Self-calibrating Quantum State Estimation

Author: Sim Jun Yan

Supervisor: Prof. Berthold-Georg Englert

A thesis submitted in partial fulfilment of the requirements for the degree of Bachelor of Science with Honours in Physics

Department of Physics

National University of Singapore

Academic Year 2014/2015

Abstract

Self-calibrating quantum state estimation is the procedure of reconstructing an unknown quantum state and certain properties of the measurement devices from the same data. We apply self-calibration to the double crosshair measurement of the BB84 scenario for reconstructing the state and detector efficiencies simultaneously. When we perform maximum-likelihood estimation, we observe multiple maxima in the likelihood function even when the state parameters and detector efficiencies are uniquely determined by the detection probabilities. This problem disappears when prior knowledge of the ratios of detector efficiencies is taken into account. Finally, the maximum-likelihood estimators are endowed with error regions to express the uncertainties associated with them.

Acknowledgements

I would like to express my gratitude to my supervisor, Professor Berthold-Georg Englert for his patience guidance and valuable advices throughout this project. I would also like to thank Dr. Shang Jiangwei for his clear explanations and helps in implementing the codes. Lastly, I would like to thank Jing Hao for some insightful discussions.

Contents

Chapter 1

Introduction

1.1 Quantum state estimation

In the usual scheme of quantum state estimation, measurements with K different outcomes are performed on N identical copies of an unknown quantum state ρ. The measurements are described by a set of positive operators Π_j which form a probability-operator measurement (POM) and satisfy

$$\sum_{k=1}^{K} \Pi_k = 1. \tag{1.1}$$

The j-th operator Π_j corresponds to the j-th outcome of the measurement and the probability for the j-th outcome to occur is given by Born's rule,

$$p_j = \mathrm{tr}\{\Pi_j \rho\} = \langle \Pi_j \rangle \,. \tag{1.2}$$

For a state ρ to be physical, it must be positive semi-definite and have unit trace. These two conditions with Born's rule ensure that all the probabilities are non-negative and they must be summed to one.

If one has the knowledge of the true probabilities, the linear relation in (1.2) can be inverted directly to solve for the unknown state ρ. However,

the true probabilities cannot be known by performing measurements on any finite number of copies. The only data one has is a set of counts,

$$D = \{n_1, ..., n_K\}, \tag{1.3}$$

where n_j denotes the number of times the j-th outcome occur. The relative frequencies are defined as

$$f_j = \frac{n_j}{N}. \tag{1.4}$$

With the measured data D or the relative frequencies, one can try to infer the unknown state ρ using various reconstruction methods.

1.2 Linear inversion

The simplest reconstruction method is linear inversion. Since the relative frequencies converge to the true probabilities when the number of copies, N, approaches infinity, the relative frequencies are good estimates of the true probabilities when N is large. The true probabilities can thus be estimated by the relative frequencies and the state can be solved directly by inverting the linear relation in (1.2). By doing linear inversion, one tries to obtain an estimator $\hat{\rho}_{\mathrm{LI}}$ which satisfy

$$\mathrm{tr}\{\hat{\rho}_{\mathrm{LI}}\Pi_k\} = f_k. \tag{1.5}$$

The advantage of using linear inversion is that it is simple and intuitive. However, the estimator $\hat{\rho}_{\mathrm{LI}}$ is unphysical quite often especially when N is small. The reason of getting an unphysical state is that the relative frequencies only satisfy one constraint,

$$\sum\nolimits_k f_k = 1, \tag{1.6}$$

whereas the probabilities that arise from a physical state satisfy further constraints due to the laws of quantum mechanics. In practice, we might not

always have a large number of copies to perform the measurements. Therefore, we need to use other reconstruction methods that always give us physical states such as maximum-likelihood estimation.

1.3 Maximum-likelihood estimation (MLE)

The main idea of maximum-likelihood estimation is to always choose the state which is most likely to generate the measured data D as our best guess. The likelihood function $\mathcal{L}$ is defined as the conditional probability of getting data D given that the true state is ρ

$$\mathcal{L}(D|\rho) = \prod_{k=1}^{K} {p_k}^{n_k}, \tag{1.7}$$

where $p_k = \mathrm{tr}\{\rho\Pi_k\}$ is the probability of measuring the k-th outcome given the state ρ. The maximum-likelihood estimator $\hat{\rho}_{\mathrm{MLE}}$ is obtained by maximizing the likelihood over the space of physical states,

$$\max_{\rho} \mathcal{L}(D|\rho) = \mathcal{L}(D|\hat{\rho}_{\mathrm{MLE}}). \tag{1.8}$$

In practice, it is more convenient to maximize the log-likelihood function

$$\log \mathcal{L}(D|\rho) = \sum_{k=1}^{K} n_k \log p_k = N \sum_{k=1}^{K} f_k \log p_k, \tag{1.9}$$

instead of the likelihood function itself.

1.4 Incomplete tomography

For a d-dimensional Hilbert space, when the POM consists of at least d^2 elements and d^2 of the elements are linearly independent, the POM is said to

be informationally complete. When the POM is informationally complete, maximal information can be extracted from the state and thus a unique estimator can always be found from the maximum-likelihood estimation. If the POM is not informationally complete, some information about the state will be unavailable and thus the state cannot be completely characterized. As a result, the estimator from the maximum-likelihood estimation is not unique. However, it is still possible to obtain a unique estimator by maximizing the likelihood and von Neumann entropy functionals [17]. In this thesis, we will only characterize the state partially for incomplete tomography by estimating some of the state parameters, the ones that are determined by the data. We do not intend to completely characterize the state for incomplete tomography.

