

# The Kernel Quantum Probabilities (KQP) Library

B. Piwowski

## Abstract

In this document, we show how the different quantities necessary to compute kernel quantum probabilities can be computed. This document forms the basis of the implementation of the Kernel Quantum Probability (KQP) open source project<sup>1</sup>.

## 1 Introduction

Quantum Probabilities correspond to one of the generalisations of standard probabilities. It is founded on the mathematical theory underlying Quantum Physics. This framework was developed in the 1930s by von Neumann and Dirac. It was recently further developed and generalised by the so-called “sequential effect algebra” [3]. The Kernel Quantum Probability library (KQP) aims to provide tools to effectively compute “quantum probabilities”, that is to compute a representation of densities, events and to update the densities when events are observed (conditionalisation). It also provides access to generalisation of standard probabilistic measures like entropy and divergence [3].

Computing quantum probabilities related quantities relies on linear algebra, and more precisely on the definition of an inner product in a Hilbert space, since this defines the probability of transition (when measuring) between possible system states.

In the machine learning community, a standard “trick” is to use a *kernel* to define the inner product [?]. That is, states can be represented in an arbitrary feature space  $\mathcal{F}$  for which there exists a mapping  $\Phi$  such that  $\Phi(x) \cdot \Phi(y)$  is a valid inner product. We call  $k(x, y) = \Phi(x) \cdot \Phi(y)$  the kernel, which can be computed without explicitly computing  $\Phi(x)$ , thus allowing to work in high or infinite spaces.

This document describes how to compute quantum probabilities related quantities relying only on the inner product definition given by the kernel. The organisation of this document is as follows:

1. In Section 2, we describe how to compute probabilities and how to update the probabilities given a subspace (or its orthogonal), both from a theoretical point of view and implementation point of view.
2. In Section 3.2, we describe how to compute an approximation of *quantum* densities or events.
3. In Section 4, we give an example of code using KQP in C++.

## Notations

We suppose that we work within a complex Hilbert space, that is that the field is  $\mathbb{C}$  unless otherwise specified. The set of complex matrices of dimensions  $n$  by  $p$  is denoted  $\mathcal{M}_{n \times p}$ .

In order to deal with kernels, following the literature, data points in the original space will be called pre-images since they are used to build a basis of the subspace containing the quantum density or event (see Section 2).

---

<sup>1</sup><http://kqp.bpiwowar.net/>

In order to use common linear algebra notations, we consider a list of pre-images as a linear map, and use uppercase calligraphic letters to denote a list of pre-images: A list  $\mathcal{X}$  of  $n$  pre-images is denoted

$$\mathcal{X} \in \mathcal{M}_{\mathcal{H}}^{(n)}$$

An arbitrary list belongs to  $\mathcal{M}_{\mathcal{H}} = \bigcup_n \mathcal{M}_{\mathcal{H}}^{(n)}$ . A linear combination of pre-images is simply denoted  $\mathcal{X}A$  where  $A \in \mathcal{M}_{n \times p}$ .

We define the adjoint operator in a natural way, i.e. it maps a list of pre-images into

$$\mathcal{X}^\dagger \in \cup_p \mathcal{L}(\mathcal{M}_{\mathcal{H}}^{(p)}; \mathcal{M}_{n \times p})$$

We denote  $k$  the kernel, i.e. we denote

$$k(\mathcal{X}, \mathcal{U}) = \mathcal{X}^\dagger \mathcal{U} \in \mathcal{M}_{n \times p}$$

with  $\mathcal{X} \in \mathcal{M}_{\mathcal{H}}^{(n)}$  and  $\mathcal{U} \in \mathcal{M}_{\mathcal{H}}^{(p)}$ .

Finally, we use the symbol  $\mathcal{P}$  to denote the composition of a linear operator with its transpose, i.e.

$$\mathcal{P}(A) = AA^\dagger$$

## 2 Computing probabilities

Readers are referred to [3, 4] for a discussion and presentation of what are quantum probabilities. Shortly, they can be defined by:

**A quantum density** is a positive semi-definite self-adjoint linear operator  $\rho$  of trace 1;

**An Observable** is a projector and corresponds to a yes-no measurement, i.e. to a quantum “event”;

**An effect** is an operator  $A$  such that  $0 \leq \langle Ax, x \rangle \leq 1$ . Note that an observable is an effect, but the reverse is not true. Effects can be considered as “fuzzy” or “imprecise” observables.

In this section, we present the formulas corresponding to the various quantities of interest (probability, conditionalisation, divergence) and how we can compute them within KQP. This section is based on the work of Gudder [3] (effects) and [6] (divergence and entropy).

In this section, we suppose we have a density  $\rho$  and an effect  $E \in \mathcal{E}(\mathcal{H})$  that can be decomposed as:

$$\rho = \mathcal{P}(\mathcal{X}_\rho Y_\rho \Sigma_\rho) \text{ and } E = \mathcal{P}(\mathcal{X}_E Y_E \Sigma_E)$$

where  $\mathcal{X}_\bullet$  belong to  $\mathcal{M}_{\mathcal{H}}$ ,  $Y_\bullet$  and  $Z_\bullet$  to  $\mathcal{M}$ .

For some operations, we need the decomposition to be in an orthonormal form, i.e. that

$$Y^\dagger \mathcal{X}^\dagger \mathcal{X} Y = Id \tag{1}$$

When using the orthonormality hypothesis, we use the symbol  $\mathbf{o}[\dots]$  over the equality. For example,

$$Y_\rho^\dagger \mathcal{X}_\rho^\dagger \mathcal{X}_\rho Y_\rho \stackrel{\mathbf{o}[\rho]}{=} Id$$

Finally, for densities we use the proportionality to denote that it should be normalised, i.e.  $\rho \propto \rho_u$  means that

$$\rho = \frac{\rho_u}{\text{tr}(\rho_u)}$$

Note that it is straightforward to compute the normalisation factor, since, using the cyclic re-ordering property of trace operators:

$$\text{tr}(\rho) = \text{tr}(\mathcal{X}_\rho Y_\rho \Sigma_\rho^2 Y_\rho^\dagger \mathcal{X}_\rho^\dagger) = \text{tr}(\Sigma_\rho^2 Y_\rho^\dagger \mathcal{X}_\rho^\dagger \mathcal{X}_\rho Y_\rho)$$

If the decomposition is orthonormal (Eq. 1), we have  $\text{tr}(\rho) = \text{tr}(\Sigma_\rho^2) = \|\Sigma_\rho\|^2$ .

For a matrix  $A$ , we denote  $A_{\bullet j}$  its  $j^{\text{th}}$  column,  $A_{i \bullet}$  its  $i^{\text{th}}$  row. If a matrix  $A$  as an subscript  $\rho$ , we use a semicolon to separate the subscript from the column/row indices, as for example in  $A_{\rho;ij}$ .

