Title: State retrieval beyond Bayes' retrodiction

Speakers: Jacopo Surace

Series: Quantum Foundations

Date: January 12, 2023 - 11:00 AM

URL: https://pirsa.org/23010098

Abstract: In the context of irreversible dynamics, the meaning of the reverse of a physical evolution can be quite ambiguous. It is a standard choice to define the reverse process using Bayes' theorem, but, in general, this is not optimal with respect to the relative entropy of recovery. In this work we explore whether it is possible to characterise an optimal reverse map building from the concept of state retrieval maps. In doing so, we propose a set of principles that state retrieval maps should satisfy. We find out that the Bayes inspired reverse is just one case in a whole class of possible choices, which can be optimised to give a map retrieving the initial state more precisely than the Bayes rule. Our analysis has the advantage of naturally extending to the quantum regime. In fact, we find a class of reverse transformations containing the Petz recovery map as a particular case, corroborating its interpretation as a quantum analogue of the Bayes retrieval.

Finally, we present numerical evidence showing that by adding a single extra axiom one can isolate for classical dynamics the usual reverse process derived from Bayes' theorem.

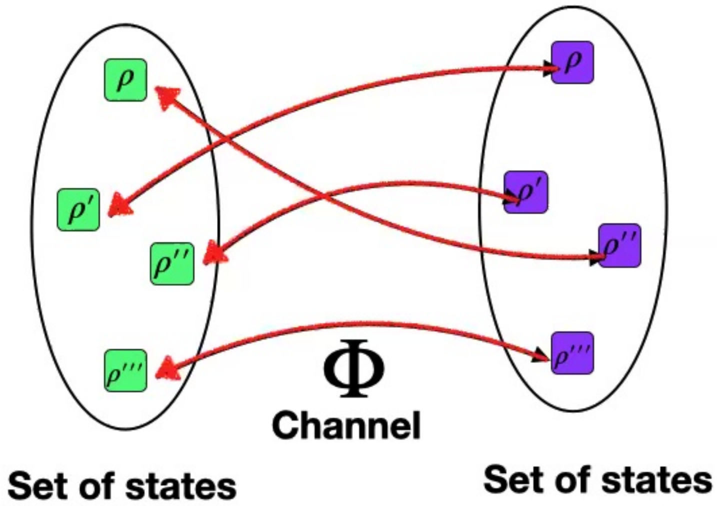
Zoom link: https://pitp.zoom.us/j/93589286500?pwd=dkZuRzR0SlhVd1lPdGNOZWFYQWtRZz09

State retrieval beyond Bayes' retrodiction and reverse processes

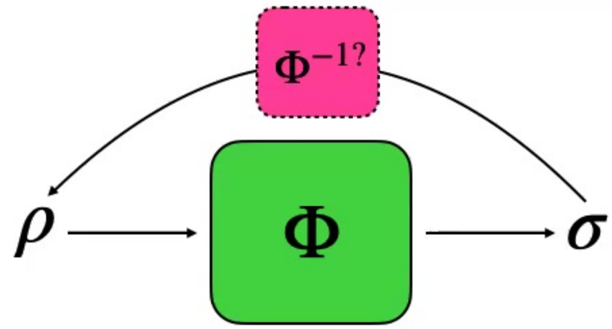
Jacopo Surace - Jan 2023

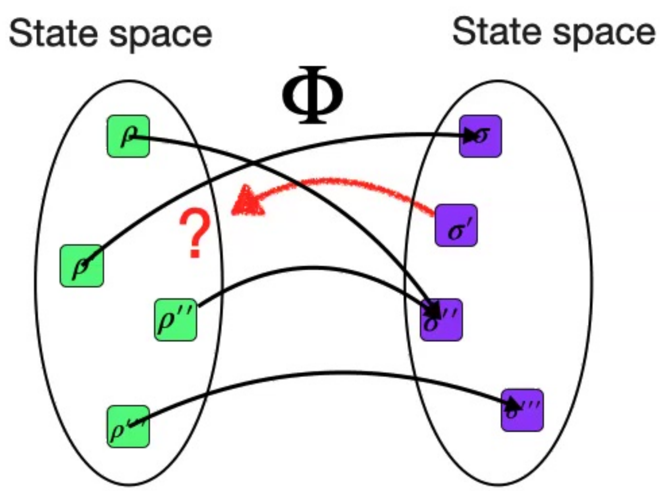
State retrieval beyond Bayes' retrodiction and reverse processes

Jacopo Surace - Jan 2023



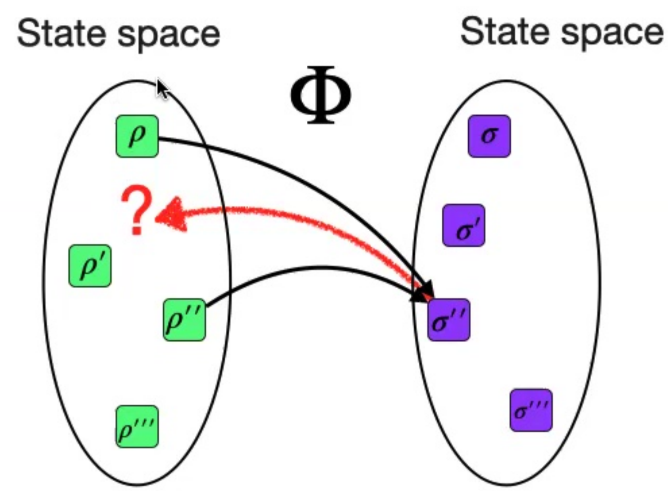
ρ
ρ'
ρ''
ρ'''
ρ
ρ'
ρ''
ρ'''
Φ
Channel
Set of states
Set of states



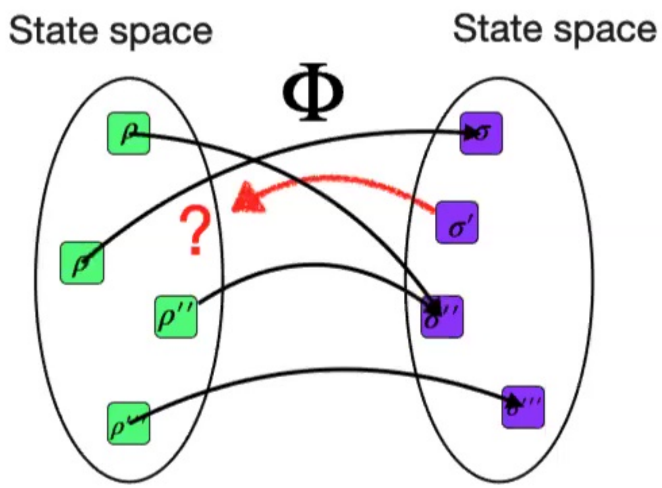


First Problem: not surjective

Well… we never reach that state in the image so it does not make sense to go back from there.

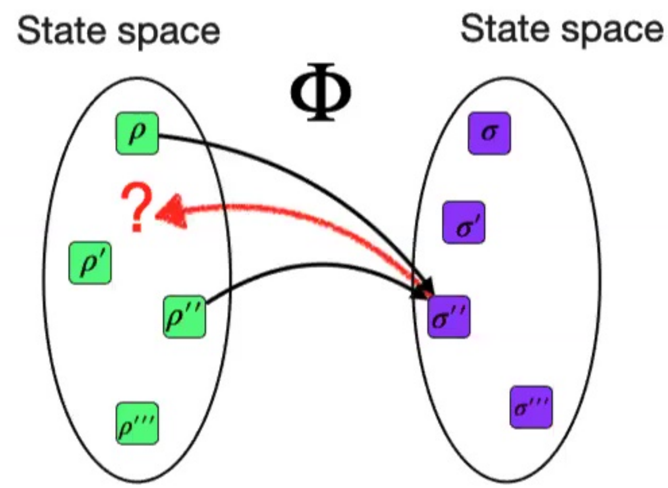


Second Problem: not injective



First Problem: not surjective

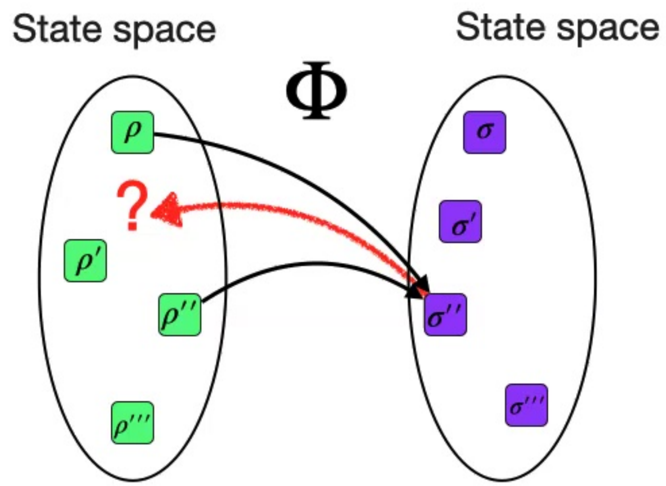
Well… we never reach that state in the image so it does not make sense to go back from there.



Second Problem: not injective

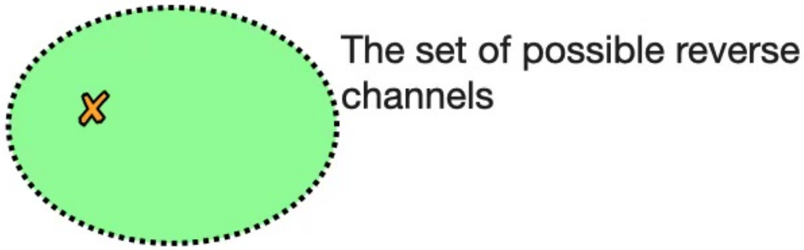
Reversing a channel is quite an ambiguous task.

Reversing a channel is quite an ambiguous task.



There are many possible choices for a reverse channel.

How to choose one?



Looking for a canonical reverse channel

Reverse channels are used already in many different field!
In thermodynamics are a fundamental tool for deriving fluctuation relations at the core of the discussion on the arrow of time.

The most commonly used reverse channel

Petz (recovery) **map**

The set of possible reverse channels.

There exists a common framework explaining while the Petz map should be THE reverse channel?

D. Petz, Sufficient subalgebras and the relative entropy of states of a von neumann algebra, Comm. Math. Phys. 105, 123 (1986).
D. Petz, Sufficiency of channels over von Neumann algebras, The Quarterly Journal of Mathematics 39, 97 (1988).

Looking for a canonical reverse channel

Satosi Watanabe

The problem of reversing a channel is the problem of **retrodicting** a state: inferring the original state from the knowledge of the channel, (possibly) some prior information and the evolved state.

Being at time t_1 you want to retrodict the state present at time $t_0 < t_1$ knowing that at time t_1 your state is σ.

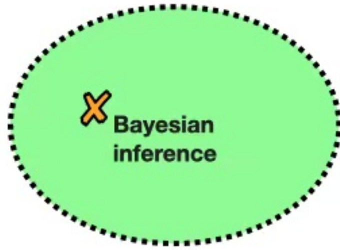
Channel: Φ
Evolved state: σ
Prior (Original Guess): π

→ Inference techniques → New Guess: π'

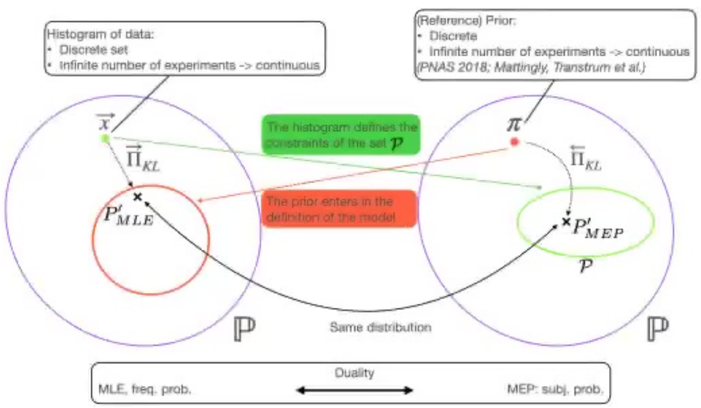
Canonical Statistical inference methods → **Petz** (recovery) **map** and **Bayes inspired reverse channel**

S. Watanabe, Symmetry of physical laws. part iii. prediction and retrodiction, Rev. Mod. Phys. 27, 179 (1955).
S. Watanabe, Conditional probabilities in physics, Progr. Theor. Phys. Suppl. E65, 135 (1965).
D. Petz, Sufficient subalgebras and the relative entropy of states of a von neumann algebra, Comm. Math. Phys. 105, 123 (1986).
D. Petz, Sufficiency of channels over von Neumann algebras, The Quarterly Journal of Mathematics 39, 97 (1988).
Buscemi, Francesco and Scarani, Valerio, Fluctuation Theorems from Bayesian Retrodiction, 10.1103/PhysRevE.103.052111
C. C. Aw, F. Buscemi, and V. Scarani, Fluctuation theorems with retrodiction rather than reverse processes, AVS Quantum Science 3, 045601 (2021), https://doi.org/10.1116/5.0060893.