1.5 Self-calibrating quantum state estimation

Quantum state estimation relies on the fact that the measurement setup is well calibrated and all the properties of the measurement devices are known exactly. In practice, it might not always be possible to calibrate the measurement setup exactly. Self-calibrating quantum state estimation deals with this situation where the quantum state and certain unknown properties of the measurement devices are reconstructed using the same data from measurement.

The first advance towards self-calibrating quantum state estimation was done by Mogilevtsev *et al.* [8] where a scheme to reconstruct the states and mismatch between signal and reference states simultaneously was proposed. Later in [7], Mogilevtsev presented a way to reconstruct the state and calibrate single-photon detectors simultaneously with some partial knowledge of the state. The first experimental realization was done by Braczyk *et al.* [5] where an unknown rotation angle of measurement basis and the state parameters are reconstructed simultaneously. In [11], Quesada *et al.* developed a full theoretical formalism to treat unknown parameters in the state and

unitary operations on an equal footing.

In this thesis, we apply self-calibration scheme to the double crosshair measurement of the BB84 [3] scenario of quantum key distribution to reconstruct the state and detector efficiencies simultaneously.

1.6 Error regions

The estimators that we get from maximum-likelihood estimation are point estimators in the state space which represents our best guess of the unknown quantum state. Like other point estimators, the maximum-likelihood estimators suffer from statistical noise. In order to express the uncertainties associated with the maximum-likelihood estimators, they have to be endowed with error regions which are the generalization of error bars in higher dimensional space. Various methods to construct the error regions have been proposed. However, they either requires a large number of data [2], involve data bootstrapping [15], or consider all the data that one could possibly have observed [4]. In [13], Shang *et al.* proposed the maximum-likelihood regions (MLR) and smallest credible regions (SCR) as the optimal error regions. The optimal error regions proposed do not have the problems stated above and can be constructed only by the data that have actually been observed.

MLR and SCR are both proved to be bounded-likelihood regions (BLR) which contains all the states with point likelihood larger than some threshold value,

$$\mathcal{R}_\lambda = \{\rho \,|\, \mathcal{L}(D|\rho) > \lambda \mathcal{L}(D|\hat{\rho}_{\mathrm{MLE}})\} \tag{1.10}$$

This simple characterization provides a simple way to report the error regions which is just by reporting the value of λ.

SCR is the region with the smallest size for a given credibility. The most natural way to define the size of an error region is by its prior probability i.e. the probability of finding the true state in the region before any data was taken. On the other hand, credibility is defined as the posterior probability

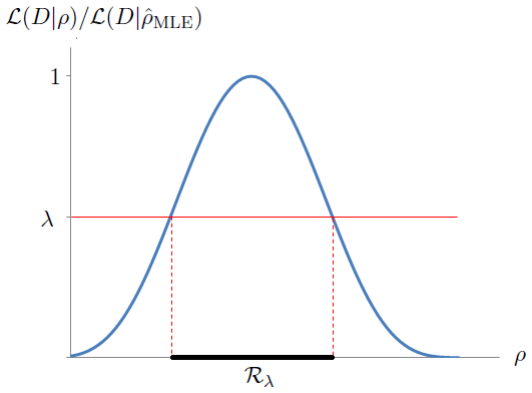


Figure 1.1: Illustration of a BLR: The BLR $\mathcal{R}_\lambda$ contains all the states with likelihood greater than $\lambda\mathcal{L}(D|\hat{\rho}_{\mathrm{MLE}})$.

i.e. the probability that the true state is contained in the region, conditioned on the measured data. In the next two subsections, we discuss how the size and credibility of a region can be computed.

1.6.1 Probability space and reconstruction space

In a K-outcome measurement, the permissible probabilities $p = (p_1, p_2, ..., p_K)$ are probabilities which there exist a state ρ such that Born's rule (1.2) holds. The probability space comprises all the permissible probabilities. For a probability to be permissible, it must satisfy some basic constraints $w_{\mathrm{basic}}(p)$ and some other constraints imposed by quantum mechanics $w_{\mathrm{qu}}(p)$ which might not be easy to find. Obviously, the basic constraint are given by

$$w_{\mathrm{basic}}(p) = \eta(p_1)\eta(p_2)...\eta(p_K)\delta\left(1 - \sum_{k=1}^{K} p_k\right). \tag{1.11}$$

The full constraints are given by

$$w_{\mathrm{cstr}}(p) = w_{\mathrm{basic}}(p)w_{\mathrm{qu}}(p). \tag{1.12}$$

In an informationally complete measurement, there is a one-to-one mapping from the probability space to state space. For informationally incomplete measurement (refer to section 1.4), each permissible p maps to a class of states. A reconstruction space $\mathcal{R}_0$ which contains exactly one ρ for each permissible p can be constructed by choosing a representative for each class of states. The point estimators refer to states in $\mathcal{R}_0$ and error regions refer to sets of states in $\mathcal{R}_0$.

