gamma^{-1/2}$$

can be read as follows. The average state Gamma describes the total region of Hilbert space occupied by the whole ensemble. If a particular signal state A_x lies in a crowded direction where many other states also have support, $\text{Gamma}^{(-1/2)}$ suppresses that direction. If it lies in a relatively distinctive direction, it is less suppressed. So the PGM does not merely project onto ρ_x. It whitens the ensemble first, then asks which whitened signal component is present.

The name “pretty good” should not be interpreted as “always almost optimal additively.” The universal theorem says

$$P_{\text{PGM}} \geq P_{\text{opt}}^2.$$

If P_{opt} is close to one, this is strong. If $P_{opt}=0.99$, then

$$P_{\text{PGM}} \geq 0.9801,$$

so the PGM error is at most about twice the optimal error. But if $P_{opt}=0.1$, the theorem only gives

$$P_{\mathrm{PGM}} \geq 0.01.$$

That lower bound is weak. The PGM is most powerful in settings where the ensemble is well designed, nearly orthogonal, symmetric, or appears as a codebook whose states become asymptotically distinguishable.

Proof of the Barnum-Knill bound

We now prove the main universal guarantee in finite dimension.

Use the subnormalized notation

$$A_x = p_x \rho_x, \qquad \Gamma = \sum_x A_x.$$

Let

$$\mu_x = M_x^{\mathrm{PGM}}$$

be the PGM effects. Let

$$\nu_x$$

be any other POVM. We will show

$$\left(\sum_x \mathrm{Tr}(\nu_x A_x) \right)^2 \leq \sum_x \mathrm{Tr}(\mu_x A_x).$$

Since this holds for every POVM \nuₓ\, it also holds for an optimal POVM. That will give

$$P_{\mathrm{opt}}^2 \leq P_{\mathrm{PGM}}.$$

For simplicity, first assume Gamma is invertible. The singular case is handled by restricting everything to supp(Gamma).

For each x, write

$$\mathrm{Tr}(\nu_x A_x) = \left\langle \Gamma^{1/4}\nu_x\Gamma^{1/4}, \Gamma^{-1/4}A_x\Gamma^{-1/4} \right\rangle,$$

where

$$\langle B, C\rangle = \mathrm{Tr}(B^\dagger C)$$

is the Hilbert-Schmidt inner product. By Cauchy-Schwarz,

$$\mathrm{Tr}(\nu_x A_x) \le \left\|\Gamma^{1/4}\nu_x\Gamma^{1/4}\right\|_2 \left\|\Gamma^{-1/4}A_x\Gamma^{-1/4}\right\|_2.$$

Summing over x and applying the ordinary Cauchy-Schwarz inequality for real vectors gives

$$\left(\sum_x \mathrm{Tr}(\nu_x A_x)\right)^2 \le \left(\sum_x \left\|\Gamma^{1/4}\nu_x\Gamma^{1/4}\right\|_2^2\right)\left(\sum_x \left\|\Gamma^{-1/4}A_x\Gamma^{-1/4}\right\|_2^2\right).$$

We estimate the first factor. Since $0 \le \nu_x \le I$, one has

$$\nu_x^2 \le \nu_x.$$

Therefore

$$\begin{aligned}\left\|\Gamma^{1/4}\nu_x\Gamma^{1/4}\right\|_2^2 &= \mathrm{Tr}\left(\Gamma^{1/4}\nu_x\Gamma^{1/2}\nu_x\Gamma^{1/4}\right)\\ &= \mathrm{Tr}\left(\nu_x\Gamma^{1/2}\nu_x\Gamma^{1/2}\right)\\ &\le \mathrm{Tr}\left(\nu_x\Gamma\right).\end{aligned}$$

Summing over x,

$$\sum_x \left\|\Gamma^{1/4}\nu_x\Gamma^{1/4}\right\|_2^2 \le \sum_x \mathrm{Tr}(\nu_x\Gamma) = \mathrm{Tr}\,\Gamma = 1.$$