Ambiguity again: why Bayesian inference?

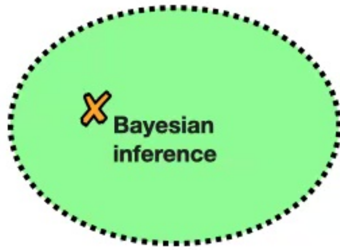


The set of possible inference methods.

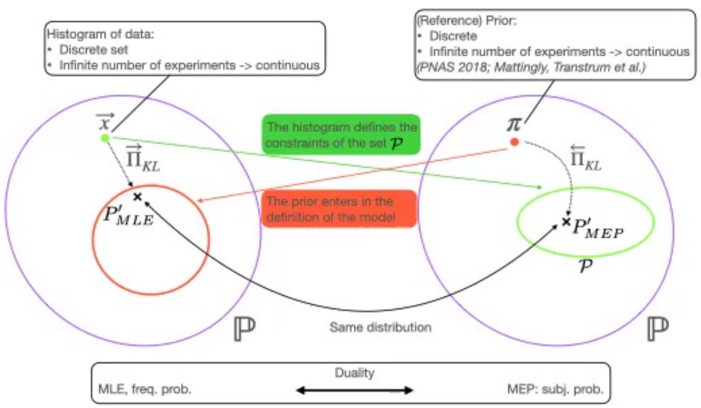


- Duality with maximum likelihood.
- Derivation from maximum entropy principles.
- Derivation from minimization of geometric distances (e.g. Kullback-Leibler).
- Derivation from principles of information goemetry (Amari)
- Derivation from geometric principles (Csizar).
- Consideration on properties of the convergence of subjective probability updates (e.g. Jeffrey, Bernando,…).
- …

Ambiguity again: why Bayesian inference?



The set of possible inference methods.



- Duality with maximum likelihood.
- Derivation from maximum entropy principles.
- Derivation from minimization of geometric distances (e.g. Kullback-Leibler).
- Derivation from principles of information goemetry (Amari)
- Derivation from geometric principles (Csizar).
- Consideration on properties of the convergence of subjective probability updates (e.g. Jeffrey, Bernando,…).
- …

There exists a common framework explaining why Bayesian inference should be THE inference method?

Why Bayesian inference?

Is it possible to characterise all the reasonable retrodiction channels?

What makes Bayes so fundamental?

Is it possible to find a better retrodiction channel?

Stochastic maps

$$\Phi = \begin{pmatrix} \phi_{1,1} & \phi_{1,2} & \cdots & \phi_{1,n} \\ \phi_{2,1} & \phi_{2,2} & \cdots & \phi_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{n,1} & \phi_{n,2} & \cdots & \phi_{n,n} \end{pmatrix}$$

$$\forall k \; \sum_{i=1}^{n} \phi_{i,k}$$

$$\forall k, i \; \phi_{i,k} \in [0,1]$$

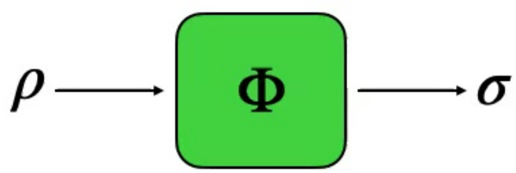
Left stochastic map

$$\rho = (\rho_1, \rho_2, \ldots, \rho_n)$$

$$\sum_{i=1}^{n} \rho_i = 1$$

$$\forall i \; \rho_i \in [0,1]$$

Probability vector



$$\sigma = \Phi \rho$$

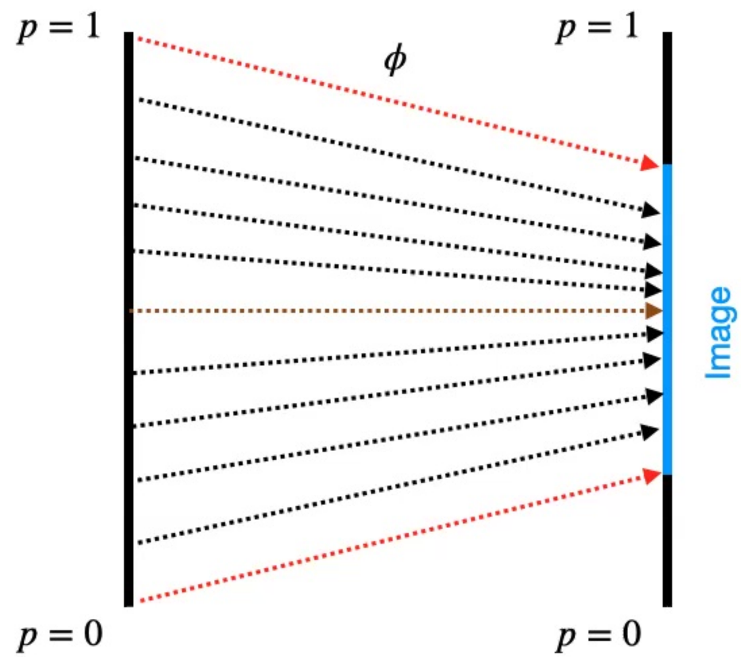
Example in 2 dimensions

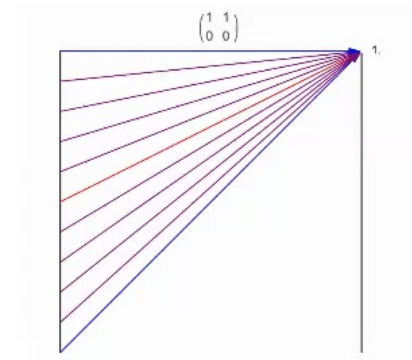
Stochastic map

$$\Phi = \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}$$

Probability vector

$$\rho = (p, 1-p) \quad p \in [0,1]$$





In the case of stochastic maps, states are probability vectors.

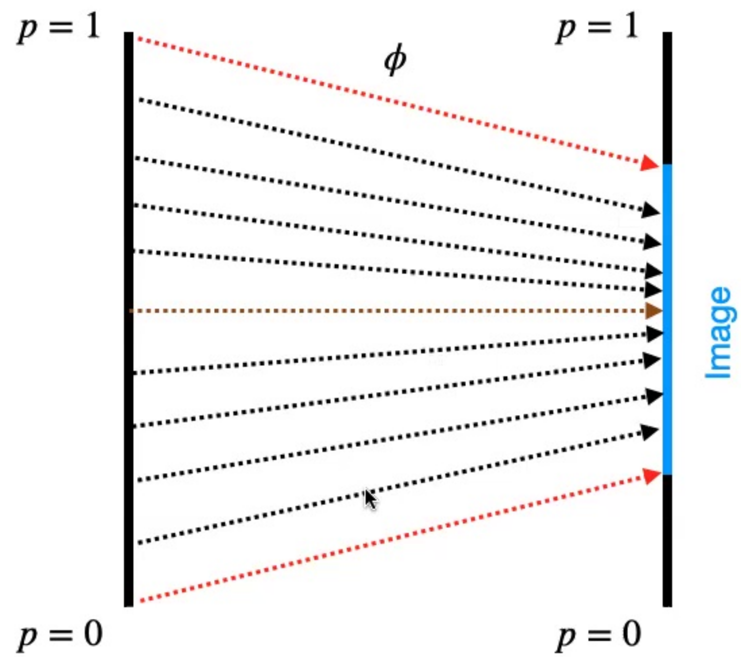
Example in 2 dimensions

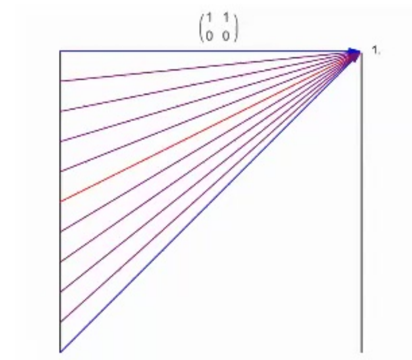
Stochastic map

$$\Phi = \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}$$

Probability vector

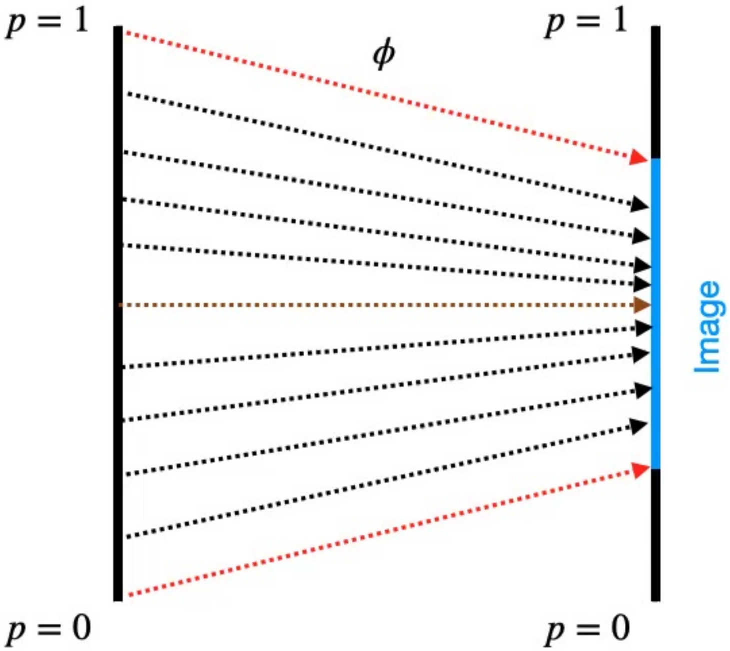
$$\rho = (p, 1-p) \quad p \in [0,1]$$



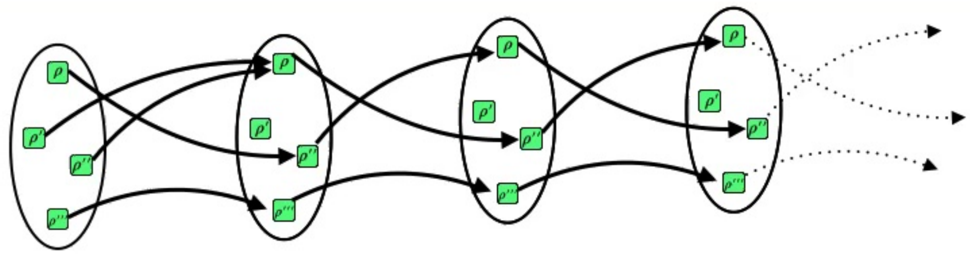


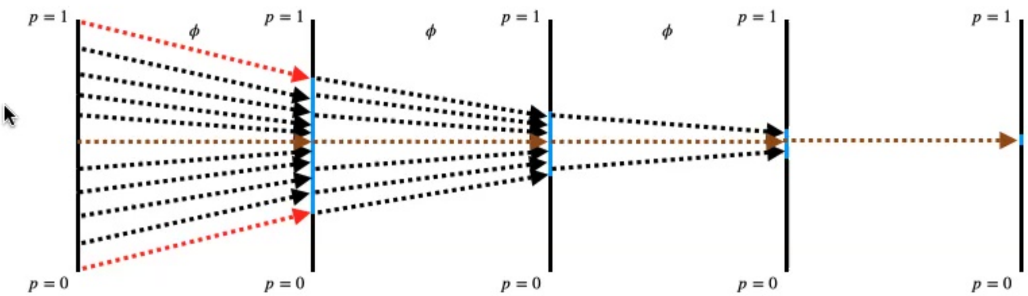
In the case of stochastic maps, states are probability vectors.