1.6.2 Size and credibility

Let $(\mathrm{d}\rho)$ denotes the volume element of infinitesimal vicinity of state ρ in the reconstruction space $\mathcal{R}_0$. The state can be parametrized in any way although it is most common to parametrize by the probabilities,

$$(\mathrm{d}\rho) = (\mathrm{d}p)w_0(p), \tag{1.13}$$

where

$$(\mathrm{d}p) = \mathrm{d}p_1\mathrm{d}p_2...\mathrm{d}p_K w_{\mathrm{cstr}}(p), \tag{1.14}$$

and $w_0(p)$ is the unnormalized prior density. The choice of prior density $w_0(p)$ is not unique and the simplest is the primitive prior,

$$w_{\mathrm{primitive}}(p) = 1. \tag{1.15}$$

By working with primitive prior, the density is uniform in the probability space. The unnormalized posterior density $w_D(p)$ can be obtained by multiplying the likelihood with the prior density,

$$w_D(p) = w_0(p)\mathcal{L}(D|p). \tag{1.16}$$

The size $s_{\mathcal{R}}$ and credibility $c_{\mathcal{R}}$ of a region $\mathcal{R}$ can then be defined in terms

of the prior and posterior density,

$$s_{\mathcal{R}} = \frac{\int_{\mathcal{R}}(\mathrm{d}\rho)}{\int_{\mathcal{R}_0}(\mathrm{d}\rho)} = \frac{\int_{\mathcal{R}}(\mathrm{d}p)w_0(p)}{\int_{\mathcal{R}_0}(\mathrm{d}p)w_0(p)}, \tag{1.17}$$

$$c_{\mathcal{R}} = \frac{\int_{\mathcal{R}}(\mathrm{d}p)w_D(p)}{\int_{\mathcal{R}_0}(\mathrm{d}p)w_D(p)}. \tag{1.18}$$

An alternative way to compute credibility is given in [13]. Once the size is known as a function of λ, credibility can also be computed as a function of λ by

$$c_\lambda = \frac{\lambda s_\lambda + \int_\lambda^1 d\lambda' s_{\lambda'}}{\int_0^1 d\lambda' s_{\lambda'}}. \tag{1.19}$$

1.6.3 Reporting the SCRs

The SCRs with different credibility can be reported once the size and credibility are calculated by (1.17) and (1.18) or (1.19). To report a SCR with credibility c, we first find the value λ_0 such that $c_{\lambda=\lambda_0} = c$. The SCR is just the BLR with $\lambda = \lambda_0$ and its size is given by $s_{\lambda=\lambda_0}$.

Chapter 2

Double crosshair measurement of BB84

Figure 2.1 shows the double crosshair measurement of the BB84 scenario of quantum key distribution. In this measurement, a photon source emits entangled pairs of photons and each of the photons passes through a 50-50 beam splitter. If the photon is transmitted, its polarization state will be measured in σ_z basis (horizontal-vertical or H/V basis). If the photon is reflected, its polarization state will be measured in σ_x basis (diagonal-antidiagonal or D/A basis). All the detectors in this measurement are imperfect and their efficiencies are denoted by η. In this measurement, we can observe either coincidence clicks or clicks on one side. It is also possible that detectors on both sides do not click even when photon pairs are emitted. We assume that the total number of photon pairs emitted is known so that the number of undetected photon pairs is also known. Thus, this measurement can result in twenty-five different outcomes. The probabilities of all the outcomes are shown in table 2.1 and (2.1).

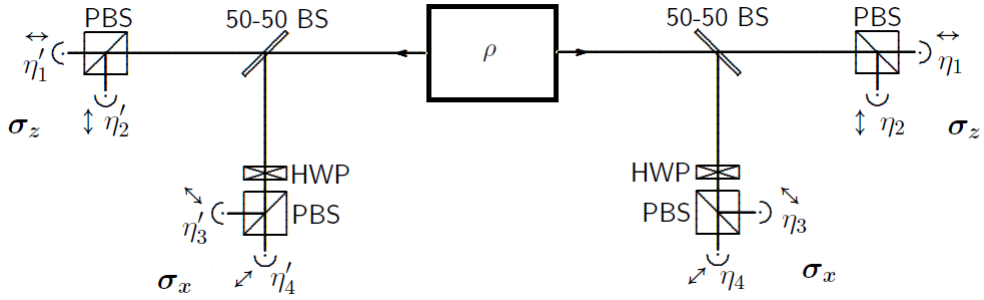


Figure 2.1: Double crosshair measurement of BB84. Measurements are done on two bases for both photons. All the detectors are imperfect and their efficiencies are denoted by η.

Left \ Right	H	V	D	A	no reading	
H	$p(HH)$	$p(HV)$	$p(HD)$	$p(HA)$	$p(H0)$	$p_L(H)$
V	$p(VH)$	$p(VV)$	$p(VD)$	$p(VA)$	$p(V0)$	$p_L(V)$
D	$p(DH)$	$p(DV)$	$p(DD)$	$p(DA)$	$p(D0)$	$p_L(D)$
A	$p(AH)$	$p(AV)$	$p(AD)$	$p(AA)$	$p(A0)$	$p_L(A)$
no reading	$p(0H)$	$p(0V)$	$p(0D)$	$p(0A)$	$p(00)$	$p_L(0)$
	$p_R(H)$	$p_R(V)$	$p_R(D)$	$p_R(A)$	$p_R(0)$	1

Table 2.1: Detection probabilities for the double crosshair measurement. The last column and last row are the marginal probabilities. The probabilities of detecting no clicks on either side are not stated in (2.1) since they have lengthy expressions. However, they can be deduced from the marginal probabilities and the probabilities of coincidence clicks.