## 2.1 Computing probabilities

The probability of an effect  $E$  is defined as

$$\Pr_\rho(E) = \text{tr}(\rho E)$$

We can compute the probability of an effect  $E$  using the re-ordering property of the trace operator

$$\Pr_\rho(E) = \text{tr}(\rho E) = \left\| \Sigma_E Y_E^\dagger k(\mathcal{X}_E, \mathcal{X}_\rho) Y_\rho \Sigma_\rho \right\|^2$$

## 2.2 Entropy

The entropy of a density  $\rho$  (with an orthonormal decomposition) can be written [6]:

$$\begin{aligned} \text{tr}(\rho \log(\rho)) &\stackrel{\text{or}[\rho]}{=} \text{tr}((Y_\rho^\dagger \mathcal{X}_\rho^\dagger \mathcal{X}_\rho Y_\rho) \Sigma_\rho (Y_\rho^\dagger \mathcal{X}_\rho^\dagger \mathcal{X}_\rho Y_\rho) \log(\Sigma_\rho^2)) \\ &= \text{tr}(\Sigma_\rho^2 \log(\Sigma_\rho^2)) \\ &= \sum_i 2\Sigma_{\rho;ii}^2 \log(\Sigma_{\rho;ii}) \end{aligned}$$

## 2.3 Divergence

Umegaki [6] proved that the equivalent of the Kullback-Leibler divergence between two densities  $\rho$  and  $\tau$  can be computed as:

$$J(\rho||\tau) = \text{tr}(\rho \log(\rho) - \rho \log(\tau))$$

The first part corresponds to the entropy, and the second part can be computed as follows. In order to deal with infinities, in practice we want to compute the divergence using  $\tau' = (1 - \epsilon)\tau + \epsilon\alpha Id$  where  $\alpha Id$  is a blank noise, i.e.  $\text{tr}(\alpha Id) = 1$ . In case of infinities,  $\alpha$  can be set to a small value. We have:

$$\begin{aligned} \text{tr}(\rho \log \tau') &= \text{tr}(\rho \log((1 - \epsilon)\tau + \epsilon\alpha Id)) \\ &\stackrel{\text{or}[\tau]}{=} \text{tr}(\rho [\mathcal{X}_\tau Y_\tau \log((1 - \epsilon)\Sigma_\tau^2 + \epsilon\alpha Id) Y_\tau^\dagger \mathcal{X}_\tau^\dagger + \log(\epsilon\alpha) (Id - \mathcal{X}_\tau Y_\tau Y_\tau^\dagger \mathcal{X}_\tau^\dagger)]) \\ &= \text{tr}(\Sigma_\rho Y_\rho^\dagger \mathcal{X}_\rho^\dagger \mathcal{X}_\rho Y_\rho \log((1 - \epsilon)\Sigma_\tau^2 + \epsilon\alpha Id) Y_\tau^\dagger \mathcal{X}_\tau^\dagger \mathcal{X}_\tau Y_\tau \Sigma_\rho) + \log(\epsilon\alpha) (\text{tr}(\rho) - \text{tr}(\rho \mathcal{X}_\tau Y_\tau Y_\tau^\dagger \mathcal{X}_\tau^\dagger)) \\ &= - \left\| \Sigma_\rho Y_\rho^\dagger \mathcal{X}_\rho^\dagger \mathcal{X}_\rho Y_\rho (\log((1 - \epsilon)\Sigma_\tau^2 + \epsilon\alpha Id))^{1/2} \right\|^2 + \log(\epsilon\alpha) \left(1 - \left\| \Sigma_\rho Y_\rho^\dagger \mathcal{X}_\rho^\dagger \mathcal{X}_\rho Y_\rho \right\|^2\right) \end{aligned}$$

## 2.4 Conditionalisation

We first give the formulas to compute the conditional quantum density when observing an effect  $E$ , and then when observing its orthogonal  $E^\perp$

### 2.4.1 Projecting on the effect $E$

If we observe the event  $E$ , the density  $\rho$  conditioned upon  $E$ , denoted  $\rho \triangleright E$ , is given by:

$$\rho \triangleright E = \frac{E^{1/2} \rho E^{1/2}}{\text{tr}(\rho E)}$$

We can focus on the numerator since we only have to normalise the resulting density afterwards. We have

$$\rho \triangleright E \propto E^{1/2} \rho E^{1/2}$$

We can distinguish two cases:

1.  $E$  is an observable: since  $E = E^{1/2}$ , we have

$$\rho \triangleright E = \mathcal{P} \left[ \mathcal{X}_E \left( Y_E \Sigma_E^2 Y_E^\dagger k(\mathcal{X}_E, \mathcal{X}_\rho) Y_\rho \right) \Sigma_\rho \right]$$

2.  $E$  is a “strict” effect: In this case, we require an orthonormal decomposition for  $E$ , and we can compute the projection as:

$$\rho \triangleright E \stackrel{\circ[E]}{=} \mathcal{P} \left[ \mathcal{X}_E \left( Y_E \Sigma_E Y_E^\dagger k(\mathcal{X}_E, \mathcal{X}_\rho) Y_\rho \right) \Sigma_\rho \right]$$

In both cases, the resulting density is not in an orthonormal form.

### 2.4.2 Projecting on the orthogonal $E^\perp$

If we observe the orthogonal of event  $E$ , we can update our knowledge on  $\rho$ , denoted  $\rho \triangleright E^\perp$ , as:

$$\rho \triangleright E^\perp = \frac{(Id - E)^{1/2} \rho (Id - E)^{1/2}}{1 - \text{tr}(\rho E)}$$

When  $E$  is in an orthonormal form, we can use the fact that  $Id - Y_E \mathcal{X}_E \mathcal{X}_E^\dagger Y_E^\dagger$  is the projector on the space orthogonal to the space spanned by the vectors of  $E$ . Thus,

$$\begin{aligned} (Id - E)^{1/2} &= \mathcal{X}_E V (Id - \Sigma^2)^{1/2} Y_E^\dagger \mathcal{X}_E^\dagger + \left( Id - \mathcal{X}_E Y_E Y_E^\dagger \mathcal{X}_E^\dagger \right) \\ &= Id - \mathcal{X}_E Y_E \left[ Id - (Id - \Sigma_E^2)^{1/2} \right] Y_E^\dagger \mathcal{X}_E^\dagger \end{aligned} \quad (2)$$

Using the above, we can write:

$$\begin{aligned} \rho \triangleright E^\perp &\propto \mathcal{P} \left( (Id - E)^{1/2} \mathcal{X}_\rho Y_\rho \Sigma_\rho \right) \\ &\propto \mathcal{P} \left[ \begin{pmatrix} \mathcal{X}_\rho & \mathcal{X}_E \end{pmatrix} \begin{pmatrix} Id \\ -Y_E \left[ Id - (Id - \Sigma_E^2)^{1/2} \right] Y_E^\dagger k(\mathcal{X}_E, \mathcal{X}_\rho) \end{pmatrix} Y_\rho \Sigma_\rho \right] \end{aligned}$$

We readily verify that when  $\Sigma_E = Id$  it gives the right formula  $\rho - \mathcal{X} Y Y^\dagger \mathcal{X}^\dagger \rho$ . We can use in those cases a direct EVD approach (section 3.1) to obtain a simplified form.