Now estimate the second factor:

$$\begin{aligned}\left\|\Gamma^{-1/4}A_x\Gamma^{-1/4}\right\|_2^2 &= \mathrm{Tr}\left(\Gamma^{-1/4}A_x\Gamma^{-1/2}A_x\Gamma^{-1/4}\right)\\ &= \mathrm{Tr}\left(\Gamma^{-1/2}A_x\Gamma^{-1/2}A_x\right)\\ &= \mathrm{Tr}(\mu_x A_x).\end{aligned}$$

Hence

$$\sum_x \left\|\Gamma^{-1/4}A_x\Gamma^{-1/4}\right\|_2^2 = \sum_x \mathrm{Tr}(\mu_x A_x) = P_{\mathrm{PGM}}.$$

Putting the two estimates together gives

$$\left(\sum_x \mathrm{Tr}(\nu_x A_x)\right)^2 \le 1 \cdot P_{\mathrm{PGM}}.$$

Thus

$$P_{\mathrm{succ}}(\nu)^2 \le P_{\mathrm{PGM}}$$

for every POVM nu. Taking nu to be an optimal measurement gives

$$P_{\mathrm{opt}}^2 \le P_{\mathrm{PGM}}.$$

This proves the Barnum-Knill bound.

The proof explains why the PGM has the form it does. The measurement arises from inserting Gamma(1/4) and Gamma(-1/4) around the arbitrary measurement and then using Cauchy-Schwarz. The square-root normalization is exactly the normalization that makes the inequality close.

Pure-state Gram-matrix form

The PGM becomes especially transparent for pure-state ensembles. Suppose

$$\rho_x = |\psi_x\rangle\langle\psi_x|.$$

Define the weighted signal vectors

$$|\phi_x\rangle = \sqrt{p_x}|\psi_x\rangle.$$

Then

$$A_x = |\phi_x\rangle\langle\phi_x|, \qquad \Gamma = \sum_x |\phi_x\rangle\langle\phi_x|.$$

The PGM vector corresponding to label x is

$$|m_x\rangle = \Gamma^{-1/2}|\phi_x\rangle,$$

and the measurement effect is

$$M_x^{\mathrm{PGM}} = |m_x\rangle\langle m_x|$$

on the support of Gamma, with the usual caveat about linear dependence and kernel completion.

Let G be the Gram matrix of the weighted vectors:

$$G_{xy} = \langle\phi_x|\phi_y\rangle = \sqrt{p_x p_y}\langle\psi_x|\psi_y\rangle.$$

Then the PGM success probability has the compact formula

$$P_{\mathrm{PGM}} = \sum_x \left(\sqrt{G}_{xx}\right)^2.$$

This formula makes the “nearly orthogonal” behavior visible. If the weighted Gram matrix is close to diagonal, then sqrt G is close to the square root of that diagonal matrix, and the PGM success probability is close to one.

Example 1: orthogonal states

Suppose Alice sends one of m orthonormal states

$$|1\rangle, \ldots, |m\rangle$$

with probabilities p_1, ...,p_m. Then

$$\Gamma = \sum_x p_x |x\rangle\langle x|.$$

For every x with $p_x > 0$,

$$\Gamma^{-1/2} p_x |x\rangle\langle x| \Gamma^{-1/2} = |x\rangle\langle x|.$$

Thus the PGM is just the projective measurement in the orthonormal signal basis. It identifies the state perfectly:

$$P_{\text{PGM}} = P_{\text{opt}} = 1.$$

This is the classical limit. When the possible quantum states are mutually orthogonal, no subtle measurement design is needed.

Example 2: identical states

Suppose every label is encoded into the same state:

$$\rho_x = \rho$$

for all x. Then no measurement can reveal anything about x. The optimal strategy is to always guess the most likely label, so

$$P_{\text{opt}} = \max_x p_x.$$

For equal priors over m labels,

$$p_x = \frac{1}{m},$$

one has

$$P_{\mathrm{opt}} = \frac{1}{m}.$$

The PGM also gives a completely uninformative measurement on the signal support and succeeds with probability

$$P_{\mathrm{PGM}} = \frac{1}{m}.$$

The universal bound says only

$$\frac{1}{m} \geq \frac{1}{m^2},$$

which is true but weak. This example reminds us that “pretty good” does not mean the ensemble is easy to distinguish. It means that the PGM competes with the optimum in a universal way.