Loss of information for stochastic maps: contractivity



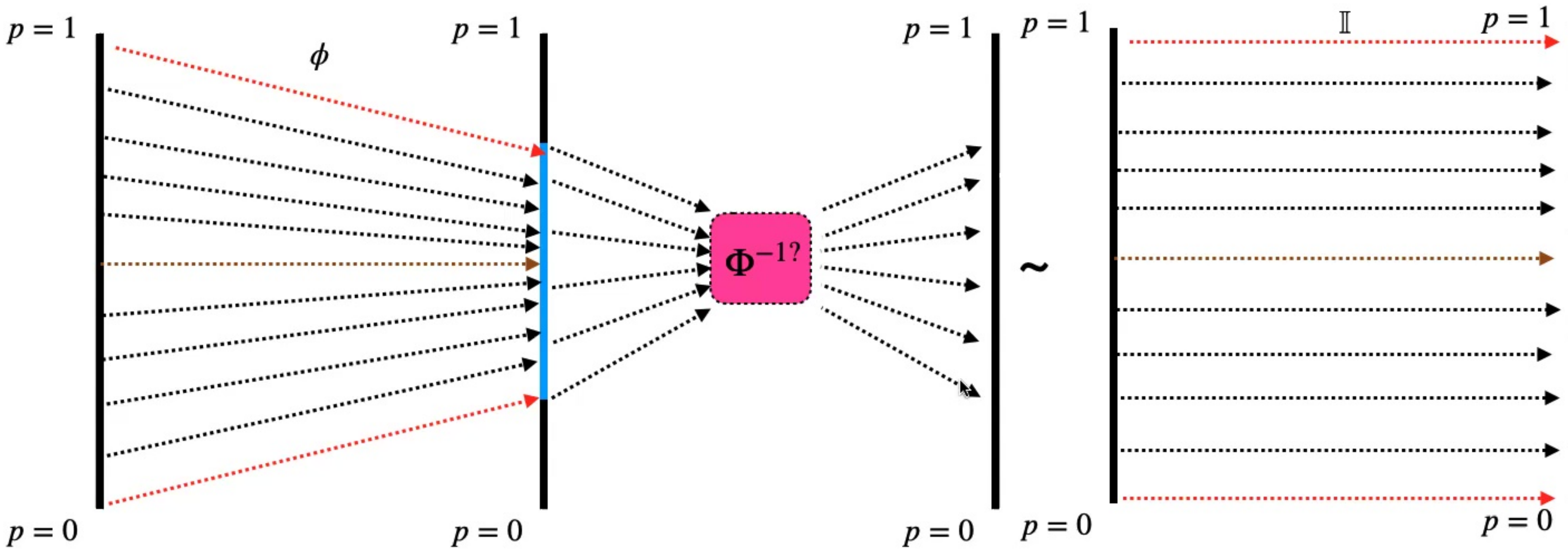
In the continuous case we find a new problem for the reversibility: channels are **contractive**.





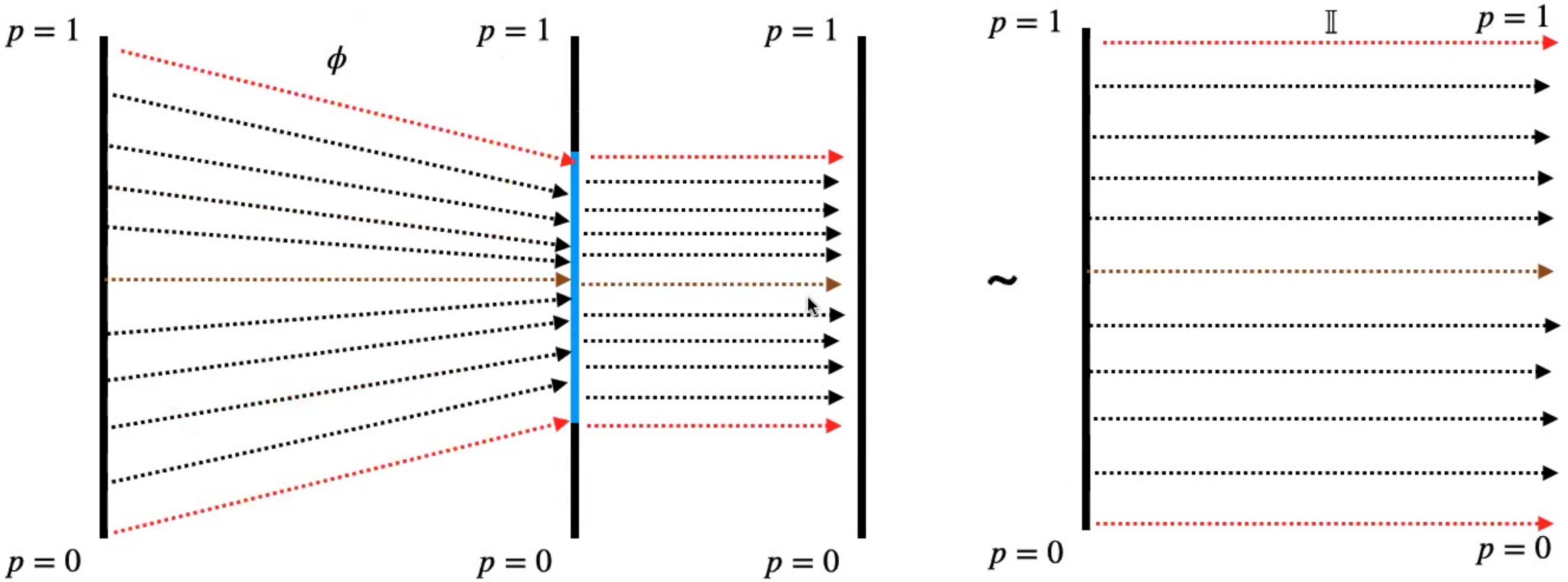
Contractivity is a property of channels. **Channels never expand.**

Reversing a channel

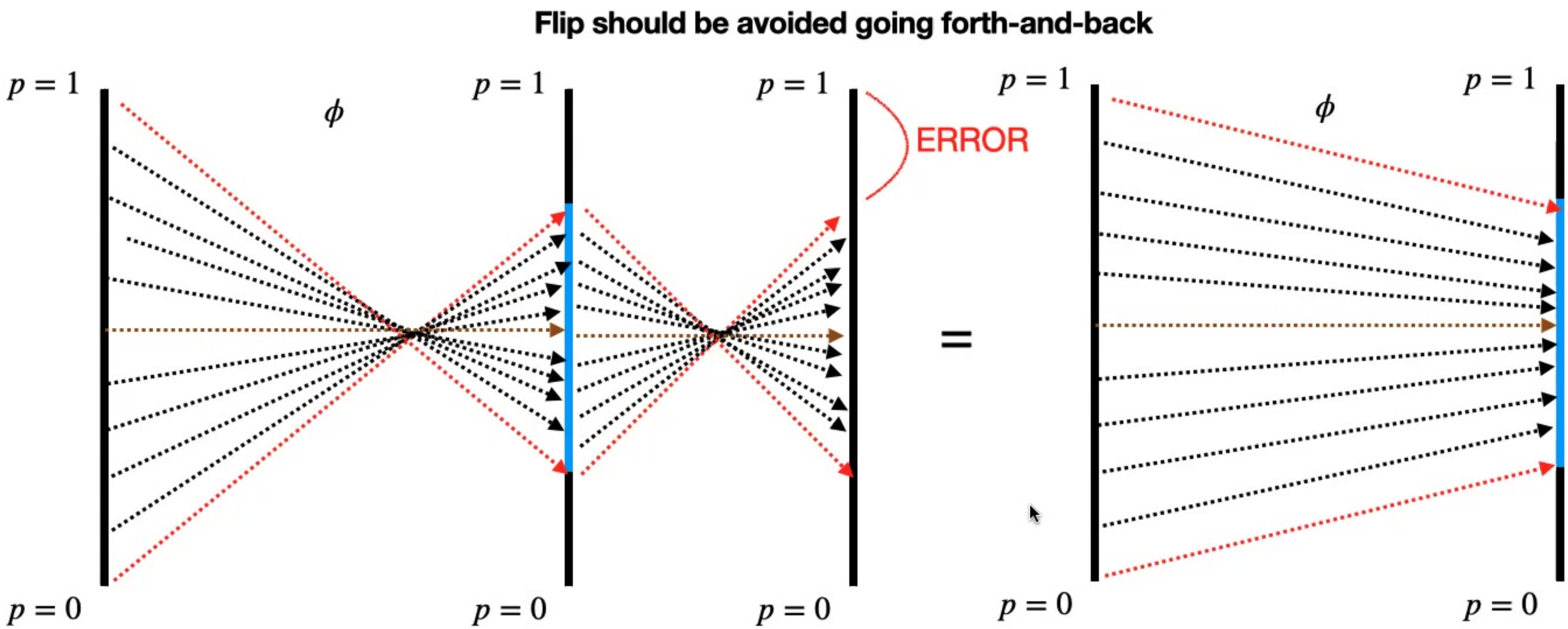


The central object to study is the forth-and-back channel

Reversing a channel

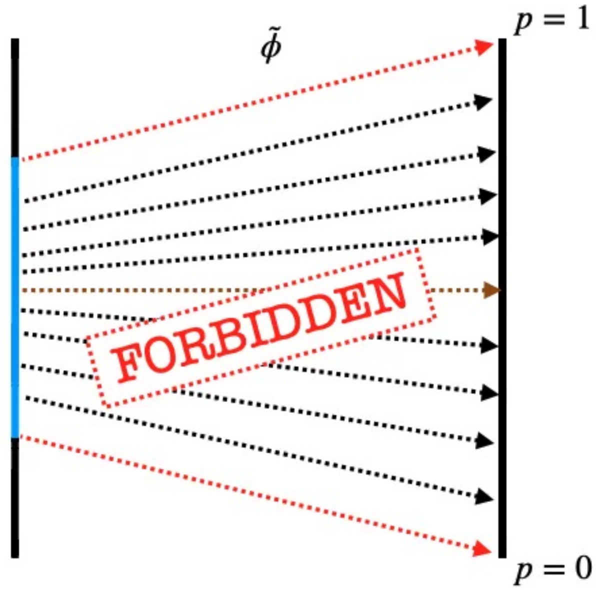


Reversing a channel



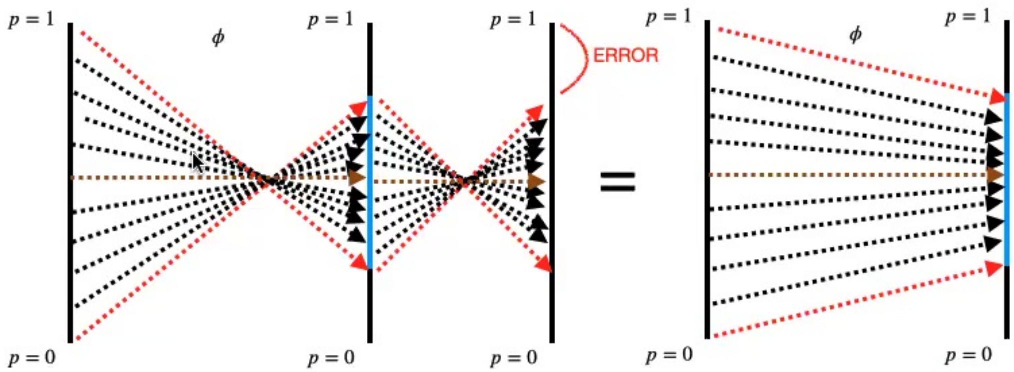
2. **The state retrieval channel should be physical.**

"It should be a meaningful retrieval map even in the single-shot scenario, not just in the full statistics case".

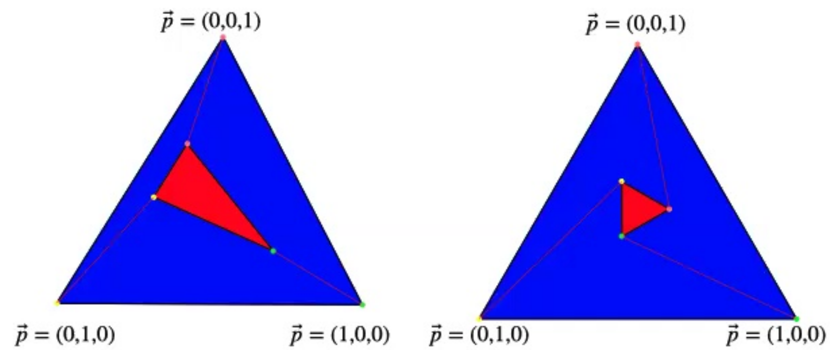


3. All the eigenvalues of the back and forth map must be positive.

"Every inversion (negative eigenvalues) or rotation (complex eigenvalues) ruins the retrieval."



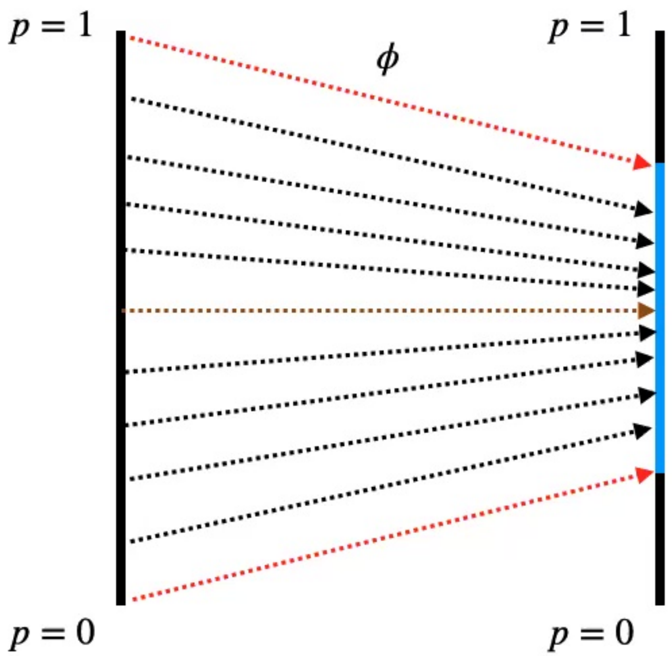
Flips

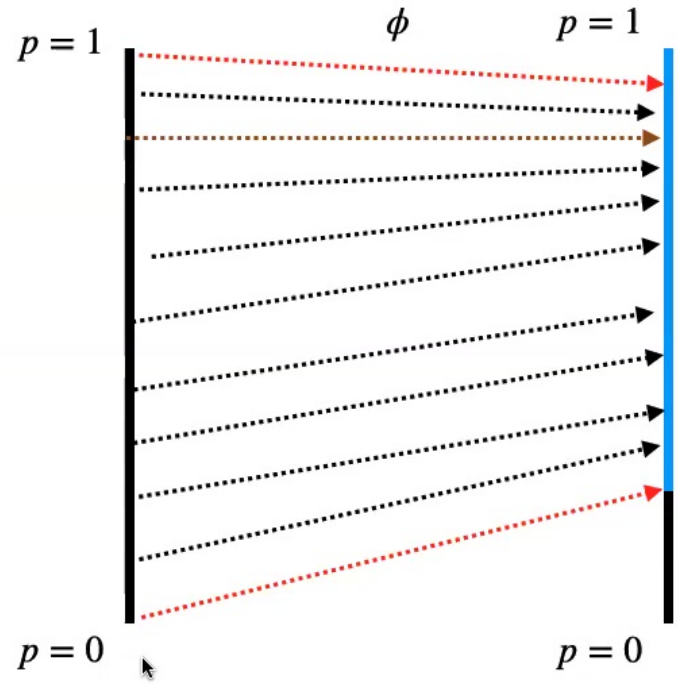


Rotations

An additional property of channels

Every channels always preserves at least one vector or state, they have at least one fixed point.