## 3 Approximating operators

In this section, we describe the techniques used to compute low-rank approximations of linear operators in the feature space. In particular, we are interested in methods where the operator can be decomposed as:

$$\sum_i \alpha_i \mathcal{U}_i A_i A_i^\dagger \mathcal{U}_i^\dagger \approx \mathcal{X} Y \Sigma Y^\dagger \mathcal{X}^\dagger \quad (3)$$

where  $\mathcal{X} Y$  is (or might be) orthonormal, i.e.  $Y^\dagger \mathcal{X}^\dagger \mathcal{X} Y$  is the identity.

In the following, we describe:

- In Section 3.1, how to get an EVD decomposition of any linear operator of the form  $\mathfrak{U} = \mathcal{X} A A^\dagger \mathcal{X}^\dagger$ . This is useful in order to e.g. lower the rank and is needed or before removing feature space vectors from  $\mathcal{X}$ .
- In Section 3.2, we show how to update the EVD of a linear operator  $\mathfrak{U}$  with a low rank operator  $\alpha_i \mathcal{U}_i A_i A_i^\dagger \mathcal{U}_i^\dagger$ .

- In Section 3.3.2, we show how to remove feature vectors from  $\mathcal{X}$  when we have an EVD  $\mathfrak{U} = \mathcal{X}\Sigma A^\dagger \mathcal{X}^\dagger$ . We use two techniques:
  - Null space method (Section 3.3.1)
  - Quadratic optimisation to find the subset of pre-images that minimise the reconstruction error (Section 3.3.2).

### 3.1 Direct EVD

In this section, we discuss how to get the orthonormal form of an operator written as  $\mathfrak{U} = \mathcal{X}ASA^\dagger \mathcal{X}^\dagger$  where  $S$  is a diagonal matrix. We first describe the case where  $S$  is positive semi-definite, before tackling the general case.

This type of approach is useful in several cases, and the builder `AccumulatorKernelEVD` in KQP relies on this decomposition, since it represents Eq. (3) as

$$(\mathcal{U}_1 \ \cdots \ \mathcal{U}_n) \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_n \end{pmatrix} \begin{pmatrix} \Sigma_1 & & \\ & \ddots & \\ & & \Sigma_n \end{pmatrix} \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_n \end{pmatrix}^\dagger (\mathcal{U}_1 \ \cdots \ \mathcal{U}_n)^\dagger$$

where  $\Sigma_i = \text{diag}(\alpha_i, \dots, \alpha_i)$ .

#### 3.1.1 Semi-positive definite case

Suppose we have  $\mathfrak{U} = \mathcal{X}AA^\dagger \mathcal{X}^\dagger$  and we wish to transform it to an orthonormal form. To achieve this, we have to compute a thin EVD

$$EDE^\dagger = A^\dagger \mathcal{X}^\dagger \mathcal{X} A$$

It is then straightforward to obtain the desired form by posing  $Y = AED^{-1/2}$  and  $\Sigma = D^{1/2}D^{1/2}$

$$\mathcal{X}Y\Sigma Y^\dagger \mathcal{X}^\dagger = \mathcal{X}AEE^\dagger A^\dagger \mathcal{X}^\dagger = \mathcal{X}ASA^\dagger \mathcal{X}^\dagger$$

where the last equality can be shown has follows. Any vector  $y \in \mathcal{H}$  can be written  $\mathcal{X}AP + \mathcal{V}Q$  where  $\mathcal{V}^\dagger \mathcal{X}A = 0$ . Then,

$$\begin{aligned} \mathcal{X}AEE^\dagger A^\dagger \mathcal{X}^\dagger y &= \mathcal{X}AEE^\dagger \underbrace{A^\dagger \mathcal{X}^\dagger \mathcal{X} A}_{EDE^\dagger} P + \underbrace{\mathcal{X}AEE^\dagger A^\dagger \mathcal{X}^\dagger \mathcal{V} Q}_0 \\ &= \mathcal{X}AA^\dagger \mathcal{X}^\dagger \mathcal{X} A P + \underbrace{\mathcal{X}AA^\dagger \mathcal{X}^\dagger \mathcal{V} Q}_0 \\ &= \mathcal{X}AA^\dagger \mathcal{X}^\dagger y \end{aligned}$$

We also can show easily that  $\mathcal{X}Y$  is an orthonormal matrix

$$Y^\dagger \mathcal{X}^\dagger \mathcal{X} Y = D^{-1/2} E^\dagger (A^\dagger \mathcal{X}^\dagger \mathcal{X} A) E D^{-1/2} = Id$$

It is then possible to remove some pre-images using techniques from Section 3.3.2.

#### 3.1.2 General case

In the general case, we have  $\mathfrak{U} = \mathcal{X}ASA^\dagger \mathcal{X}^\dagger$  which can be rewritten  $\mathfrak{U} = \mathcal{X}AS^{1/2}S^{1/2}A^\dagger \mathcal{X}^\dagger$ . That is, unless we use a real field and  $S$  is not semidefinite positive. In that case, we can still write

$$\mathfrak{U} = \mathcal{X}BB^\dagger \mathcal{X}^\dagger - 2\mathcal{X}CC^\dagger \mathcal{X}^\dagger$$

where

$$\begin{aligned} B &= A(S_+ + S_-)^{1/2} \\ C &= AS_-^{1/2} \end{aligned}$$

where  $S_{\pm}$  is the  $S$  matrix where negative (resp. positive) values are set to 0. We then use the approach above to compute an orthonormal decomposition of  $\mathcal{X}BB^\dagger\mathcal{X}^\dagger$ , and then, using the fact that the space defined by  $\mathcal{X}YY\mathcal{X}^\dagger$  contains  $\mathcal{X}C$ ,

$$\begin{aligned} \mathfrak{U} &= \mathcal{X}Y\Sigma Y^\dagger\mathcal{X}^\dagger - 2\mathcal{X}CC^\dagger\mathcal{X}^\dagger \\ &= \mathcal{X}Y[\Sigma - 2ZZ^\dagger]Y^\dagger\mathcal{X}^\dagger \end{aligned}$$

with  $Z = Y^\dagger\mathcal{X}^\dagger\mathcal{X}C$ . We then have to compute another EVD for  $\Sigma - 2ZZ^\dagger$ , which will give the final form of  $\mathfrak{U}$ .

### 3.2 Low-rank update of operators

The problem is to compute a low rank approximation of

$$\mathfrak{U} = \sum_i \alpha_i \mathcal{U}_i A_i A_i^\dagger \mathcal{U}_i^\dagger$$

where  $\mathcal{U}_i \in \mathcal{M}_{\mathcal{H}}$ .