Example 3: two equally likely pure states

Let

$$\rho = |0\rangle\langle 0|, \qquad \sigma = |+\rangle\langle +|,$$

where

$$|+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}},$$

and let the priors be equal. The overlap is

$$c = |\langle 0|+\rangle| = \frac{1}{\sqrt{2}}.$$

For two equally likely pure states, the PGM success probability is

$$P_{\mathrm{PGM}} = \frac{1}{2}\left(1 + \sqrt{1 - c^2}\right).$$

Here this gives

$$P_{\mathrm{PGM}} = \frac{1}{2}\left(1 + \frac{1}{\sqrt{2}}\right) \approx 0.8536.$$

This equals the Helstrom optimum for two equally likely pure states. In this special case, “pretty good” is actually optimal.

The important lesson is not that PGM is always optimal. It is not. The lesson is that for simple symmetric binary pure-state discrimination, the square-root normalization naturally produces the same sign structure as the Helstrom measurement.

Example 4: why many copies help

Suppose we have m pure states

$$|\psi_1\rangle, \ldots, |\psi_m\rangle$$

whose pairwise overlaps are bounded by

$$|\langle\psi_i|\psi_j\rangle| \le c < 1 \qquad (i \neq j).$$

If we are given k copies, the states become

$$|\psi_i\rangle^{\otimes k}.$$

Their pairwise overlaps shrink exponentially:

$$\left|\left\langle \psi_i^{\otimes k} \middle| \psi_j^{\otimes k} \right\rangle\right| = |\langle\psi_i|\psi_j\rangle|^k \le c^k.$$

Thus the Gram matrix becomes closer and closer to diagonal as k grows. The PGM becomes increasingly reliable because the ensemble becomes increasingly close to an orthogonal ensemble.

This is one of the reasons PGM appears naturally in coding and learning arguments. One does not need to solve the exact optimal measurement. It is enough to show that the relevant code states or hypothesis states become nearly orthogonal, and then the PGM succeeds with high probability. Montanaro proved useful bounds showing that for sets of mixed states with pairwise fidelities bounded away from one, logarithmically many copies can be enough for high-probability discrimination in worst-case settings.

PGM in classical-quantum channel coding

A classical-quantum channel maps a classical input x to a quantum output state ρ_x. A codebook consists of many input codewords, and each codeword produces a quantum state at the receiver. Decoding the message is exactly a many-state quantum discrimination problem.

For a codebook m=1,...,M, one often constructs operators A_m associated with the codeword states or their typical projections and then defines the square-root decoder

$$M_m = \left(\sum_{j=1}^{M} A_j\right)^{-1/2} A_m \left(\sum_{j=1}^{M} A_j\right)^{-1/2}.$$

This is a PGM. It is central in achievability proofs for transmitting classical information through quantum channels. The historical path runs through the Holevo-Schumacher-Westmoreland theorem and related random-coding proofs. Modern one-shot proofs often use the Hayashi-Nagaoka operator inequality to bound the error of the square-root decoder; more recent work shows that the PGM itself can sometimes play the role of a quantum union bound in coding analyses.

The coding intuition is simple. A good random code makes the possible output states nearly distinguishable in the typical subspace. The PGM then acts as a universal decoder: it compares each candidate codeword against the total codeword cloud and assigns the outcome according to the whitened candidate operators.

PGM in hidden-subgroup-type problems

The hidden subgroup problem can often be converted into a state discrimination problem. The oracle produces coset states or hidden subgroup states, and the task is to identify the subgroup from these quantum states.

In several hidden-subgroup settings, the pretty-good measurement is not merely a convenient heuristic. It is optimal or sample-optimal. Bacon, Childs, and van Dam studied hidden subgroup states over certain semidirect product groups and showed that the PGM is optimal in those settings, with its success probability and implementation connected to average-case algebraic problems. Hayashi, Kawachi, and Kobayashi proved general sample-complexity bounds for hidden subgroup identification and showed that an upper bound is attained by the PGM, giving identification with O(log|mathcal H|) samples for a candidate family of hidden subgroups.