We want to exploit this property.

Stochastic channel:	Φ
Prior:	π

Basic ingredients

5. **The prior should be an equilibrium state for the back-and-forth channel.** *“The back-and-forth channel should be time symmetric on the fixed point”.*

Taming the ambiguity: the space of retrieval channels

Stochastic channel:	Φ
Prior:	π

Basic ingredients

$\tilde{\Phi}$	Must be a (left-)stochastic matrix
$\tilde{\Phi} \circ \Phi$	The back-and-forth channel must have positive eigenvalues
$\tilde{\Phi} : \Phi\pi \to \pi$	$\tilde{\Phi}$ have this transition fixed.
$(\tilde{\Phi}\Phi)_{j,i}\pi_i = (\tilde{\Phi}\Phi)_{i,j}\pi_j$	The prior state is the equilibrium state for the back-and-forth channel

- Convex Set
- Finite set of vertices
- Algorithm for computing vertices by Jurkat and Ryser

- The forth-and-back channel is a positive semidefinite matrix

Taming the ambiguity: the space of retrieval channels

Stochastic channel: Φ

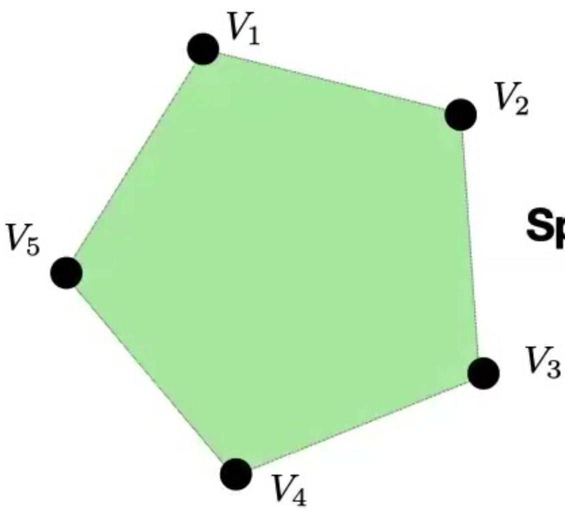
Prior: π

Basic ingredients

$\tilde{\Phi}$	Must be a (left-)stochastic matrix
$\tilde{\Phi} \circ \Phi$	The back-and-forth channel must have positive eigenvalues
$\tilde{\Phi} : \Phi\pi \to \pi$	$\tilde{\Phi}$ have this transition fixed.
$(\tilde{\Phi}\Phi)_{j,i}\pi_i = (\tilde{\Phi}\Phi)_{i,j}\pi_j$	The prior state is the equilibrium state for the back-and-forth channel

- Convex Set
- Finite set of vertices
- Algorithm for computing vertices by Jurkat and Ryser

- The forth-and-back channel is a positive semidefinite matrix



Space of retrieval channels

Taming the ambiguity: the space of retrieval channels

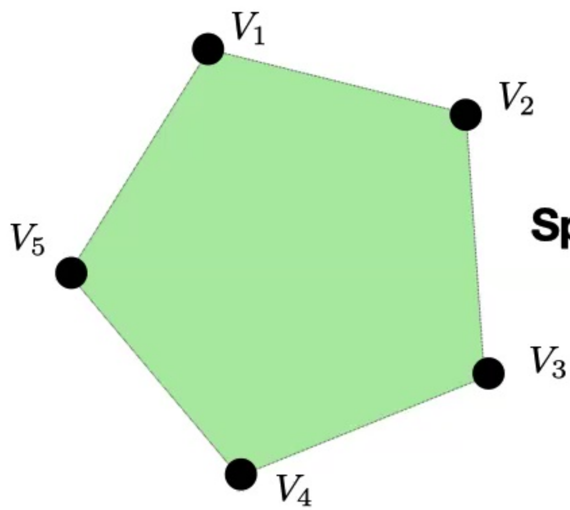
Stochastic channel:	Φ
Prior:	π

Basic ingredients

$\tilde{\Phi}$	Must be a (left-)stochastic matrix
$\tilde{\Phi} \circ \Phi$	The back-and-forth channel must have positive eigenvalues
$\tilde{\Phi} : \Phi\pi \to \pi$	$\tilde{\Phi}$ have this transition fixed.
$(\tilde{\Phi}\Phi)_{j,i}\pi_i = (\tilde{\Phi}\Phi)_{i,j}\pi_j$	The prior state is the equilibrium state for the back-and-forth channel

- Convex Set
- Finite set of vertices
- Algorithm for computing vertices by Jurkat and Ryser

- The forth-and-back channel is a positive semidefinite matrix



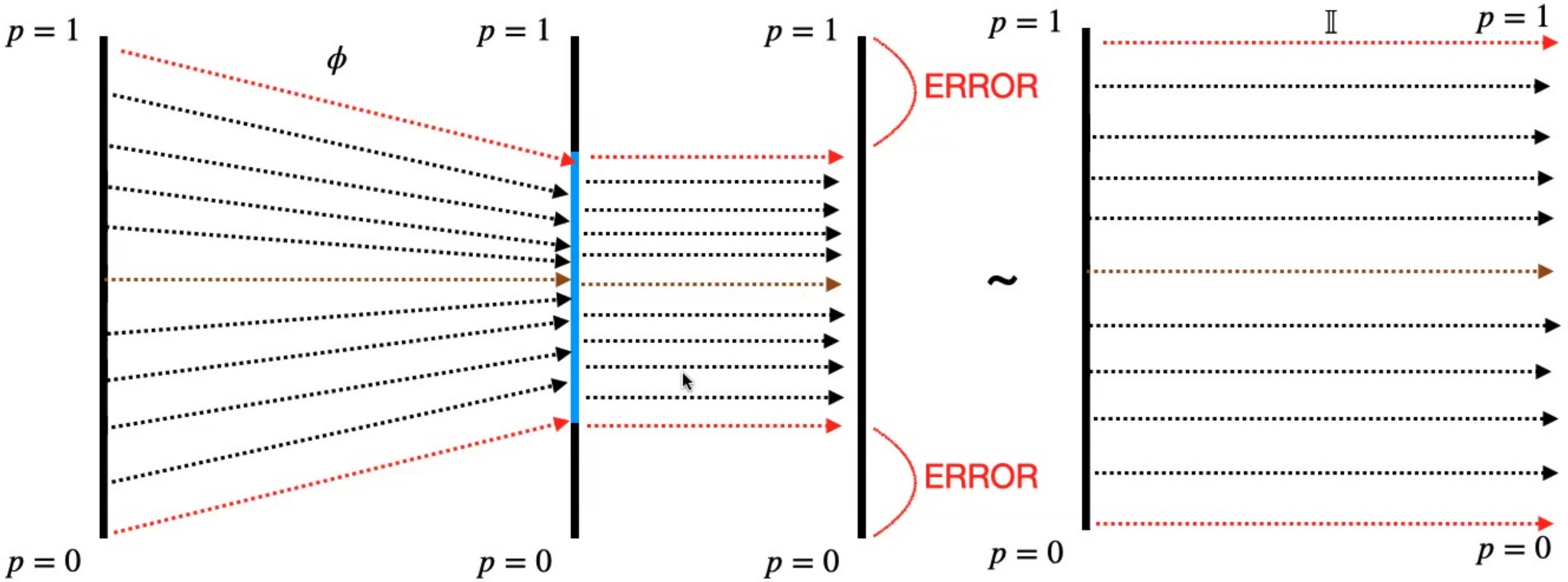
Space of retrieval channels

Each reverse channel is completely characterised by the vector of coefficients.
This is a probability vector.
Each reverse channel is characterised by a probability vector.

$$\tilde{\Phi} \leftrightarrow \vec{\lambda}^{\tilde{\Phi}}$$

Reversing a channel

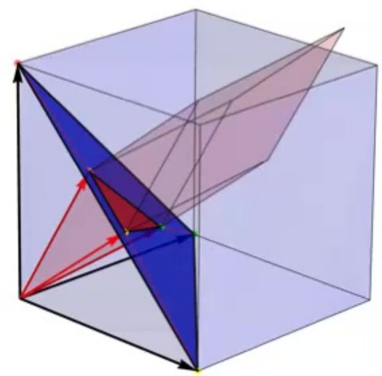
IDEA: Contract least possible



IDEA: minimal contraction

Optimisation criterion: **The optimal retrieval map is the one that maximise the determinant of the forth-and-back channel.**

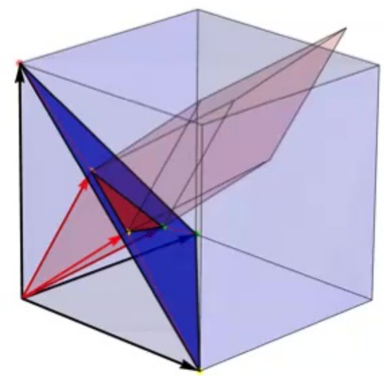
$$\tilde{\Phi}_O = \max_{\tilde{\Phi}\text{ state retrieval}} \det \tilde{\Phi}\Phi$$



IDEA: minimal contraction

Optimisation criterion: **The optimal retrieval map is the one that maximise the determinant of the forth-and-back channel.**

$$\tilde{\Phi}_O = \max_{\tilde{\Phi}\ \text{state retrieval}} \det \tilde{\Phi}\Phi$$



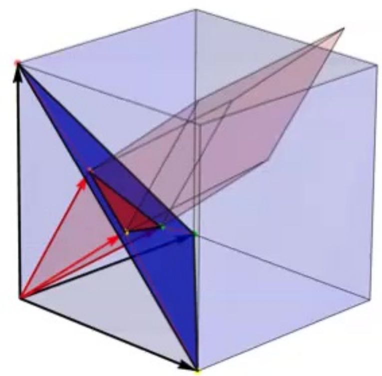
Second reason for choosing this optimisation criterion

$$D(\tilde{\Phi}\Phi \| \mathbb{1}) = \mathrm{Tr}[\mathbb{1}(\log \mathbb{1} - \log \tilde{\Phi}\Phi)] =$$
$$= -\,\mathrm{Tr}[\log \tilde{\Phi}\Phi] = \log \det(\tilde{\Phi}\Phi)^{-1}$$

IDEA: minimal contraction

Optimisation criterion: **The optimal retrieval map is the one that maximise the determinant of the forth-and-back channel.**

$$\tilde{\Phi}_O = \max_{\tilde{\Phi}\ \text{state retrieval}} \det \tilde{\Phi}\Phi$$



Second reason for choosing this optimisation criterion

$$D(\tilde{\Phi}\Phi || \mathbb{1}) = \mathrm{Tr}[\mathbb{1}(\log \mathbb{1} - \log \tilde{\Phi}\Phi)] =$$
$$= -\mathrm{Tr}[\log \tilde{\Phi}\Phi] = \log \det(\tilde{\Phi}\Phi)^{-1}$$

Practical reason

Encyclopedia of Optimization pp 3375–3380 | Cite as

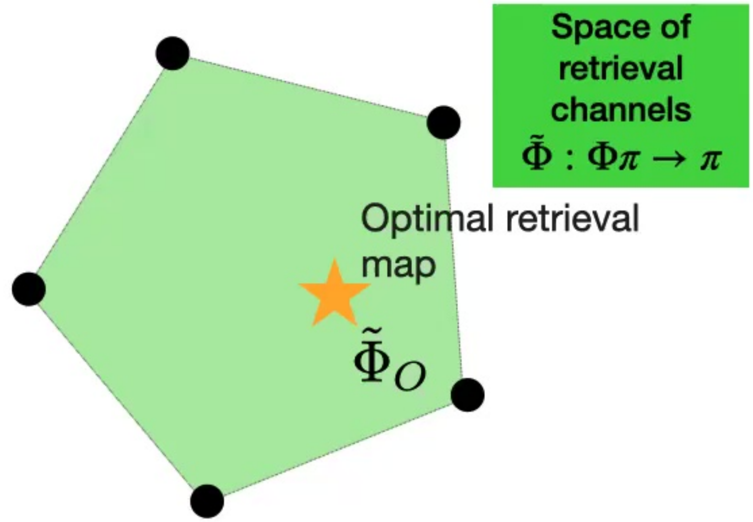
Semidefinite Programming and Determinant Maximization

Lieven Vandenberghe, Stephen Boyd & Shao-Po Wu

Stochastic channel: Φ

Prior: π

Basic ingredients

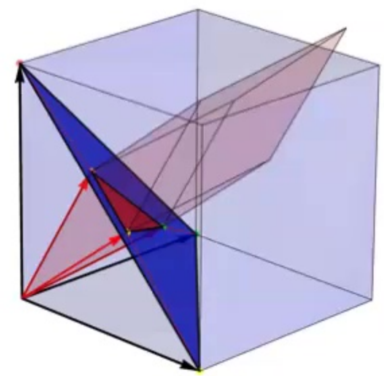


The optimal retrieval map is found!