In the following, we consider just one update and we drop the  $i$  for more clarity. We further assume that we have a current approximation decomposition expressed as

$$\mathfrak{U} = \mathcal{X}YZ\Sigma Z^\dagger Y^\dagger \mathcal{X}$$

where

- $\mathcal{X} \in \mathcal{M}_{\mathcal{H}}^{(n)}$  and  $Y$  is a  $n \times r$  matrix such that  $\mathcal{X}Y$  is orthonormal;
- $Z$  is a  $r \times r$  unitary matrix. This matrix is used in order to avoid updating the potentially larger matrix  $Y$  when the list of pre-images remain the same;
- $\Sigma$  is a diagonal matrix of rank  $r$

In order to be able to process incrementally the set of vectors  $\mathcal{U}$ , we wish to compute at each step a rank one update of  $U$

$$\widehat{\mathfrak{U}} = \mathfrak{U} + \alpha \mathcal{U} A A^\dagger \mathcal{U}^\dagger \approx \widetilde{\mathfrak{U}} = \mathcal{X}' Y' Z' \Sigma (\mathcal{X}' Y' Z')^\dagger$$

This problem is related to [2] that deals with incremental Kernel SVD, and we follow mainly the same approach. We use the following constraints:

1. Keep the (relative) error  $\epsilon = \left\| \widehat{\mathfrak{U}} - \widetilde{\mathfrak{U}} \right\| / \|\mathfrak{U}\|$  below a limit  $\eta$  (if possible, see below);
2. Keep the rank  $r$  below the limit  $r_{\max}$  ;
3. Keep the number of pre-images below a number  $cr$  where  $c \geq 1$ .

#### 3.2.1 Pre-computations

We can write  $\mathcal{U}$  as the direct sum

$$\mathcal{U}A = \underbrace{(Id - \mathcal{X}Y Y^\dagger \mathcal{X}^\dagger) \mathcal{U}A}_V + \mathcal{X}Y \underbrace{Y^\dagger \mathcal{X}^\dagger \mathcal{U}A}_W \quad (4)$$

The operator  $W$  can be computed explicitly as:

$$W = Y^\dagger \mathcal{X}^\dagger \mathcal{U}A = Y^\dagger k(\mathcal{X}, \mathcal{U}) A \quad (5)$$

**General case** We can compute  $\mathcal{V}^\dagger \mathcal{V}$  as

$$\mathcal{V}^\dagger \mathcal{V} = A^\dagger k(\mathcal{U}, \mathcal{U}) A - W^\dagger W$$

which can in turn be used to compute<sup>2</sup> the (full) EVD of  $\mathcal{V}\mathcal{V}^\dagger$ .

**Special case  $\mathcal{U} = \mathcal{X}$**  When  $\mathcal{U}A$  is a linear combination of kernel vectors. In this case, we have  $\mathcal{V} = 0$  and

$$W = Y^\dagger \mathcal{X}^\dagger \mathcal{X} A = Y^\dagger k(\mathcal{X}, \mathcal{X}) A$$

**Updating the operator** Let us express  $\mathcal{U}A$  as the direct sum of its projection onto the subspace spanned by  $\mathcal{X}Y Y^\dagger \mathcal{X}^\dagger$  and its orthogonal. Since by definition  $W = (\mathcal{X}Y)^\dagger \mathcal{U}A$ , we can write  $\mathcal{U}A$  as:

$$\mathcal{U}A A^\dagger \mathcal{U}^\dagger = \mathcal{P}(\mathcal{X}Y Y^\dagger \mathcal{X}^\dagger \mathcal{U}A + \mathcal{V}) = \mathcal{P}\left(\begin{pmatrix} \mathcal{X} & \mathcal{V} \end{pmatrix} \begin{pmatrix} Y & 0 \\ 0 & Q \end{pmatrix} \begin{pmatrix} WQD^{1/2} & WQ_0 \\ D^{1/2} & \mathbf{0} \end{pmatrix}\right)$$

where  $Q, Q_0$  and  $D$  such that  $\begin{pmatrix} Q & Q_0 \end{pmatrix}$  is unitary and

$$\mathcal{V}QDQ^\dagger \mathcal{V}^\dagger = \mathcal{V}\mathcal{V}^\dagger \quad (6)$$

We can write

$$\mathfrak{U} = \mathcal{P}\left(\mathcal{X}Y Z \Sigma^{1/2} \quad 0\right) = \mathcal{P}\left[\begin{pmatrix} \mathcal{X} & \mathcal{V} \end{pmatrix} \begin{pmatrix} Y & 0 \\ 0 & Q \end{pmatrix} \begin{pmatrix} Z & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \Sigma^{1/2} & 0 \\ 0 & 0 \end{pmatrix}\right]$$

and hence:

$$\begin{aligned} \mathfrak{U} + \alpha \mathcal{U}A A^\dagger \mathcal{U}^\dagger &= \begin{pmatrix} \mathcal{X} & \mathcal{V} \end{pmatrix} \begin{pmatrix} Y & 0 \\ 0 & Q \end{pmatrix} \left[ \begin{pmatrix} Z & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Z & 0 \\ 0 & 0 \end{pmatrix}^\dagger + \right. \\ &\quad \left. \alpha \begin{pmatrix} WQD^{1/2} & WQ_0 \\ D^{1/2} & \mathbf{0} \end{pmatrix} \begin{pmatrix} WQD^{1/2} & WQ_0 \\ D^{1/2} & \mathbf{0} \end{pmatrix}^\dagger \right] \begin{pmatrix} Y & 0 \\ 0 & Q \end{pmatrix}^\dagger \begin{pmatrix} \mathcal{X} & \mathcal{V} \end{pmatrix}^\dagger \quad (7) \end{aligned}$$

**Computing  $Q$**  Since  $\begin{pmatrix} \mathcal{X} & \mathcal{V} \end{pmatrix} \begin{pmatrix} Y & 0 \\ 0 & Q \end{pmatrix}$  should be an orthonormal matrix, we should have:

$$\begin{aligned} \left(\begin{pmatrix} \mathcal{X} & \mathcal{V} \end{pmatrix} \begin{pmatrix} Y & 0 \\ 0 & Q \end{pmatrix}\right)^\dagger \begin{pmatrix} \mathcal{X} & \mathcal{V} \end{pmatrix} \begin{pmatrix} Y & 0 \\ 0 & Q \end{pmatrix} &= \begin{pmatrix} (\mathcal{X}Y)^\dagger \\ Q^\dagger \mathcal{V}^\dagger \end{pmatrix} \begin{pmatrix} \mathcal{X}Y & \mathcal{V}Q \end{pmatrix} \\ &= \begin{pmatrix} Id & 0 \\ 0 & Q^\dagger \mathcal{V}^\dagger \mathcal{V}Q \end{pmatrix} \end{aligned}$$

where we used the fact that  $(\mathcal{X}Y)^\dagger \mathcal{V} = 0$  from Eq. (4). Hence, orthonormality is equivalent to

$$Q^\dagger \mathcal{V}^\dagger \mathcal{V}Q = Id \quad (8)$$

Using the result of Section 3.1, we can compute the thin EVD  $CDC^\dagger$  of  $k(\mathcal{V}, \mathcal{V})$  and pose  $Q = CD^{-1/2}$  which will verify both Eqs. (6) and (8).  $Q_0$  corresponds to the basis of the null space obtained using the same decomposition (note that if  $\mathcal{V} = 0$ ,  $Q_0$  is the identity).