This is why PGM appears in hidden-subgroup algorithms. The measurement has the right symmetry. When the ensemble is group-covariant, the square-root construction often respects that symmetry and can coincide with the optimal measurement.

Why the PGM is often optimal for symmetric ensembles

Suppose a group G acts transitively on the labels and the states are generated by this group action. Informally, all labels then look equally difficult. No label has a privileged position. In many such cases, the optimal measurement can be chosen to have the same covariance symmetry as the ensemble.

The PGM also has this covariance property because it is built canonically from the ensemble average

$$\Gamma = \sum_x p_x \rho_x .$$

If the ensemble has a symmetry, Gamma has that symmetry, and the square-root normalization respects it. This is the structural reason PGM becomes optimal in many group-covariant problems.

This does not mean that symmetry always makes the PGM optimal. The precise optimality conditions depend on the ensemble. But symmetry explains why PGM is so often the first measurement to try in coding, hidden subgroup states, geometrically uniform states, and many representation-theoretic discrimination problems.

Relation to optimality conditions

For a general ensemble, a POVM M_x\ is optimal if and only if it satisfies the Holevo-Yuen-Kennedy-Lax optimality conditions. One convenient form is that the operator

$$Y = \sum_x A_x M_x$$

must be Hermitian and must satisfy

$$Y \geq A_x$$

for every x, together with complementary slackness conditions.

The PGM is obtained from a formula, not by directly solving these optimality conditions. Sometimes the PGM satisfies them, and then it is optimal. Sometimes it does not, and then it is merely near-optimal in the Barnum-Knill sense or useful because additional structure gives sharper bounds.

So the logical relation is:

$$\text{PGM is canonical and closed form,}$$

but

$$\text{optimality requires extra structure.}$$

How to use the theorem

Given an ensemble p_x, ρ_x\, first form

$$A_x = p_x \rho_x$$

and

$$\Gamma = \sum_x A_x.$$

Then compute

$$M_x^{\mathrm{PGM}} = \Gamma^{-1/2} A_x \Gamma^{-1/2}$$

on supp(Gamma). The success probability is

$$P_{\mathrm{PGM}} = \sum_x \mathrm{Tr}(M_x^{\mathrm{PGM}} A_x).$$

Immediately, without solving the optimal SDP, one has

$$P_{\mathrm{opt}}^2 \le P_{\mathrm{PGM}} \le P_{\mathrm{opt}}.$$

The upper bound is trivial because the PGM is one particular measurement. The lower bound is the useful theorem.

If $P_{(}PGM_{)}$ is close to one, then the ensemble is reliably distinguishable. If P_{opt} is known or can be lower-bounded close to one, the Barnum-Knill inequality implies that PGM is also good. If the states are pure, use the weighted Gram matrix formula. If the states arise from a code or a group action, exploit the structure of Gamma and the symmetry of the ensemble.

Common mistakes

A common mistake is to define the PGM without the priors. The correct construction uses

$$A_x = p_x \rho_x,$$

not just ρ_x. Changing the priors changes the ensemble, the average state, and therefore the PGM.

A second mistake is to ignore the support of Gamma. If Gamma is singular, $Gamma^{(}-1/2^{)}$ must be interpreted on supp(Gamma), or as a Moore-Penrose pseudoinverse. The measurement may be completed arbitrarily on ker(Gamma), because no input state ever occupies that subspace.

A third mistake is to think that “pretty good” means “always close to optimal by a small additive error.” The universal guarantee is multiplicative in success probability:

$$P_{\mathrm{PGM}} \ge P_{\mathrm{opt}}^2.$$

This gives a strong error comparison only when P_{opt} is close to one.

A fourth mistake is to confuse the PGM with the Helstrom measurement. For two states, the Helstrom measurement is the exact optimum. For many states, the PGM is a canonical closed-form measurement. It may be optimal in special cases, but not in general.