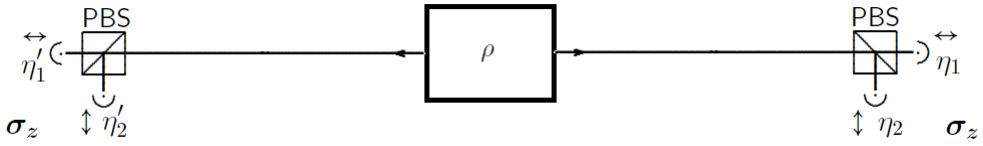


Figure 2.2: Simplified version of the double crosshair measurement. Measurements are done on one basis for both photons. All the detectors are imperfect and their efficiencies are denoted by η.

$$
\begin{aligned}
p(HH) &= \frac{\eta_1\eta_1'}{4}\,\langle\frac{1+\sigma_z}{2}\otimes\frac{1+\sigma_z}{2}\rangle\,, & p(HV) &= \frac{\eta_2\eta_1'}{4}\,\langle\frac{1+\sigma_z}{2}\otimes\frac{1-\sigma_z}{2}\rangle\,,\\
p(HD) &= \frac{\eta_3\eta_1'}{4}\,\langle\frac{1+\sigma_z}{2}\otimes\frac{1+\sigma_x}{2}\rangle\,, & p(HA) &= \frac{\eta_4\eta_1'}{4}\,\langle\frac{1+\sigma_z}{2}\otimes\frac{1-\sigma_x}{2}\rangle\,,\\
p(VH) &= \frac{\eta_1\eta_2'}{4}\,\langle\frac{1-\sigma_z}{2}\otimes\frac{1+\sigma_z}{2}\rangle\,, & p(VV) &= \frac{\eta_2\eta_2'}{4}\,\langle\frac{1-\sigma_z}{2}\otimes\frac{1-\sigma_z}{2}\rangle\,,\\
p(VD) &= \frac{\eta_3\eta_2'}{4}\,\langle\frac{1-\sigma_z}{2}\otimes\frac{1+\sigma_x}{2}\rangle\,, & p(VA) &= \frac{\eta_4\eta_2'}{4}\,\langle\frac{1-\sigma_z}{2}\otimes\frac{1-\sigma_x}{2}\rangle\,,\\
p(DH) &= \frac{\eta_1\eta_3'}{4}\,\langle\frac{1+\sigma_x}{2}\otimes\frac{1+\sigma_z}{2}\rangle\,, & p(DV) &= \frac{\eta_2\eta_3'}{4}\,\langle\frac{1+\sigma_x}{2}\otimes\frac{1-\sigma_z}{2}\rangle\,,\\
p(DD) &= \frac{\eta_3\eta_3'}{4}\,\langle\frac{1+\sigma_x}{2}\otimes\frac{1+\sigma_x}{2}\rangle\,, & p(DA) &= \frac{\eta_4\eta_3'}{4}\,\langle\frac{1+\sigma_x}{2}\otimes\frac{1-\sigma_x}{2}\rangle\,,\\
p(AH) &= \frac{\eta_1\eta_4'}{4}\,\langle\frac{1-\sigma_x}{2}\otimes\frac{1+\sigma_z}{2}\rangle\,, & p(AV) &= \frac{\eta_2\eta_4'}{4}\,\langle\frac{1-\sigma_x}{2}\otimes\frac{1-\sigma_z}{2}\rangle\,,\\
p(AD) &= \frac{\eta_3\eta_4'}{4}\,\langle\frac{1-\sigma_x}{2}\otimes\frac{1+\sigma_x}{2}\rangle\,, & p(AA) &= \frac{\eta_4\eta_4'}{4}\,\langle\frac{1-\sigma_x}{2}\otimes\frac{1-\sigma_x}{2}\rangle\,,\\
p_L(H) &= \frac{\eta_1'}{2}\,\langle\frac{1+\sigma_z}{2}\otimes 1\rangle\,, & p_L(V) &= \frac{\eta_2'}{2}\,\langle\frac{1-\sigma_z}{2}\otimes 1\rangle\,,\\
p_L(D) &= \frac{\eta_3'}{2}\,\langle\frac{1+\sigma_x}{2}\otimes 1\rangle\,, & p_L(A) &= \frac{\eta_4'}{2}\,\langle\frac{1-\sigma_x}{2}\otimes 1\rangle\,,\\
p_R(H) &= \frac{\eta_1}{2}\,\langle 1\otimes\frac{1+\sigma_z}{2}\rangle\,, & p_R(V) &= \frac{\eta_2}{2}\,\langle 1\otimes\frac{1-\sigma_z}{2}\rangle\,,\\
p_R(D) &= \frac{\eta_3}{2}\,\langle 1\otimes\frac{1+\sigma_x}{2}\rangle\,, & p_R(A) &= \frac{\eta_4}{2}\,\langle 1\otimes\frac{1-\sigma_x}{2}\rangle\,. \qquad (2.1)
\end{aligned}
$$

We could also consider a simplified version of the double crosshair measurement of BB84 which is shown in figure 2.2. Similarly, entangled pairs of photons are emitted by the source. However, polarization state of photons is only measured in σ_z basis (horizontal-vertical or H/V basis). This measurement can result in nine different outcomes. The probabilities of all the outcomes are shown in table 2.2 and (2.2).