<sup>2</sup>Note that we could use a Cholesky decomposition  $k(\mathcal{U}, \mathcal{U}) = LL^\dagger$  followed by a generalised SVD on  $L^\dagger A$  and  $W$  to find the EVD of  $\mathcal{V}\mathcal{V}^\dagger$ .

**Updating** We can use standard rank-one update techniques to update the decomposition; since  $Z$  is unitary, we can write

$$\begin{aligned} & \begin{pmatrix} Z & 0 \\ 0 & Id \end{pmatrix} \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Z & 0 \\ 0 & Id \end{pmatrix}^\dagger + \alpha \begin{pmatrix} WQD^{1/2} & WQ_0 \\ D^{1/2} & \mathbf{0} \end{pmatrix} \begin{pmatrix} WQD^{1/2} & WQ_0 \\ D^{1/2} & \mathbf{0} \end{pmatrix}^\dagger \\ &= \begin{pmatrix} Z & 0 \\ 0 & Id \end{pmatrix} \left( \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} + \alpha \begin{pmatrix} Z^\dagger WQD^{1/2} & WQ_0 \\ D^{1/2} & \mathbf{0} \end{pmatrix} \begin{pmatrix} Z^\dagger WQD^{1/2} & WQ_0 \\ D^{1/2} & \mathbf{0} \end{pmatrix}^\dagger \right) \begin{pmatrix} Z & 0 \\ 0 & Id \end{pmatrix}^\dagger \end{aligned}$$

which is a rank  $p$  update of a diagonal matrix. Note that, as we cannot really compute  $\mathcal{V}$ , we have to get back to an expression where  $\mathcal{U}$  appears in the first matrix:

$$\begin{aligned} (\mathcal{X} \ \mathcal{V}) \begin{pmatrix} Y & 0 \\ 0 & Q \end{pmatrix} &= (\mathcal{X} \ \mathcal{U}A - \mathcal{X}YW) \begin{pmatrix} Y & 0 \\ 0 & Q \end{pmatrix} \\ &= (\mathcal{X} \ \mathcal{U}) \begin{pmatrix} Y & -YWQ \\ 0 & AQ \end{pmatrix} \end{aligned}$$

### 3.3 Reducing the pre-image set

In this section, we describe the techniques used to reduce the number of pre-images. When a linear combination of pre-images is possible, it is better to use the direct EVD approach described in Section 3.1.

#### 3.3.1 Null space method

Suppose we have an operator  $\mathcal{U}$  defined as

$$\mathcal{U} = \mathcal{X}Y\Sigma Y^\dagger \mathcal{X}^\dagger$$

and we wish to reduce the set of pre-images in  $\mathcal{X}$  without loss. We suppose  $\Sigma$  is full rank.

1. Remove the pre-images for which a line of  $Y\Sigma Y^\dagger$  is null (and remove the corresponding column of  $Y$ ).
2. If the decomposition is not orthonormal, use a QR or LU decomposition to find the null space of  $\mathcal{X}^\dagger \mathcal{X}$ , i.e. a full rank  $Z$  such that  $\mathcal{X}^\dagger \mathcal{X}Z = 0$ . We then remove  $n$  pre-images (see below) where  $n$  is the rank of  $Z$ .
3. Finally, we remove non used pre-images like in step (1).

**Null space and pre-images** We want to find  $\mathcal{X}'$  and  $A$  such that  $\mathcal{X} = (\mathcal{X}' \ \mathcal{X}'A)P$  where  $P$  is a permutation matrix.

We have a basis  $Z$  for the null subspace of  $\mathcal{X}^\dagger \mathcal{X}$ . If  $z$  is in the null subspace, then

$$\forall i \ \mathcal{X}_i^\dagger \sum_j z_j \mathcal{X}_j = 0 \implies \sum_j z_j \mathcal{X}_j = 0$$

since  $\sum_j z_j \mathcal{X}_j$  belongs to the span of  $\mathcal{X}$ .

To chose among the pre-images, we chose to remove first those that are the less used, i.e. those for which  $\|Y_{j\bullet}\| \|\mathcal{X}_j\|$  is minimum. We also have to ensure that  $z_j$  is not too small, i.e. is above  $\delta \|z\|_\infty$ . We then remove entries one by one using the pivoted Gauss algorithm.



### 3.3.2 Quadratic Programming approach (L1-optimisation)

Another way to remove some pre-images is to try to directly optimise the cost using an  $L_1$  regulariser to set some rows of  $Y$  close to 0. Denoting  $A = YZ$ , we seek at minimising the difference

$$E = \left\| \mathcal{X}A\Sigma A^\dagger \mathcal{X}^\dagger - \mathcal{X}BTB^\dagger \mathcal{X}^\dagger \right\|^2 + \lambda \sum_i \|B_{i\bullet}\|_\infty$$

Using  $L_1$  regularisation ensures that  $B$  is sparser than  $A$  – in particular, rows of  $B$  are close to 0 (which means that the corresponding pre-images  $\mathcal{X}_i$  can be removed).

In the following, we suppose that  $A$  is of dimension  $r \times n$  (i.e.  $r$  basis vectors and  $n$  feature vectors). Using the link between the trace and the Frobenius norm, we have

$$E = \text{tr} \left( \mathcal{P} \left( \mathcal{X}A\Sigma A^\dagger \mathcal{X}^\dagger - \mathcal{X}BTB^\dagger \mathcal{X}^\dagger \right) \right)$$

Denoting  $a_i = \mathcal{X}A_{\bullet i}\Sigma^{1/2}$  and  $b_i = \mathcal{X}B_{\bullet i}T^{1/2}$  the two sets of vectors (in the feature space), we can then rewrite  $E$  as

$$\begin{aligned} E &= \text{tr} \left( \sum_{i,j} a_i a_i^\dagger a_j a_j^\dagger + b_i b_i^\dagger b_j b_j^\dagger - 2a_i a_i^\dagger b_j b_j^\dagger \right) \\ &= \sum_{i,j} \left( a_i^\dagger a_j \right)^2 + \left( b_i^\dagger b_j \right)^2 - 2 \left| a_i^\dagger b_j \right|^2 \end{aligned}$$

The problem we want to solve is linked to the “reduced set” approach proposed in [5], where one seeks to minimise the following cost function (with  $L_1$  regularisation):

$$\begin{aligned} \text{minimise } E_{RS} &= \sum_j \nu_j \|a_j - b_j\|^2 + \lambda \xi_i \\ \text{subject to } \forall i \xi_i &\geq \max_j |B_{ij}| \end{aligned} \tag{9}$$

The role of  $\xi_i$  is to regularise the importances of feature vectors; we need to set  $\lambda$  appropriately so that some rows of  $B$  are close to 0 at the end of the optimisation. Finally, and differently from other approaches, we added a new constant,  $\nu_i$ , that ensures that  $K \times E_{RS} \geq E$  for some  $K \geq 0$ . We discuss both in the following.

**Relation with the reduced set approach** We first check that minimising  $E_{RS}$  solves our problem. The main difference is that  $E$  contains terms of the form  $b_i^\dagger b_j$  and  $a_i^\dagger b_j$ . However, they will tend to be close to 0 since the feature vectors will be approximately orthogonal.