IDEA: minimal contraction

Optimisation criterion: **The optimal retrieval map is the one that maximise the determinant of the forth-and-back channel.**

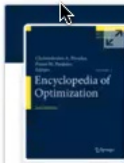
$$\tilde{\Phi}_O = \max_{\tilde{\Phi} \text{ state retrieval}} \det \tilde{\Phi}\Phi$$



Second reason for choosing this optimisation criterion

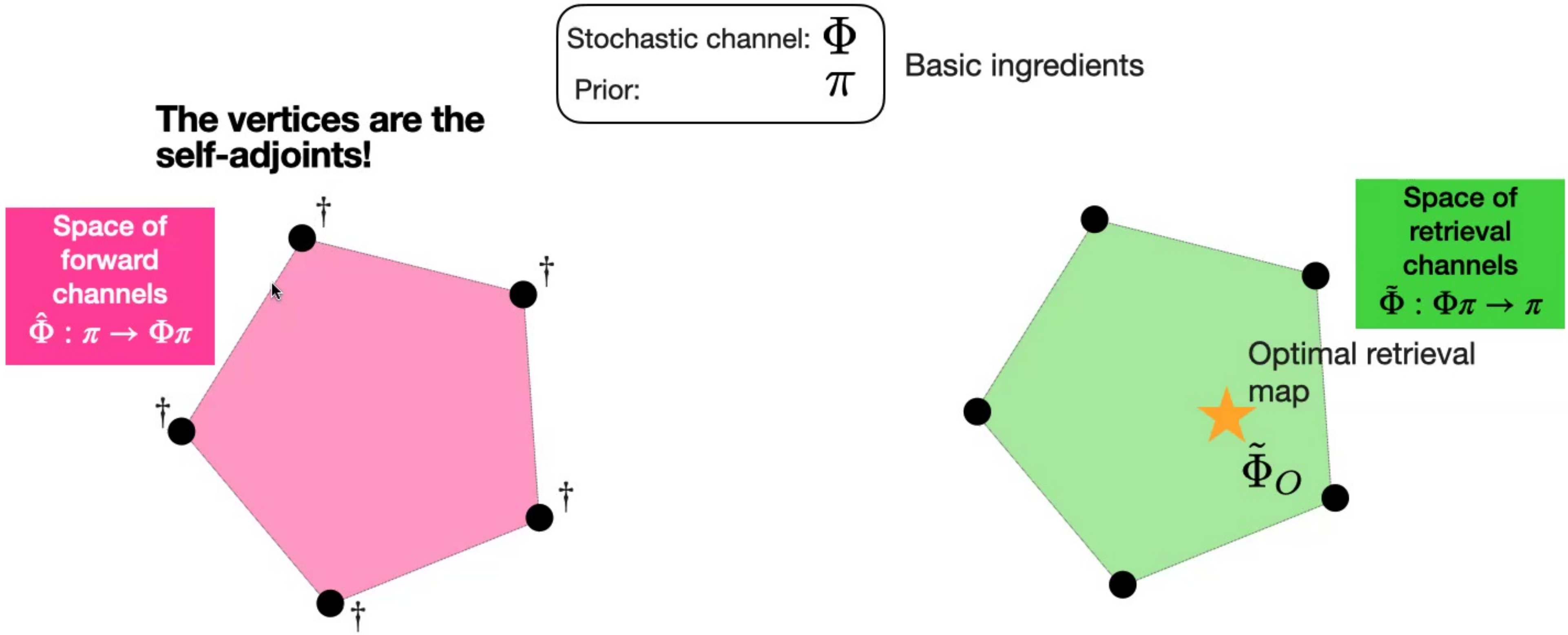
$$D(\tilde{\Phi}\Phi || \mathbb{1}) = \mathrm{Tr}[\mathbb{1}(\log \mathbb{1} - \log \tilde{\Phi}\Phi)] =$$
$$= -\mathrm{Tr}[\log \tilde{\Phi}\Phi] = \log \det(\tilde{\Phi}\Phi)^{-1}$$

Practical reason


Encyclopedia of Optimization pp 3375–3380 | Cite as

Semidefinite Programming and Determinant Maximization

Lieven Vandenberghe, Stephen Boyd & Shao-Po Wu

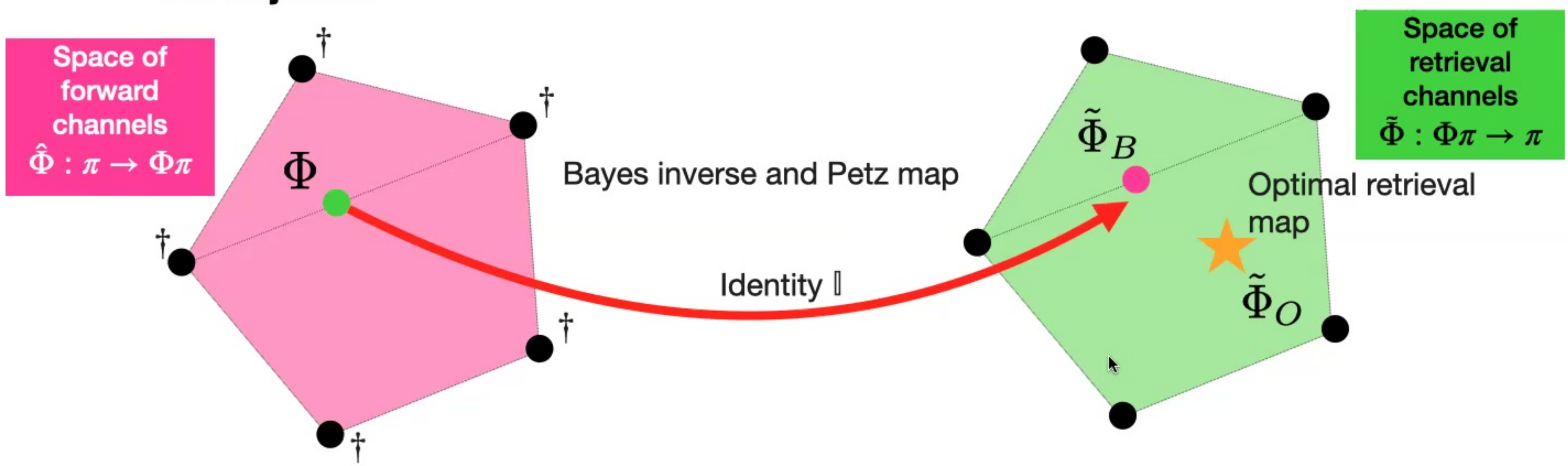

Stochastic channel: Φ
Prior: π
Basic ingredients
The vertices are the self-adjoints!
Space of forward channels $\hat{\Phi} : \pi \to \Phi\pi$
†
†
†
†
†
Space of retrieval channels $\tilde{\Phi} : \Phi\pi \to \pi$
Optimal retrieval map
$\tilde{\Phi}_O$

Stochastic channel: Φ

Prior: π

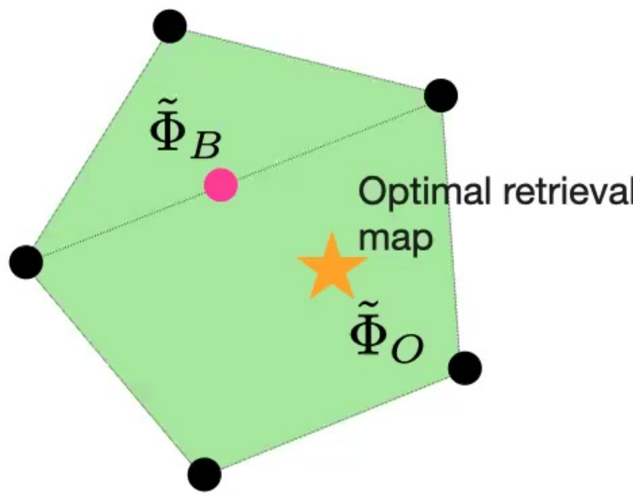
Basic ingredients

The vertices are the self-adjoints!



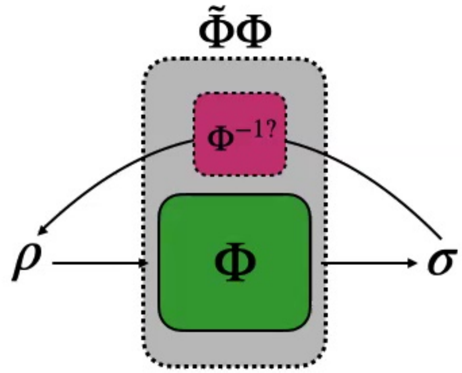
$\Phi \leftrightarrow \vec{\lambda}^{\Phi}$ → $\tilde{\Phi} \leftrightarrow \vec{\lambda}^{\tilde{\Phi}} = \mathbb{I}\vec{\lambda}^{\Phi} = \vec{\lambda}^{\Phi}$

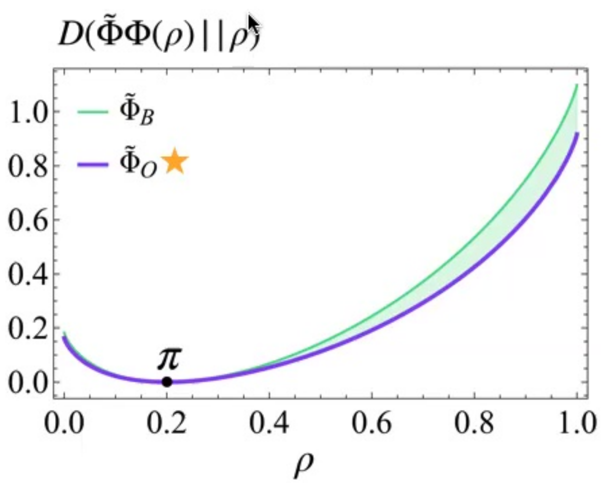
Reversion as a linear transformation between these two spaces

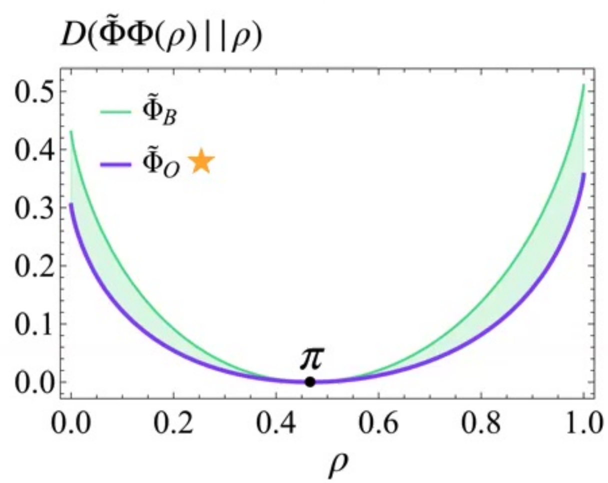


Bayes inverse and the optimal retrieval map are not the same!
Petz map and the optimal retrieval map are not the same!