Left \ Right	H	V	no reading	
H	$p(HH)$	$p(HV)$	$p(H0)$	$p_L(H)$
V	$p(VH)$	$p(VV)$	$p(V0)$	$p_L(V)$
no reading	$p(0H)$	$p(0V)$	$p(00)$	$p_L(0)$
	$p_R(H)$	$p_R(V)$	$p_R(0)$	1

Table 2.2: Detection probabilities for the simplified measurement. The last column and last row are the marginal probabilities. The probabilities of detecting no clicks on either side are not stated in (2.2) since they have lengthy expressions. However, they can be deduced from the marginal probabilities and the probabilities of coincidence clicks.

$$
\begin{aligned}
p(HH) &= \eta_1\eta_1' \langle \frac{1+\sigma_z}{2} \otimes \frac{1+\sigma_z}{2} \rangle , \qquad & p(HV) &= \eta_2\eta_1' \langle \frac{1+\sigma_z}{2} \otimes \frac{1-\sigma_z}{2} \rangle , \\
p(VH) &= \eta_1\eta_2' \langle \frac{1-\sigma_z}{2} \otimes \frac{1+\sigma_z}{2} \rangle , \qquad & p(VV) &= \eta_2\eta_2' \langle \frac{1-\sigma_z}{2} \otimes \frac{1-\sigma_z}{2} \rangle , \\
p_L(H) &= \eta_1' \langle \frac{1+\sigma_z}{2} \otimes 1 \rangle , & p_L(V) &= \eta_2' \langle \frac{1-\sigma_z}{2} \otimes 1 \rangle , \\
p_R(H) &= \eta_1 \langle 1 \otimes \frac{1+\sigma_z}{2} \rangle , & p_R(V) &= \eta_2 \langle 1 \otimes \frac{1-\sigma_z}{2} \rangle . \qquad (2.2)
\end{aligned}
$$

2.1 Determining the feasibility of self-calibration

The first step to determine whether the self-calibration scheme is feasible is to find out whether it is possible to solve for all the unknowns uniquely given that the exact detection probabilities are known. The expressions of all the detection probabilities for the double crosshair measurement are shown in (2.1). All the state parameters can then be expressed in terms of the detection probabilities and the detector efficiencies in (2.4). These expressions could then turn into some linear equations of reciprocal of efficiencies. For example, from the first and fifth expression of $\langle 1 \otimes \sigma_z \rangle$ in (2.4a), we get

$$2\frac{p(HH)}{\eta_1'} + 2\frac{p(VH)}{\eta_2'} = p_R(H). \tag{2.3}$$

All these linear equations can be written in matrix forms in (2.5).

$$\begin{aligned} < 1 \otimes \sigma_z > &= \frac{8p(HH)}{\eta_1\eta_1'} + \frac{8p(VH)}{\eta_1\eta_2'} - 1 \\ &= \frac{8p(DH)}{\eta_1\eta_3'} + \frac{8p(AH)}{\eta_1\eta_4'} - 1 \\ &= 1 - \frac{8p(HV)}{\eta_2\eta_1'} - \frac{8p(VV)}{\eta_2\eta_2'} \\ &= 1 - \frac{8p(DV)}{\eta_2\eta_3'} - \frac{8p(AV)}{\eta_2\eta_4'} \\ &= \frac{4p_R(H)}{\eta_1} - 1 \\ &= 1 - \frac{4p_R(V)}{\eta_2}, \end{aligned} \tag{2.4a}$$

$$\begin{aligned} < 1 \otimes \sigma_x > &= \frac{8p(HD)}{\eta_3\eta_1'} + \frac{8p(VD)}{\eta_3\eta_2'} - 1 \\ &= \frac{8p(DD)}{\eta_3\eta_3'} + \frac{8p(AD)}{\eta_3\eta_4'} - 1 \\ &= 1 - \frac{8p(HA)}{\eta_4\eta_1'} - \frac{8p(VA)}{\eta_4\eta_2'} \\ &= 1 - \frac{8p(DA)}{\eta_4\eta_3'} - \frac{8p(AA)}{\eta_4\eta_4'} \\ &= \frac{4p_R(D)}{\eta_3} - 1 \\ &= 1 - \frac{4p_R(A)}{\eta_4}, \end{aligned} \tag{2.4b}$$

$$
\begin{aligned}
<\sigma_z \otimes 1> &= \frac{8p(HH)}{\eta_1\eta_1'} + \frac{8p(HV)}{\eta_2\eta_1'} - 1 \\
&= \frac{8p(HD)}{\eta_3\eta_1'} + \frac{8p(HA)}{\eta_4\eta_1'} - 1 \\
&= 1 - \frac{8p(VH)}{\eta_1\eta_2'} - \frac{8p(VV)}{\eta_2\eta_2'} \\
&= 1 - \frac{8p(VD)}{\eta_3\eta_2'} - \frac{8p(VA)}{\eta_4\eta_2'} \\
&= \frac{4p_L(H)}{\eta_1'} - 1 \\
&= 1 - \frac{4p_L(V)}{\eta_2'}, \qquad (2.4c)
\end{aligned}
$$