A fifth mistake is to assume that a closed-form POVM is automatically easy to implement. The PGM requires applying Gamma(-1/2), or implementing a measurement equivalent to that operator normalization. In algorithmic settings such as hidden subgroup problems, the circuit implementation of the PGM can be as important as the success probability.

Final mental image

The pretty-good measurement is the measurement obtained by whitening the ensemble and then testing each whitened signal component. The ensemble average

$$\Gamma = \sum_x p_x \rho_x$$

describes the total cloud of possible received states. The PGM effect

$$M_x^{\mathrm{PGM}} = \Gamma^{-1/2} p_x \rho_x \Gamma^{-1/2}$$

asks how much of the whitened received state points in the direction of hypothesis x.

Its universal guarantee is

$$P_{\mathrm{PGM}} \geq P_{\mathrm{opt}}^2.$$

Thus, if the ensemble can be decoded well by any measurement, the PGM cannot be catastrophically bad. And when the states are nearly orthogonal, symmetric, code-like, or group-covariant, the PGM is often not just pretty good but essentially optimal.

In one sentence:

> PGM is the canonical square-root decoder whose success is at least the square o

That is why it appears throughout quantum information theory, especially in state discrimination, classical-quantum coding, and hidden-subgroup-type quantum algorithms.

References

Hausladen, Paul, and William K. Wootters. “A ‘Pretty Good’ Measurement for Distinguishing Quantum States.” Journal of Modern Optics 41, no. 12 (1994): 2385–2390. DOI: 10.1080/09500349414552221. https://doi.org/10.1080/09500349414552221

Hausladen, Paul, Richard Jozsa, Benjamin Schumacher, Michael Westmoreland, and William K. Wootters. “Classical Information Capacity of a Quantum Channel.” Physical Review A 54, no. 3 (1996): 1869–1876.

Barnum, Howard, and Emanuel Knill. “Reversing Quantum Dynamics with Near-Optimal Quantum and Classical Fidelity.” Journal of Mathematical Physics 43 (2002): 2097–2106. arXiv:quant-ph/0004088. https://arxiv.org/abs/quant-ph/0004088

Watrous, John. The Theory of Quantum Information. Cambridge University Press, 2018. See the section on quantum state discrimination and Theorem 3.10 on the Barnum-Knill bound. https://cs.uwaterloo.ca/~watrous/TQI/TQI.pdf

Hayashi, Masahito, and Hiroshi Nagaoka. “General Formulas for Capacity of Classical-Quantum Channels.” IEEE Transactions on Information Theory 49, no. 7 (2003): 1753–1768.

Bacon, Dave, Andrew M. Childs, and Wim van Dam. “From Optimal Measurement to Efficient Quantum Algorithms for the Hidden Subgroup Problem over Semidirect Product Groups.” FOCS 2005, 469–478. arXiv:quant-ph/0504083. https://arxiv.org/abs/quant-ph/0504083

Hayashi, Masahito, Akinori Kawachi, and Hirotada Kobayashi. “Quantum Measurements for Hidden Subgroup Problems with Optimal Sample Complexity.” Quantum Information and Computation 8 (2008): 345–358. arXiv:quant-ph/0604174. https://arxiv.org/abs/quant-ph/0604174

Montanaro, Ashley. “Pretty Simple Bounds on Quantum State Discrimination.” arXiv:1908.08312, 2019. https://arxiv.org/abs/1908.08312

Cheng, Hao-Chung. “Simple and Tighter Derivation of Achievability for Classical Communication over Quantum Channels.” arXiv:2208.02132, 2022. https://arxiv.org/abs/2208.02132

Document information

Pretty-Good Measurement Bounds

Project	[QIT 002] State Distinguishability and Measurement Theorems
Document	Primary document
Author	mujirin
Verifier	Not verified
Downloaded	July 04, 2026 22:26 KST
Status	Working
Document link	https://www.theorytrace.com/projects/state-distinguishability-and-measurement-theorems/documents/untitled-document-d11bd3/