Posing  $b_i = \mu_i (a_i + c_i)$  with  $a_i \perp c_i$ , we can first show that  $\mu_i$  must be equal to  $\|a_i\|^{-2} (\|a_i\|^2 + \|c_i\|^2)$  when  $E_{RS}$  is minimised. Then, we can show that

$$6 \|a_i\|^2 \|a_i - b_i\|^2 \geq \left( a_i^\dagger a_i \right)^2 + \left( b_i^\dagger b_i \right)^2 - 2 \left| a_i^\dagger b_i \right|^2$$

Now we have to prove that all cross terms ( $i \neq j$ ) are minimised if we minimise the new objective function, which intuitively is ensured by the fact that  $a_i \perp a_j$ . Denoting  $\Delta_i = a_i - b_i$ , we have for  $i \neq j$ :

$$\left( a_i^\dagger a_j \right)^2 + \left( b_i^\dagger b_j \right)^2 - 2 \left| a_i^\dagger b_j \right|^2 = \left( a_i^\dagger \Delta_j + \Delta_i^\dagger a_j + \Delta_i^\dagger \Delta_j \right)^2 - 2 \left| a_i^\dagger \Delta_j \right|^2$$

which is clearly bounded by  $K \max_i \|\Delta_i\|^2$  and hence by  $K' E_{RS}$ .

Our problem is thus to optimise Eq. (9) with  $\nu_j = \sigma_j = \|a_j\|^2 = |\Sigma_{jj}|$ , or equivalently, by posing  $B'_{\bullet j} = \sigma_j B_j$ , we can reformulate the optimisation problem as:

$$\begin{aligned} & \text{minimise } E_{RS} = \sum_j \|\sigma_j a_j - b'_j\|^2 + \lambda \xi_j \\ & \text{subject to } \forall i \xi_i \geq \max_j \sigma_j^{-1} |B'_{ij}| \end{aligned} \tag{10}$$

**Setting**  $\lambda$  If  $B = A$  (trivial solution when  $\lambda = 0$ ), then to minimise the above equation, we set  $\xi_i = \max_j \sigma_j^{-1} |\sigma_j A_{ij} \sigma_j^{1/2}|$  and

$$E_{RS}^{(0)} = \lambda \sum_i \max_j |\sigma_j^{1/2} A_{ij}|$$

If we remove the  $i^{\text{th}}$  pre-image the error becomes

$$E_{RS}^{(i)} = E_{RS}^{(0)} - \lambda \max_j |\sigma_j^{1/2} A_{ij}| + \sum_j \underbrace{\left\| \sigma_j^{3/2} (\mathcal{X} A_{\bullet j} - \mathcal{X} A_{\bullet j}^{(i)}) \right\|^2}_{\sigma_j^3 |K_{ii} A_{ij}|^2}$$

where  $A^{(i)}$  is  $A$  with the  $i^{\text{th}}$  row set to zero. Hence, in order to remove the  $i^{\text{th}}$  pre-image, we need to set  $\lambda$  such that

$$E_{RS}^{(i)} - E_{RS}^{(0)} = |K_{ii}|^2 \sum_{j=1}^r \sigma_j^3 A_{\bullet j}^\dagger K A_{\bullet j} |A_{ij}|^2 - \lambda \max_j |\sigma_j^{1/2} A_{ij}| \geq 0$$

where  $K = k(\mathcal{X}, \mathcal{X})$ . If we want to remove (at least)  $m$  pre-images whose indices are in  $M$ , we want to have

$$\lambda \geq \frac{\sum_{i \in M} |K_{ii}|^2 \sum_{j=1}^r \sigma_j^3 A_{\bullet j}^\dagger K A_{\bullet j} |A_{ij}|^2}{\sum_{i \in M} \max_j |\sigma_i A_{ij}|}$$

As an heuristic, we set  $M$  to be the set of indices of pre-images with minimum  $E_{RS}^{(i)} - E_{RS}^{(0)}$ .

**Quadratic optimisation** The quadratic programming problem can be solved using quadratic cone optimisation. This is detailed in Appendix A.

**Re-estimation of parameters** We project the old operator into the new space in order to minimise the error, i.e.

$$\mathfrak{U} = (\mathcal{Y} B B^\dagger \mathcal{Y}^\dagger \mathcal{X} A) \Sigma A^\dagger \mathcal{X}^\dagger B \mathcal{Y} \mathcal{Y}^\dagger B^\dagger$$

## 4 Example

```
#include <kqp/feature_matrix/dense.hpp>
#include <kqp/kernel_evd/incremental.hpp>
#include <kqp/probabilities.hpp>
```

```
int main(int, const char**) {
```

```
    // --- Compute a density at random
```

```
    // Definitions
```

```
    using namespace kqp;
```

```

typedef Eigen::Matrix<double, Eigen::Dynamic, Eigen::Dynamic> Matrix;
int dim = 10;

// Creating an incremental builder
IncrementalKernelEVD<DenseMatrix<double>> kevd;

// Add 10 vectors with  $\alpha_i = 1$ 
for(int i = 0; i < 10; i++) {
    // Adds a random  $\varphi_i$ 
    Matrix m = Matrix::Random(dim, 1);
    kevd.add(DenseMatrix<double>(m));
}

// Get the result  $\rho \approx X Y D Y^\dagger X^\dagger$ 
DenseMatrix<double> mX;
typename AltDense<double>::type mY;
Eigen::Matrix<double, Eigen::Dynamic, 1> mD;

kevd.get_decomposition(mX, mY, mD);

// --- Compute a kEVD for a subspace

IncrementalKernelEVD<DenseMatrix<double>> kevd_event;
for(int i = 0; i < 3; i++) {
    // Adds a random  $\varphi_i$ 
    Matrix m = Matrix::Random(dim, 1);
    kevd_event.add(DenseMatrix<double>(m));
}

// --- Compute some probabilities

// Setup densities and events
Density<DenseMatrix<double>> rho(kevd);
Event<DenseMatrix<double>> event(kevd_event);

// Compute the probability
std::cout << "Probability = " << rho.probability(event) << std::endl;

// Conditional probability
Density<DenseMatrix<double>> rho_cond = event.project(kevd).normalize();

// Conditional probability (orthogonal event)
Density<DenseMatrix<double>> rho_cond_orth = event.project(kevd, true).normalize();

return 0;
}

```

## 5 Conclusion

This document described the Kernel Quantum Probability Library, that can be used to compute quantum events and density in the an arbitrary feature space and relies only on the definition of a kernel, i.e. of the inner product between any two feature vectors.