$$
\begin{aligned}
<\sigma_z \otimes \sigma_z> &= \frac{8p(HH)}{\eta_1\eta_1'} + \frac{8p(VV)}{\eta_2\eta_2'} - 1 \\
&= 1 - \frac{8p(HV)}{\eta_2\eta_1'} - \frac{8p(VH)}{\eta_1\eta_2'}, \qquad (2.4e)
\end{aligned}
$$

$$
\begin{aligned}
<\sigma_z \otimes \sigma_x> &= \frac{8p(HD)}{\eta_3\eta_1'} + \frac{8p(VA)}{\eta_4\eta_2'} - 1 \\
&= 1 - \frac{8p(HA)}{\eta_4\eta_1'} - \frac{8p(VD)}{\eta_3\eta_2'}, \qquad (2.4f)
\end{aligned}
$$

$$
\begin{aligned}
<\sigma_x \otimes 1> &= \frac{8p(DH)}{\eta_1\eta_3'} + \frac{8p(DV)}{\eta_2\eta_3'} - 1 \\
&= \frac{8p(DD)}{\eta_3\eta_3'} + \frac{8p(DA)}{\eta_4\eta_3'} - 1 \\
&= 1 - \frac{8p(AH)}{\eta_1\eta_4'} - \frac{8p(AV)}{\eta_2\eta_4'} \\
&= 1 - \frac{8p(AD)}{\eta_3\eta_4'} - \frac{8p(AA)}{\eta_4\eta_4'} \\
&= \frac{4p_L(D)}{\eta_3'} - 1 \\
&= 1 - \frac{4p_L(A)}{\eta_4'}, \qquad (2.4d)
\end{aligned}
$$

$$
\begin{aligned}
<\sigma_x \otimes \sigma_x> &= \frac{8p(DD)}{\eta_3\eta_3'} + \frac{8p(AA)}{\eta_4\eta_4'} - 1 \\
&= 1 - \frac{8p(DA)}{\eta_4\eta_3'} - \frac{8p(AD)}{\eta_3\eta_4'}, \qquad (2.4g)
\end{aligned}
$$

$$
\begin{aligned}
<\sigma_x \otimes \sigma_z> &= \frac{8p(DH)}{\eta_1\eta_3'} + \frac{8p(AV)}{\eta_2\eta_4'} - 1 \\
&= 1 - \frac{8p(DV)}{\eta_2\eta_3'} - \frac{8p(AH)}{\eta_1\eta_4'}. \qquad (2.4h)
\end{aligned}
$$

$$
2 \begin{pmatrix} p(HH) & p(VH) \\ p(HV) & p(VV) \\ p(HD) & p(VD) \\ p(HA) & p(VA) \\ p_L(H) & p_L(V) \end{pmatrix} \begin{pmatrix} \frac{1}{\eta_1'} \\ \frac{1}{\eta_2'} \end{pmatrix} = \begin{pmatrix} p_R(H) \\ p_R(V) \\ p_R(D) \\ p_R(A) \\ 1 \end{pmatrix}, \qquad (2.5a)
$$

$$2\begin{pmatrix} p(DH) & p(AH) \\ p(DV) & p(AV) \\ p(DD) & p(AD) \\ p(DA) & p(AA) \\ p_L(D) & p_L(A) \end{pmatrix}\begin{pmatrix} \frac{1}{\eta_3'} \\ \frac{1}{\eta_4'} \end{pmatrix} = \begin{pmatrix} p_R(H) \\ p_R(V) \\ p_R(D) \\ p_R(A) \\ 1 \end{pmatrix}, \tag{2.5b}$$

$$2\begin{pmatrix} p(HH) & p(HV) \\ p(VH) & p(VV) \\ p(DH) & p(DV) \\ p(AH) & p(AV) \\ p_R(H) & p_R(V) \end{pmatrix}\begin{pmatrix} \frac{1}{\eta_1} \\ \frac{1}{\eta_2} \end{pmatrix} = \begin{pmatrix} p_L(H) \\ p_L(V) \\ p_L(D) \\ p_L(A) \\ 1 \end{pmatrix}, \tag{2.5c}$$

$$2\begin{pmatrix} p(HD) & p(HA) \\ p(VD) & p(VA) \\ p(DD) & p(DA) \\ p(AD) & p(AA) \\ p_R(D) & p_R(A) \end{pmatrix}\begin{pmatrix} \frac{1}{\eta_3} \\ \frac{1}{\eta_4} \end{pmatrix} = \begin{pmatrix} p_L(H) \\ p_L(V) \\ p_L(D) \\ p_L(A) \\ 1 \end{pmatrix}. \tag{2.5d}$$

At first glance, it seems that we have more equations than unknowns i.e. an overdetermined system. However, this overdetermined system is consistent if the probabilities in this system of linear equations are the exact detection probabilities. Thus, all the detector efficiencies can be solved uniquely.[1] Once the detector efficiencies are known, all the state parameters can then be solved uniquely using (2.4) or treating (2.1) as a system of linear equations of the state parameters.

For the simplified measurement, all the state parameters and detector efficiencies can also be solved using the same approach. From the expressions of the detection probabilities in (2.2), the expressions for the state parameters, and the system of linear equations of reciprocal of efficiencies can be derived in (2.6) and (2.7).

[1]The solution of (2.5) is not unique if the two columns of the coefficient matrix are parallel to each other. In practice, the two columns will not be exactly parallel to each other. Thus, we can safely assume that the solution is unique.