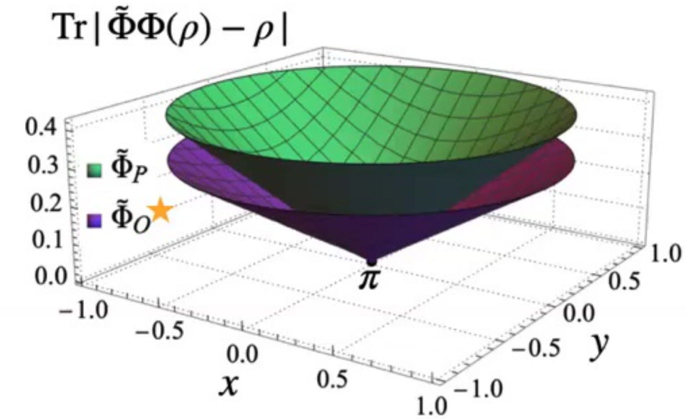
Comparison of the state retrieval with Bayes and Petz







Stochastic maps and Bayes



Quantum channels and Petz

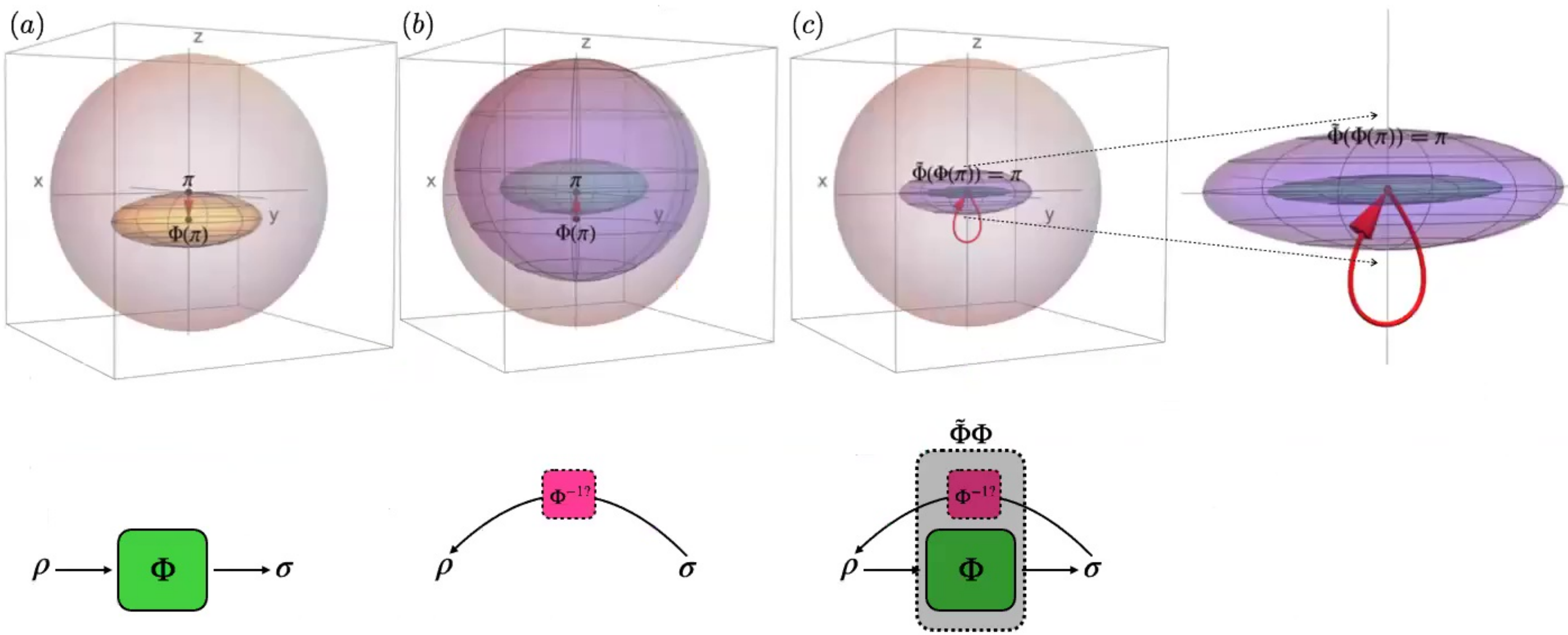
Comparison of the optimal state retrieval with Petz

Stochastic channel: Φ
Prior: π

Basic ingredients

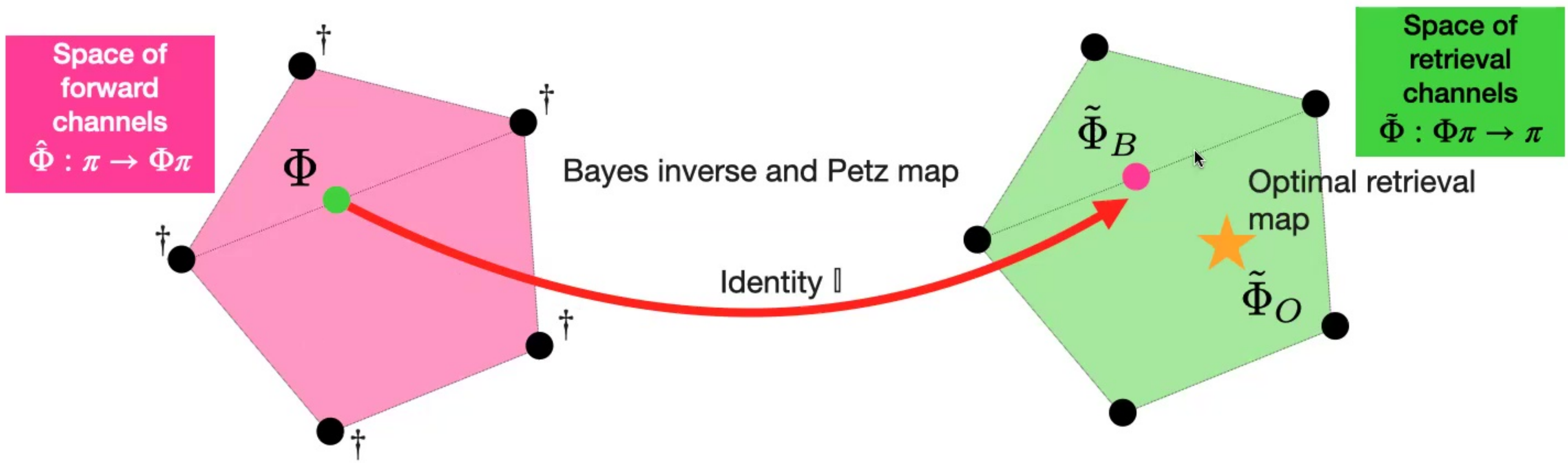
$\tilde{\Phi}_P$

$\tilde{\Phi}_O$ ★

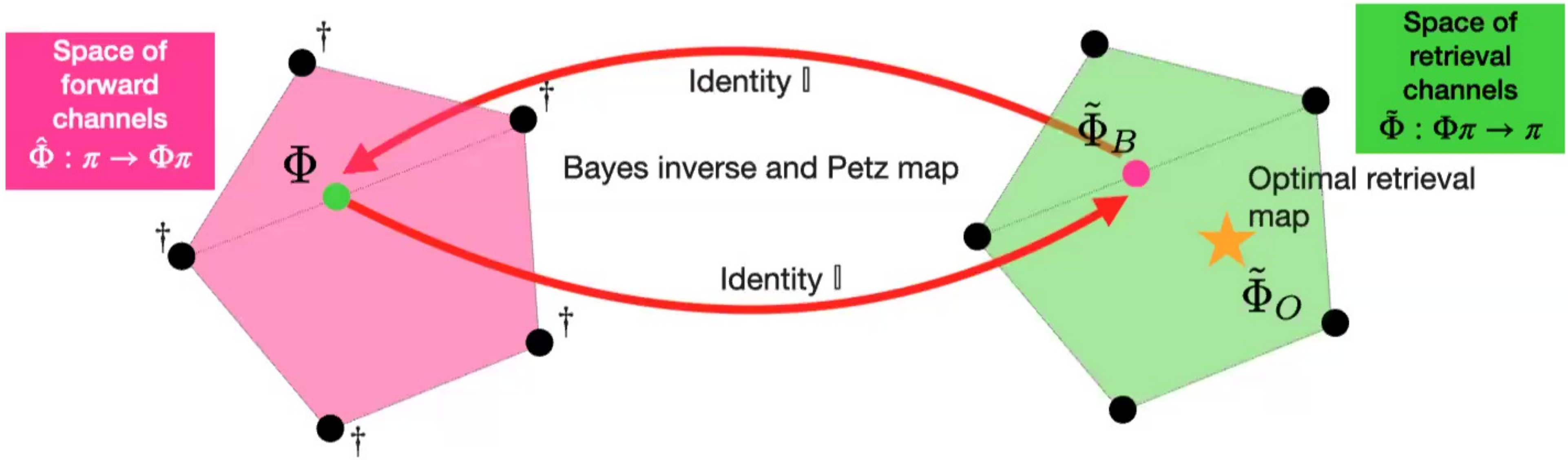


Where does Bayes come from?

Can we add a property that shrinks the whole space of retrieval maps to a single point?



Simplicity is a nice criterion



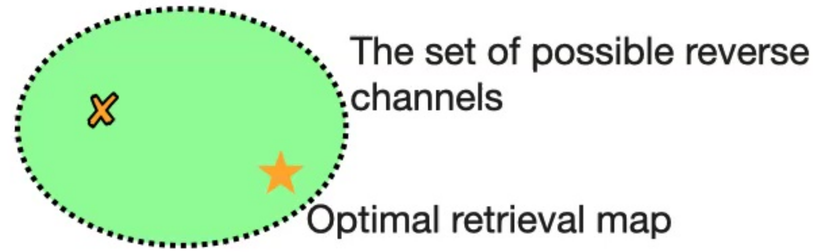
Can we add a property that isolates Bayes and Petz?

Maybe.

Numerical evidence of a sufficient 6th property (involutivity) that isolates Bayes and Petz.

Conclusions

- We dealt with the ambiguity in defining a reverse channel by characterising a set of admissible retrieval channels.
- We gave a computable criterion for choosing the optimal retrieval channel.
- The “canonical” reverse channel (Bayes inspired) is an admissible retrieval channels.
- The Bayes and Petz maps maybe can be isolated by asking for involutivity.



https://arxiv.org/abs/2201.09899

Thank you

Exploiting our prior knowledge, last 2 principles for a reverse channel

Taming the ambiguity: the space of retrieval channels

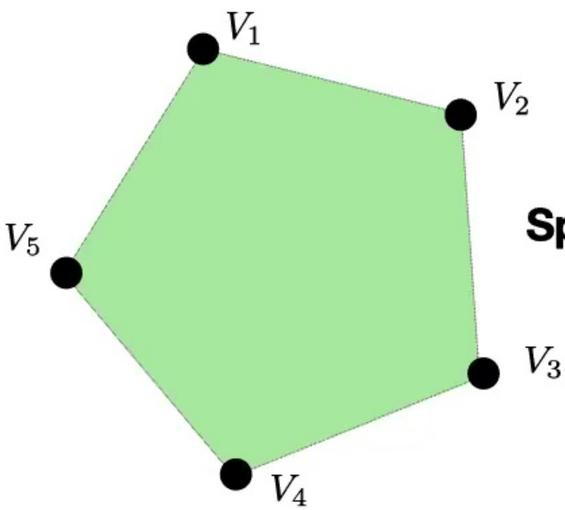
Stochastic channel:	Φ
Prior:	π

Basic ingredients

$\tilde{\Phi}$	Must be a (left-)stochastic matrix
$\tilde{\Phi} \circ \Phi$	The back-and-forth channel must have positive eigenvalues
$\tilde{\Phi} : \Phi\pi \to \pi$	$\tilde{\Phi}$ have this transition fixed.
$(\tilde{\Phi}\Phi)_{j,i}\pi_i = (\tilde{\Phi}\Phi)_{i,j}\pi_j$	The prior state is the equilibrium state for the back-and-forth channel

- Convex Set
- Finite set of vertices
- Algorithm for computing vertices by Jurkat and Ryser

- The forth-and-back channel is a positive semidefinite matrix



Space of retrieval channels

Each reverse channel is completely characterised by the vector of coefficients.
This is a probability vector.
Each reverse channel is characterised by a probability vector.