## References

- [1] M. Andersen, J. Dahl, Z. Liu, and L. Vandenberghe. Interior-point methods for large-scale cone programming. In S. N. S. J. W. S Sra, editor, *Optimization for Machine Learning*. MIT Press, Dec. 2011. [12](#)
- [2] T. Chin, K. Schindler, and D. Suter. Incremental Kernel SVD for Face Recognition with Image Sets. In *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pages 461–466, Southampton, UK, 2006. [6](#)
- [3] S. Gudder. Quantum probability. In K. Engesser, D. Gabbay, and D. Lehmann, editors, *Handbook of quantum logic and quantum structures*, pages 121–146. Elsevier, 2007. [1](#), [2](#)
- [4] I. Pitowsky. *Quantum Probability - Quantum Logic*, volume 321 of *Lecture Notes in Physics*. Springer-Verlag, Berlin/Heidelberg, 1989. [2](#)
- [5] B. Schölkopf, S. Mika, P. Knirsch, Müller, G. Rätsch, and A. Smola. Input space vs. feature space in kernel-based methods. *IEEE transactions on neural networks*, pages 1000–1017, 1999. [9](#)
- [6] H. Umegaki. Conditional expectation in an operator algebra. IV. Entropy and information. *Kodai Mathematical Seminar Reports*, 14(2):59–85, 1962. Notion of divergence in quantum probabilities  $K(r||s) = \text{tr} ( r ( \ln(r) + \ln(s) ) )$ . [2](#), [3](#)

## A QP Approach

In this section, we derive a computationally efficient way to optimise Eq. (9). We show here how to transform this optimisation problem into a cone quadratic programming approach proposed in [1].

### A.1 Precomputations

Writing  $\alpha_i = \sigma_i A_{\bullet,i}$ ,  $\beta_i = \tau_i B_{\bullet,i}$  and  $K = \mathcal{X}^\dagger \mathcal{X}$  the gram matrix ( $r$  is the rank of the opeator,  $n$  is the number of pre-images), we have

$$\begin{aligned} E_{RS} &= \sum_{q=1}^r \beta_q^\dagger \beta_q - 2\nu_q \Re(\alpha_q^\dagger \beta_q) + \lambda \xi_q \\ &= \sum_{q=1}^r \beta_q^\dagger K \beta_q - 2\Re(\alpha_q^\dagger K \beta_q) + \lambda \xi_q \end{aligned}$$

We get back to a real case by posing  $\beta_q = \beta'_q + i\beta''_q$  and  $\alpha_q = \alpha'_q + i\alpha''_q$ . Dropping  $q$  for clarity, we have

$$\begin{aligned} \beta^\dagger K \beta &= \beta'^\dagger K \beta' + \beta''^\dagger K \beta'' + i(\beta'^\dagger K \beta'' - \beta''^\dagger K \beta') \\ &= \beta'^\dagger \Re(K) \beta' + \beta''^\dagger \Re(K) \beta'' + i(\beta'^\dagger K \beta'' - \beta''^\dagger K \beta') \\ &= \beta'^\dagger \Re(K) \beta' + \beta''^\dagger \Re(K) \beta'' - 2\beta'^\dagger \Im(K) \beta'' \end{aligned}$$

and

$$\begin{aligned}\Re(\alpha^\dagger K \beta) &= \Re(\alpha^\dagger K \beta' + \alpha''^\dagger K \beta'' + i\alpha'^\dagger K \beta'' - i\alpha''^\dagger K \beta') \\ &= (\alpha'^\dagger \Re(K) + \alpha''^\dagger \Im(K)) \beta' + (\alpha''^\dagger \Re(K) - \alpha'^\dagger \Im(K)) \beta''\end{aligned}$$

If we pose

$$x = (\beta_1^\dagger \ \cdots \ \beta_r^\dagger \ \xi_1 \ \cdots \ \xi_n)^\dagger$$

with  $\beta_i = \begin{pmatrix} \beta_i' \\ \beta_i'' \end{pmatrix}$  in the complex case and  $\beta_i = \beta_i'$  in the real one.

We require that both the real and imaginary part be inferior to  $\xi$ , i.e. that

$$\begin{aligned}\forall i \in 1 \dots n, \forall q \in 1 \dots r, S_{qq} \beta_{qi}' + \xi_i \geq 0 \text{ and } -S_{qq} \beta_{qi}' + \xi_i \geq 0 \\ S_{qq} \beta_{qi}'' + \xi_i \geq 0 \text{ and } -S_{qq} \beta_{qi}'' + \xi_i \geq 0\end{aligned}$$

where  $S$  is a diagonal matrix of size  $r$  such that  $S_{jj} = \sigma_j^{-1}$ , and  $x$  has a length  $n \times (r' + 1)$ . Our problem can be expressed as a cone quadratic problem

$$\begin{aligned}\text{minimise } x^\dagger H x + 2c^\dagger x \\ \text{subject to } G x \leq 0\end{aligned}$$

Denoting  $Id_n^{(\cdot)}$  the matrix  $(Id_n \ \cdots \ Id_n)^\dagger$  and  $\mathbf{1}^{(\cdot)}$  the matrix  $(1 \ \cdots \ 1)^\dagger$ , we can identify:

$$\begin{aligned}H &= \begin{pmatrix} \text{diag}_r K' & \\ & \mathbf{0}_n \end{pmatrix} \\ c &= \begin{pmatrix} -K'' \alpha_1 \\ \vdots \\ -K'' \alpha_r \\ \frac{\lambda}{2} \mathbf{1}^{(n)} \end{pmatrix} \\ G &= \begin{pmatrix} -\text{diag}_{r'} S & -Id_n^{(r')} \\ \text{diag}_{r'} S & -Id_n^{(r')} \end{pmatrix}\end{aligned}$$

where  $\text{diag}_r$  repeats the matrix  $r$  times in the diagonal.

In the case where  $\mathbb{K} = \mathbb{R}$ , we have  $K' = K'' = K$  and  $r' = r$ ; in the case where  $\mathbb{K} = \mathbb{C}$ , we have  $r' = 2r$  and

$$\begin{aligned}K' &= \begin{pmatrix} \Re(K) & -\Im(K) \\ -\Im(K) & \Re(K) \end{pmatrix} \\ K'' &= \begin{pmatrix} \Re(K) & \Im(K) \\ -\Im(K) & \Re(K) \end{pmatrix}\end{aligned}$$

## A.2 Pre-solving the system

In order to speed up, we need to solve the linear systems defined by

$$\begin{pmatrix} H & G^\dagger \\ G & V \end{pmatrix}$$

where  $V$  is a diagonal negative matrix. With a bit of re-ordering, this gives

$$\begin{pmatrix} \text{diag}_r K' & -Id_{nr'} & Id_n^{(r')} & \mathbf{0}_n^{(r')} \\ -\text{diag}_{r'} S & -U & \mathbf{0}_{nr'} & -Id_n^{(r')} \\ \text{diag}_{r'} S & \mathbf{0}_{2nr} & -V & -Id_n^{(r')} \\ \mathbf{0}_n^{(r')\dagger} & -Id_n^{(r')\dagger} & -Id_n^{(2r)\dagger} & \mathbf{0} \end{pmatrix} \begin{pmatrix} x \\ z \\ t \\ y \end{pmatrix} = \begin{pmatrix} a \\ b \\ d \\ c \end{pmatrix}$$

where  $U$  and  $V$  are positive semi-definite (diagonal) matrices of size  $2nr$ .