We have thus shown that the state parameters and the detector efficiencies can be solved uniquely for both measurements if the exact detection probabilities are known. This implies that the self-calibration scheme for both measurements is feasible theoretically. However, we do not know the exact detection probabilities in practice. Thus, we would need to use reconstruction methods such as those stated in sections 1.2 and 1.3 to estimate the state parameters and detector efficiencies.

$$\begin{aligned} < 1 \otimes \sigma_z > &= \frac{2p(HH)}{\eta_1\eta_1'} + \frac{2p(VH)}{\eta_1\eta_2'} - 1 \\ &= 1 - \frac{2p(HV)}{\eta_2\eta_1'} - \frac{2p(VV)}{\eta_2\eta_2'} \\ &= \frac{2p_R(H)}{\eta_1} - 1 \\ &= 1 - \frac{2p_R(V)}{\eta_2}, \end{aligned} \tag{2.6a}$$

$$\begin{aligned} < \sigma_z \otimes 1 > &= \frac{2p(HH)}{\eta_1\eta_1'} + \frac{2p(HV)}{\eta_2\eta_1'} - 1 \\ &= 1 - \frac{2p(VH)}{\eta_1\eta_2'} - \frac{2p(VV)}{\eta_2\eta_2'} \\ &= \frac{2p_L(H)}{\eta_1'} - 1 \\ &= 1 - \frac{2p_L(V)}{\eta_2'}, \end{aligned} \tag{2.6b}$$

$$\begin{aligned} < \sigma_z \otimes \sigma_z > &= \frac{2p(HH)}{\eta_1\eta_1'} + \frac{2p(VV)}{\eta_2\eta_2'} - 1 \\ &= 1 - \frac{2p(HV)}{\eta_2\eta_1'} - \frac{2p(VH)}{\eta_1\eta_2'}. \end{aligned} \tag{2.6c}$$

$$\begin{pmatrix} p(HH) & p(VH) \\ p(HV) & p(VV) \\ p_L(H) & p_L(V) \end{pmatrix} \begin{pmatrix} \frac{1}{\eta_1'} \\ \frac{1}{\eta_2'} \end{pmatrix} = \begin{pmatrix} p_R(H) \\ p_R(V) \\ 1 \end{pmatrix}, \tag{2.7a}$$

$$\begin{pmatrix} p(HH) & p(HV) \\ p(VH) & p(VV) \\ p_R(H) & p_R(V) \end{pmatrix} \begin{pmatrix} \frac{1}{\eta_1} \\ \frac{1}{\eta_2} \end{pmatrix} = \begin{pmatrix} p_L(H) \\ p_L(V) \\ 1 \end{pmatrix}. \tag{2.7b}$$

Chapter 3

Point estimators

3.1 Linear inversion

As stated in section 1.2, linear inversion is the simplest reconstruction method that one could use. In doing linear inversion, all the detection probabilities in (2.5) and (2.7) are estimated by the relative frequencies measured and the state parameters and detector efficiencies are solved directly using the same method in last chapter. When the detection probabilities are estimated by the relative frequencies, the overdetermined systems of linear equations in (2.5) and (2.7) are inconsistent. However, an approximate solution can still be obtained using the least-square method. For example, let's denote the leftmost matrix and the matrix on right hand side in (2.7a) as A and B respectively,

$$A \begin{pmatrix} \frac{1}{\eta_1'} \\ \\ \frac{1}{\eta_2'} \end{pmatrix} = B, \tag{3.1}$$

the least-square solution[1] is given by

$$\begin{pmatrix} \frac{1}{\eta_1'} \\ \\ \frac{1}{\eta_2'} \end{pmatrix} = (A^T A)^{-1} A^T B. \tag{3.2}$$

As stated in section 1.2, the estimated state and the detector efficiencies from linear inversion might be unphysical sometimes. Figure 3.1 shows the result from doing linear inversion for 1000 simulations with different number of photon pairs measured, N. From the figure, the estimated state is unphysical quite often when N is small. The percentage discrepancies of the estimated efficiencies are also quite large when N is small because we often get some unphysical values for the estimated efficiencies when N is small. Since we might not always have a large of number of photon pairs to perform the measurements, linear inversion might not be the suitable reconstruction method to use.