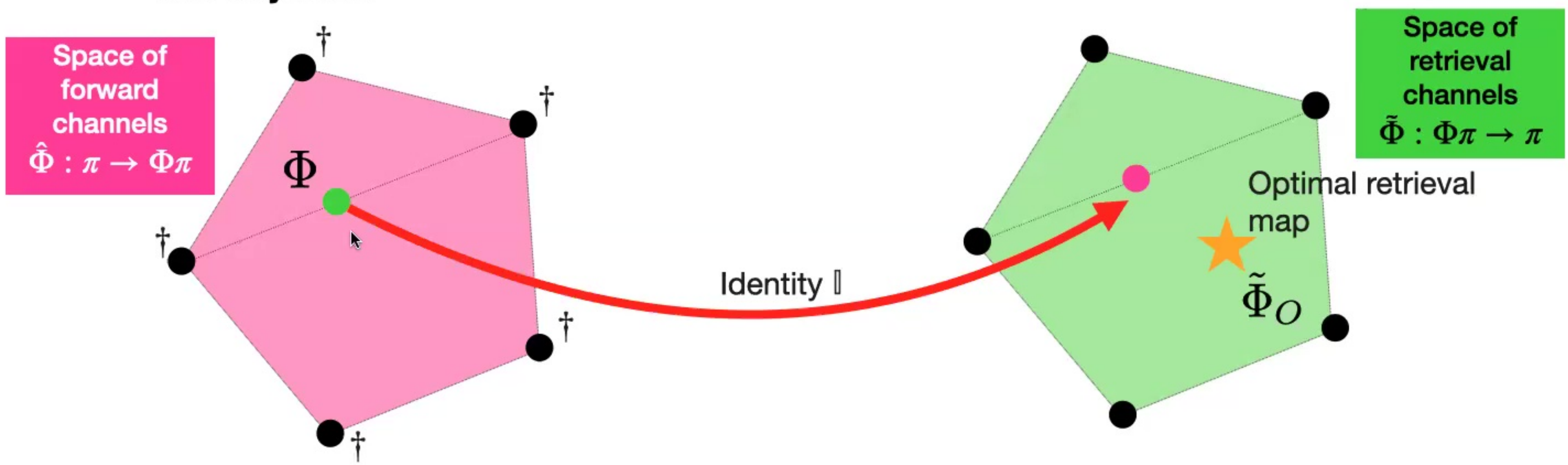
$$\tilde{\Phi} \leftrightarrow \vec{\lambda}^{\tilde{\Phi}}$$

Stochastic channel: Φ

Prior: π

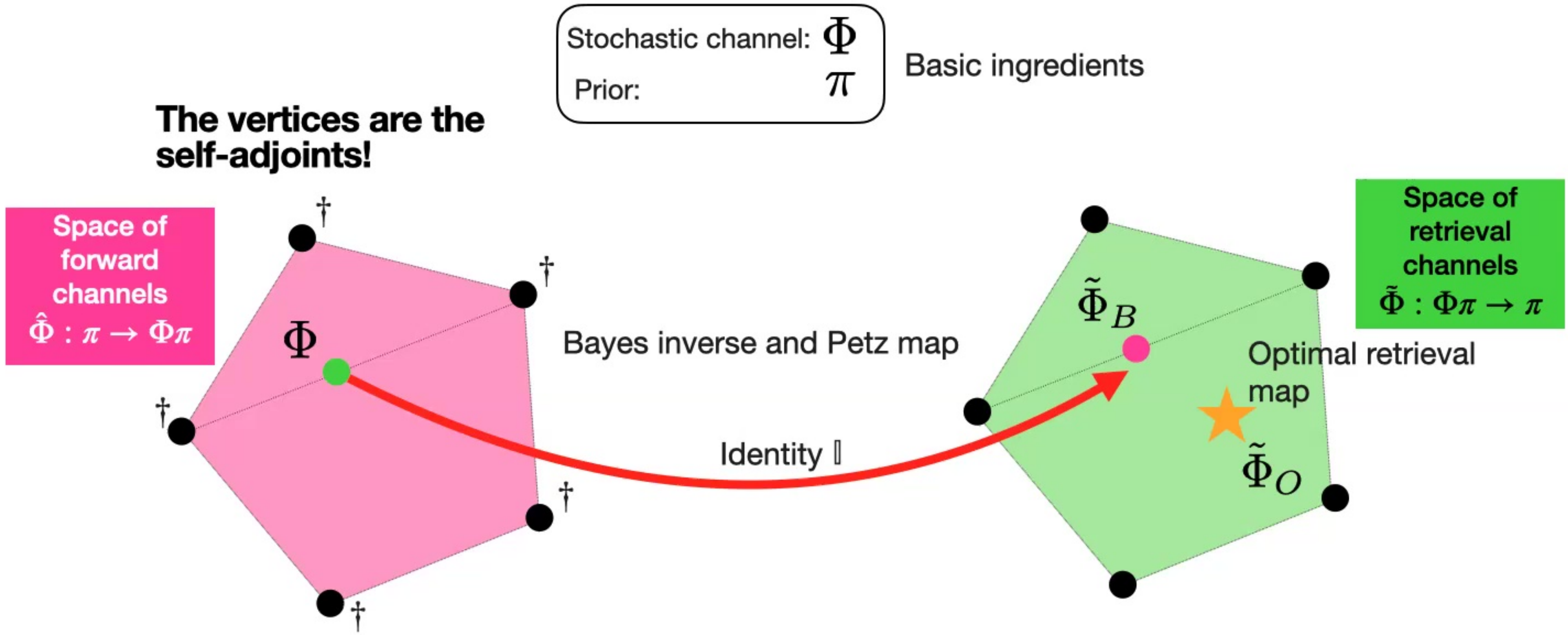
Basic ingredients

The vertices are the self-adjoints!



$$\Phi \leftrightarrow \vec{\lambda}^{\Phi}$$

Reversion as a linear transformation between these two spaces

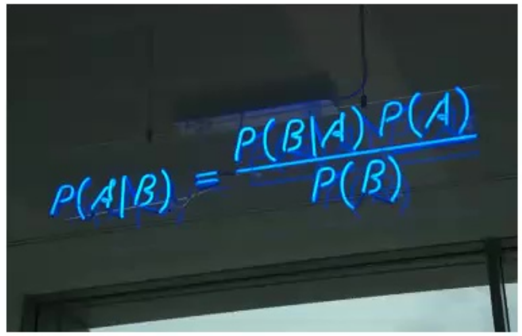


$$\Phi \leftrightarrow \vec{\lambda}^{\Phi} \longrightarrow \tilde{\Phi} \leftrightarrow \vec{\lambda}^{\tilde{\Phi}} = \mathbb{I}\vec{\lambda}^{\Phi} = \vec{\lambda}^{\Phi}$$

Reversion as a linear transformation between these two spaces

Bayesian inference: minimal working knowledge

Bayes Theorem



Bayes' theorem spelt out in blue neon at the offices of Autonomy in Cambridge.

Step 1.

Model: $P(x \mid \theta)$

Prior: $\hat{P}(\theta)$

Basic ingredients

Step 2.

$\tilde{x}$

Receive a piece of information

Step 3.

$$\hat{P}_{updated}(\theta) = \tilde{P}(\theta \mid x) = \frac{P(x \mid \theta)\hat{P}(\theta)}{\sum_\theta P(x \mid \theta)\hat{P}(\theta)}$$

Update your knowledge on the parameter

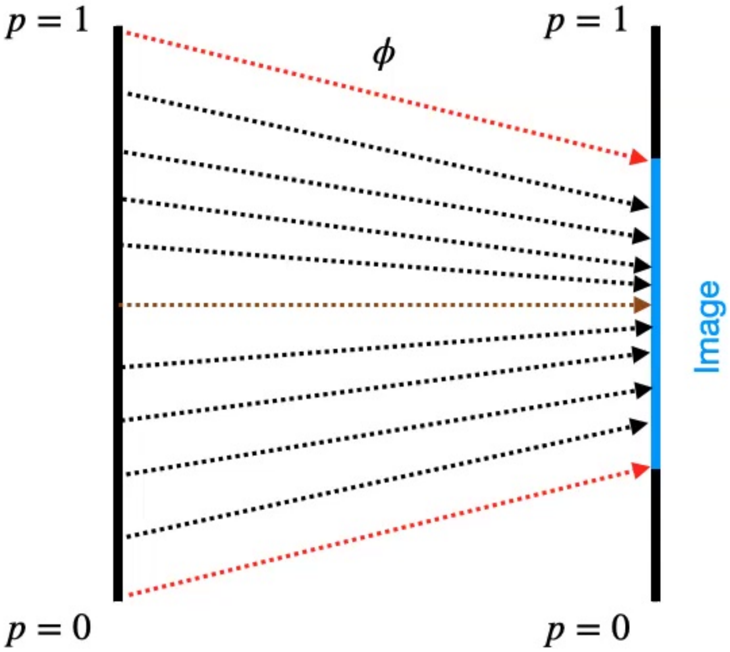
Step 1 updated

Model: $P(x \mid \theta)$

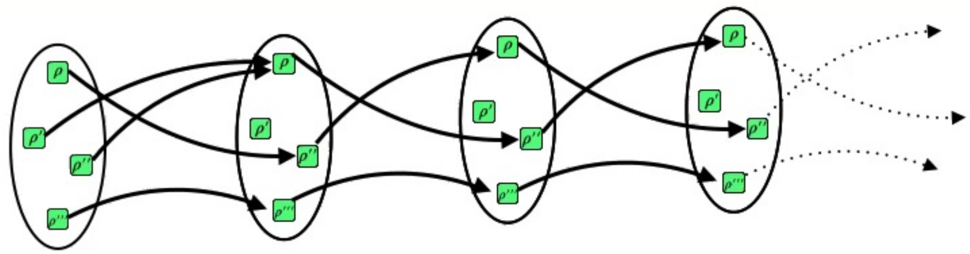
Prior: $\hat{P}_{updated}(\theta)$

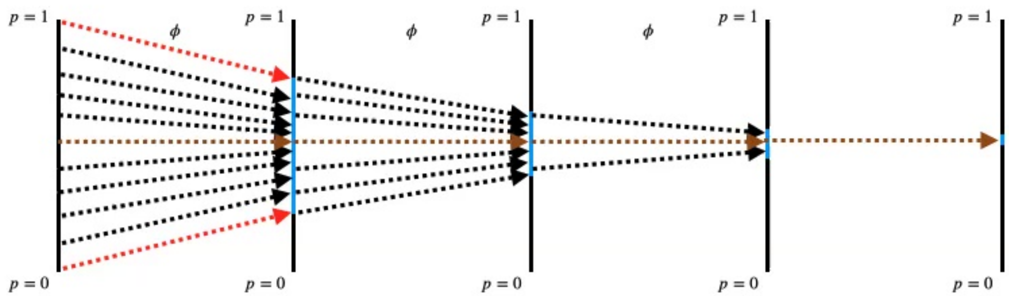
Basic ingredients updated

Loss of information for stochastic maps: contractivity



In the continuous case we find a new problem for the reversibility: channels are **contractive**.





Contractivity is a property of channels. **Channels never expand.**

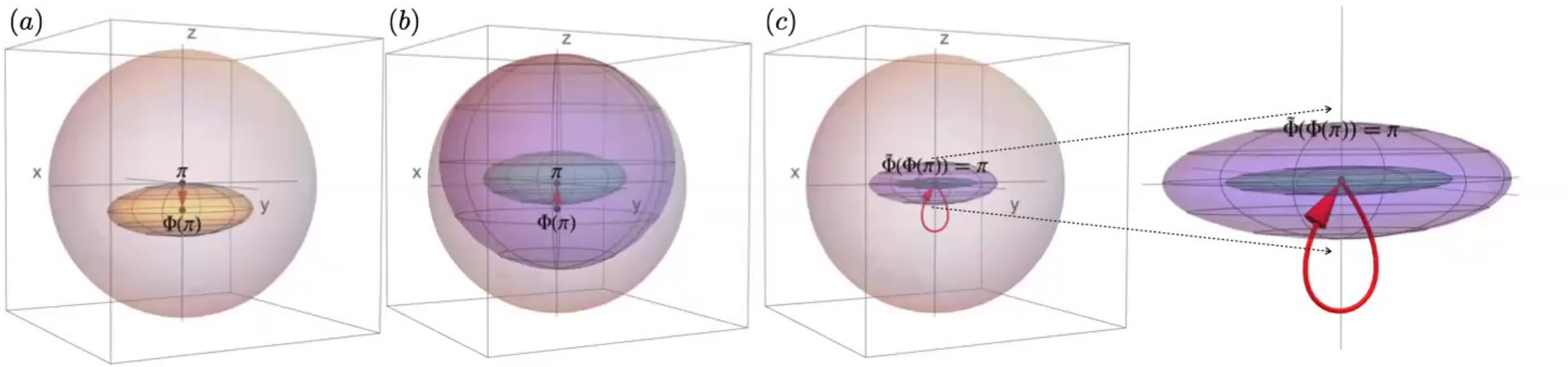
Comparison of the optimal state retrieval with Petz