We want to perform a  $LDL^\dagger$  decomposition of this matrix (a D-Cholesky). Given the structure of the above matrix, we decompose these matrices as

$$L = \begin{pmatrix} L_{11} & & & \\ L_{21} & L_{22} & & \\ L_{31} & L_{32} & L_{33} & \\ L_{41} & L_{42} & L_{43} & L_{44} \end{pmatrix} \text{ and } D = \begin{pmatrix} D_1 & & & \\ & D_2 & & \\ & & D_3 & \\ & & & D_4 \end{pmatrix}$$

A Cholesky decomposition of  $K'$ ,  $AA^\dagger = K'$  gives

$$L_{11} = \text{diag} ( A \ \cdots \ A ) \text{ and } D_1 = Id$$

Note that in the complex field case, we can decompose the problem into

$$\begin{aligned} A_{11}A_{11}^\dagger &= \Re(K) \\ A_{21}A_{11}^\dagger &= -\Im(K) \\ A_{22}A_{22}^\dagger &= \Re(K) - A_{21}A_{21}^\dagger \end{aligned}$$

We now have

$$\begin{pmatrix} L_{21} \\ L_{31} \end{pmatrix} L_{11}^\dagger = \begin{pmatrix} -\text{diag}_{r'} S \\ -\text{diag}_{r'} S \end{pmatrix}$$

which can be solved straightforwardly by first solving  $BA^\dagger = S$ <sup>3</sup>.

$$L_{21} = \text{diag} ( -B \ \cdots \ -B ) \text{ and } L_{31} = \text{diag} ( B \ \cdots \ B )$$

We have  $L_{21}L_{21}^\dagger + L_{22}D_{22}^\dagger L_{22} = -U$ . Since  $U_i + BB^\dagger$  is negative definite. It is sufficient to solve the  $r$  Cholesky decompositions  $L_{22}^{(i)}L_{22}^{(i)\dagger} = U_i + BB^\dagger$ , and then

$$L_{22} = \text{diag} \left( L_{22}^{(1)} \ \cdots \ L_{22}^{(r)} \right) \text{ and } D_2 = -Id_{nr}$$

Then, we find  $L_{32}$  by solving  $L_{31}L_{21}^\dagger + L_{32}D_2L_{22}^\dagger = 0$  which can be solved by solving the  $r$  systems  $L_{32}^{(i)}L_{22}^{(i)\dagger} = -BB^\dagger$ , and then

$$L_{32} = \text{diag} \left( L_{32}^{(1)} \ \cdots \ L_{32}^{(r)} \right)$$

For  $L_{33}$ , we have

$$L_{33}D_3L_{33}^\dagger - L_{32}L_{32}^\dagger + L_{31}L_{31}^\dagger = -V$$

which can be solved by  $r$  D-Cholesky decomposition (all the matrices are block diagonal):

$$L_{33}^{(i)}D_3^{(i)}L_{33}^{(i)\dagger} = -V_i - BB^\dagger + L_{32}^{(i)}L_{32}^{(i)\dagger} \quad (11)$$

---

<sup>3</sup>Note that  $BB^\dagger = K'^{-1}$  since  $BB^\dagger = A^{-\dagger}A^{-1} = (AA^\dagger)^{-1}$  and hence  $BB^\dagger$  is positive definite

In order to simplify the computation, let us prove that the right hand side is negative definite; we have

$$\begin{aligned} L_{32}^{(i)} L_{32}^{(i)\dagger} &= -BB^\dagger L_{22}^{(i)-1} L_{22}^{(i)-\dagger} BB^\dagger \\ &= -BB^\dagger (U_i + BB^\dagger)^{-1} BB^\dagger \\ &= BB^\dagger - U_i + U_i (U_i + BB^\dagger)^{-1} U_i \end{aligned}$$

We have

$$U_i^{-1} (U_i + BB^\dagger) U_i^{-1} = U_i^{-1} + U_i^{-1} BB^\dagger U_i^{-1} \geq U_i^{-1} > 0$$

using the partial order of definite matrices and its properties. This implies that

$$U_i \geq U_i (U_i + BB^\dagger)^{-1} U_i > 0$$

and hence  $-U_i + U_i (U_i + BB^\dagger)^{-1} U_i$  is positive semi-definite, which in turn implies that the right hand side of Eq. (11) is negative definite. This shows that we can find  $L_{33}$  using a Cholesky decomposition

$$L_{33}^{(i)} L_{33}^{(i)\dagger} = V_i + BB^\dagger - L_{32}^{(i)} L_{32}^{(i)\dagger}$$

and that  $D_3 = Id_{rn}$ .

For the fourth row, we first have trivially  $L_{41} = 0$ .

To solve  $L_{42} D_2 L_{22}^\dagger = (-Id_n \ \cdots \ -Id_n)$ , we have to solve  $L_{42}^{(i)} L_{22}^{(i)\dagger} = Id_n$  for  $i = 1 \dots r$ , and then

$$L_{42} = \begin{pmatrix} L_{42}^{(1)} & \cdots & L_{42}^{(r)} \end{pmatrix}$$

To find  $L_{43}$ , we have to solve  $L_{43}^{(i)} L_{33}^{(i)\dagger} = Id_n - L_{42}^{(i)} L_{32}^{(i)\dagger}$  for  $i = 1 \dots r$  and

$$L_{43} = \begin{pmatrix} L_{43}^{(1)} & \cdots & L_{43}^{(r)} \end{pmatrix}$$

Finally, the last equation  $L_{42} D_2 L_{42}^\dagger + L_{43} D_3 L_{43}^\dagger + L_{44} D_4 L_{44}^\dagger = 0$  can be solved by computing the Cholesky decomposition

$$L_{44} L_{44}^\dagger = \sum_{i=1}^r L_{42}^{(i)} L_{42}^{(i)\dagger} + L_{43}^{(i)} L_{43}^{(i)\dagger} \text{ with } D_4 = Id_n$$

Here is the final structure of the decomposition:

$$L = \begin{pmatrix} A_r & & & & \\ -B_r & L_{22} & & & \\ B_r & L_{32} & L_{33} & & \\ \mathbf{0} & L_{42} & L_{43} & L_{44} & \\ & & & & \end{pmatrix}$$

$$D = \begin{pmatrix} Id_{nr} & & & & \\ & -Id_{2rn} & & & \\ & & & & Id_n \end{pmatrix}$$

which gives the following systems to solve:

$$\begin{aligned} Ax'_i &= a_i \\ L_{22}^{(i)} z'_i &= -Bx'_i - c_i \\ L_{33}^{(i)} t'_i &= Bx'_i - L_{32}^{(i)} z'_i - d_i \\ L_{44} y' &= b + \sum_{i=1}^r L_{43}^{(i)} t'_i + L_{42}^{(i)} z'_i \end{aligned}$$

and finally

$$\begin{aligned}L_{44}^\dagger y &= y' \\L_{33}^{(i)\dagger} t_i &= t'_i - L_{43}^{(i)\dagger} y \\L_{22}^{(i)\dagger} y_i &= y'_i - L_{32}^{(i)\dagger} t_i - L_{42}^{(i)\dagger} y \\A^\dagger x_i &= x'_i - B^\dagger t_i + B^\dagger z_i\end{aligned}$$