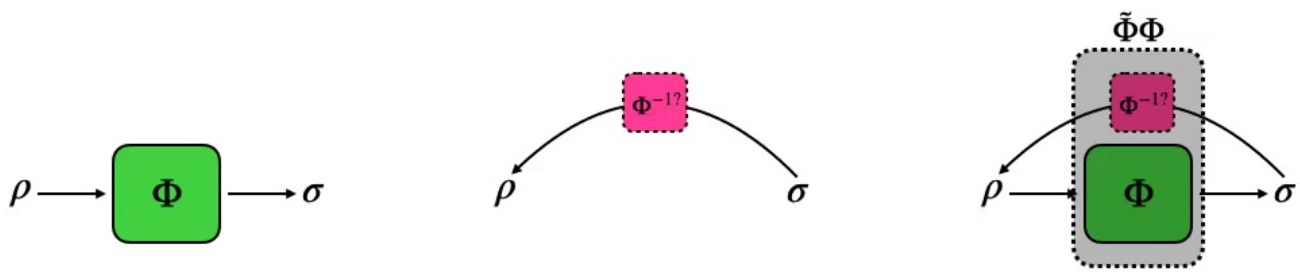
Stochastic channel:	Φ
Prior:	π

Basic ingredients

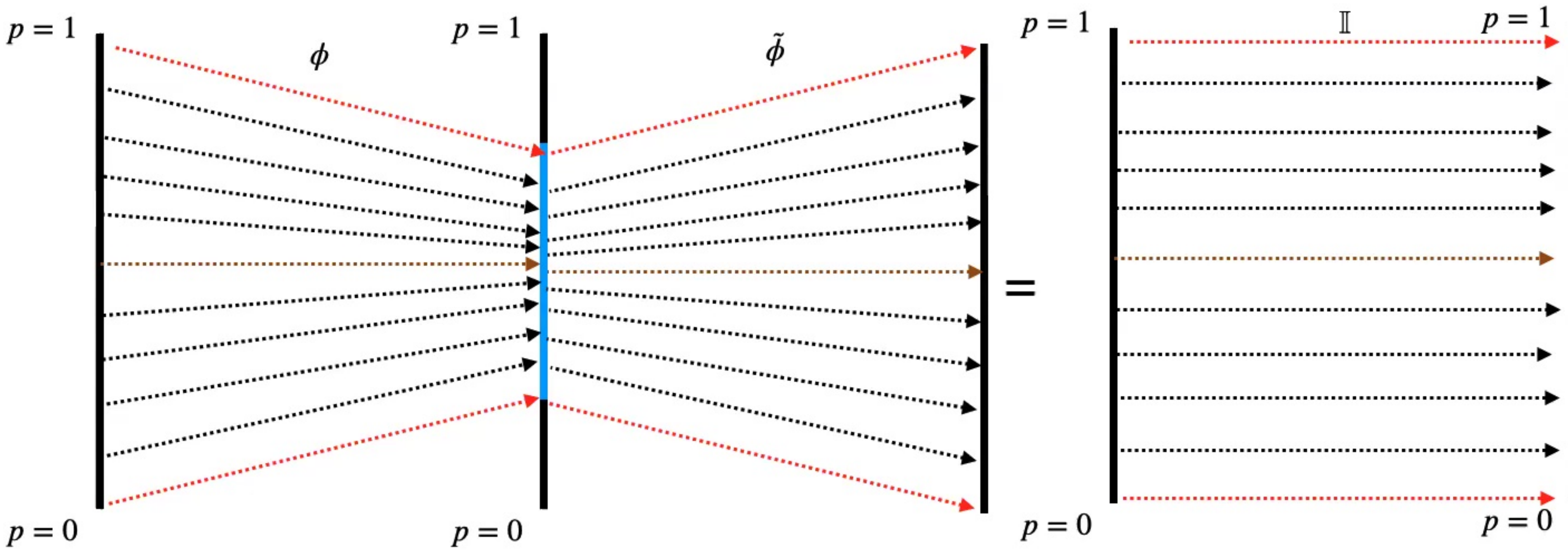
$\tilde{\Phi}_P$

$\tilde{\Phi}_O$ ★

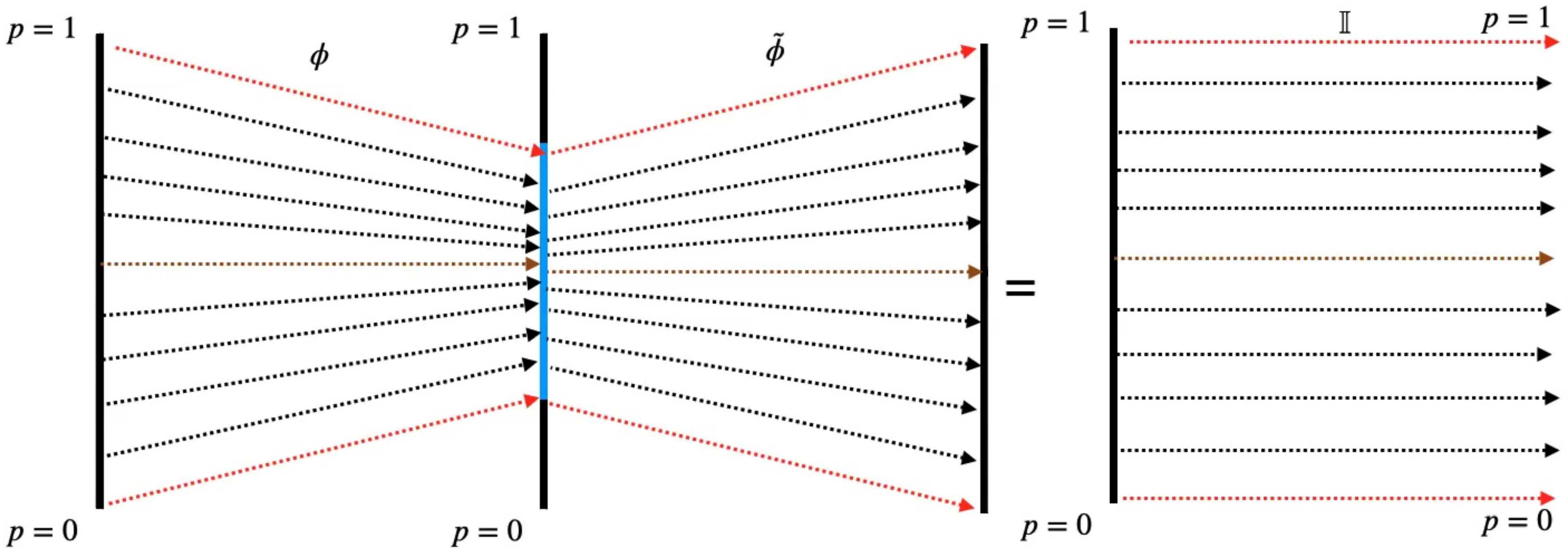




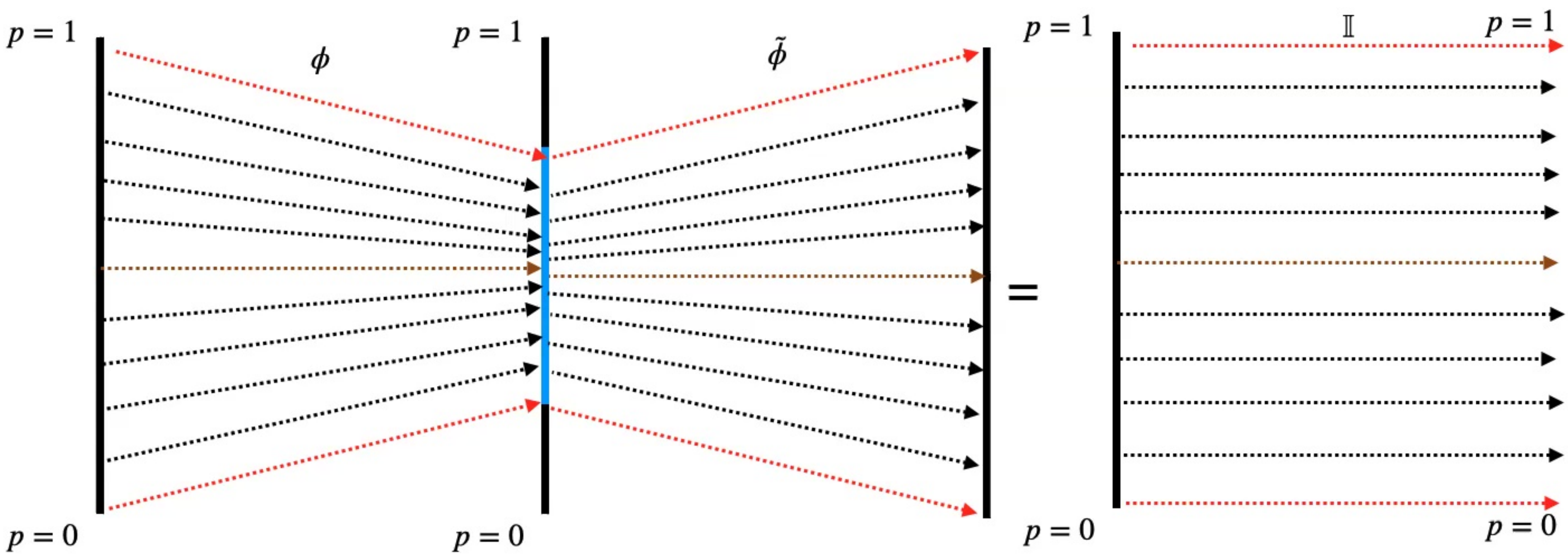
Reversing a channel



Reversing a channel



Reversing a channel



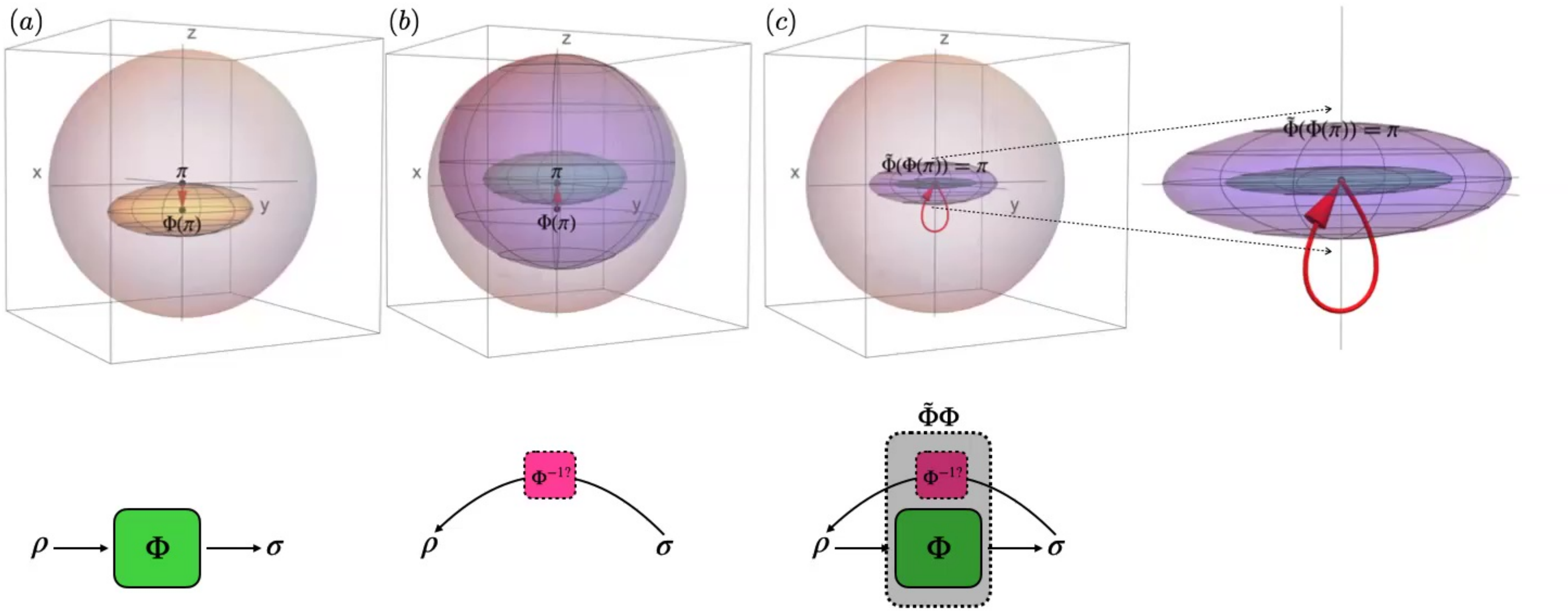
Comparison of the optimal state retrieval with Petz

Stochastic channel: Φ
Prior: π

Basic ingredients

■ $\tilde{\Phi}_P$

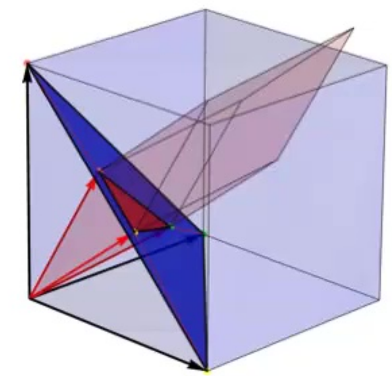
■ $\tilde{\Phi}_O$ ★



IDEA: minimal contraction

Optimisation criterion: **The optimal retrieval map is the one that maximise the determinant of the forth-and-back channel.**

$$\tilde{\Phi}_O = \max_{\tilde{\Phi}\ \text{state retrieval}} \det \tilde{\Phi}\Phi$$



Second reason for choosing this optimisation criterion

$$D(\tilde{\Phi}\Phi||\mathbb{1}) = \mathrm{Tr}[\mathbb{1}(\log \mathbb{1} - \log \tilde{\Phi}\Phi)] =$$
$$= -\mathrm{Tr}[\log \tilde{\Phi}\Phi] = \log\det(\tilde{\Phi}\Phi)^{-1}$$

Practical reason

Encyclopedia of Optimization pp 3375–3380 | Cite as

Semidefinite Programming and Determinant Maximization

Lieven Vandenberghe, Stephen Boyd & Shao-Po Wu