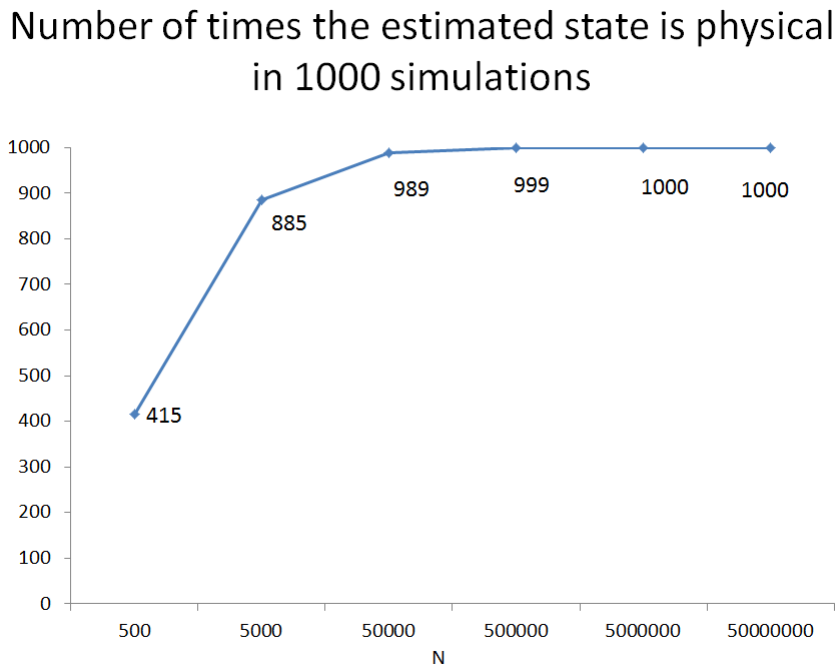
3.2 Maximum-likelihood estimation

Maximum-likelihood estimation which is introduced in section 1.3 is a reconstruction method which guarantees to give a physical estimated state and physical detector efficiencies. In this self-calibration scheme, the likelihood function in (1.7) is now defined as a function of both state and efficiencies since the detection probabilities depend on both state and efficiencies. The maximum-likelihood estimator for state $\hat{\rho}_{\mathrm{MLE}}$ and efficiencies $\hat{\boldsymbol{\eta}}_{\mathrm{MLE}}$ are obtained by maximizing the log-likelihood over the space of physical states and efficiencies,

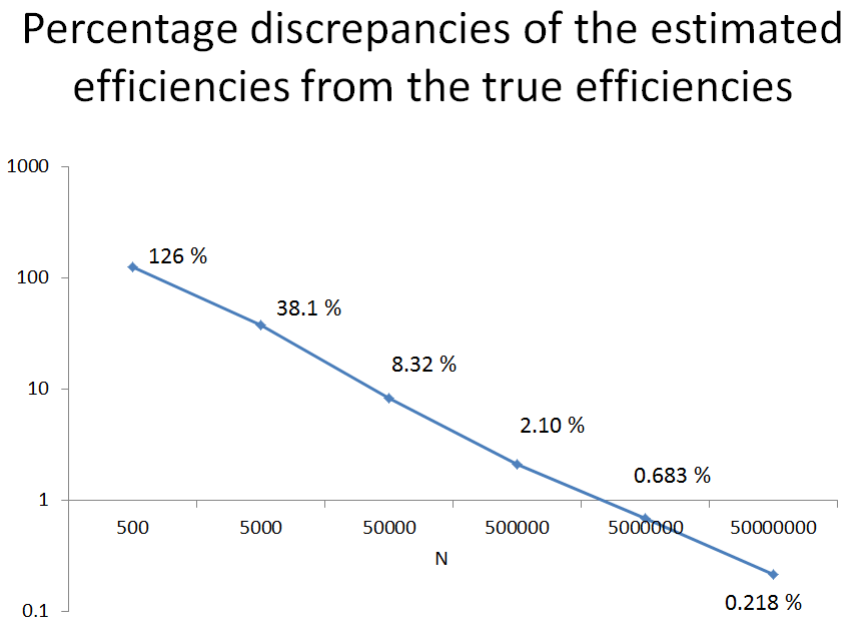
$$\max_{\rho,\boldsymbol{\eta}} \log \mathcal{L}(D|\rho,\boldsymbol{\eta}) = \log \mathcal{L}(D|\hat{\rho}_{\mathrm{MLE}}, \hat{\boldsymbol{\eta}}_{\mathrm{MLE}}). \tag{3.3}$$

In general, it is hard to find a closed formula for the maximum-likelihood

[1] The derivation of this expression is given in appendix A.



(a)



(b)

Figure 3.1: Result from doing linear inversion for the double crosshair measurement in 1000 simulations with different number of photon pairs measured N.

estimator $\hat{\rho}_{\mathrm{MLE}}$ and $\hat{\boldsymbol{\eta}}_{\mathrm{MLE}}$. In the case of ideal detectors, the log-likelihood function is concave and thus the problem of finding the maximum of the log-likelihood function is a convex optimization problem which can be solved by the direct gradient method. The algorithm to find the maximum of the log-likelihood function using direct gradient method has been developed by Hradil [6, 16].

When the detectors have imperfect efficiencies, the likelihood function have a much more complicated expression and it is very hard to determine whether the log-likelihood is concave. At this stage, we develop the algorithm to find the maximum-likelihood estimator using direct gradient method even though we are not sure whether the log-likelihood is concave. There are only slight modifications to the algorithm in [16].

All the efficiencies are parametrized by some parameters y to guarantee that they are always some values between zero and one. For the simplified measurement,

$$\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_1', \eta_2')^T = (\sin^2(y_1), \sin^2(y_2), \sin^2(y_3), \sin^2(y_4))^T, \tag{3.4}$$

whereas for the double crosshair measurement,

$$\boldsymbol{\eta} = (\eta_1, ..., \eta_4, \eta_1', ..., \eta_4')^T = (\sin^2(y_1), ..., \sin^2(y_4), \sin^2(y_5), ..., \sin^2(y_8))^T. \tag{3.5}$$

The derivatives of log-likelihood with respect to the y parameters for simplified measurement and double crosshair measurement are derived in (3.6) and (3.7) respectively.

$$\frac{\partial \log \mathcal{L}}{\partial y_1} = 2 \cot y_1 \Big(n_{HH} + n_{VH} + n_{0H} - \frac{n_{H0}}{p_{H0}} p_{HH} - \frac{n_{V0}}{p_{V0}} p_{VH} - \frac{n_{00}}{p_{00}} p_{0H} \Big), \tag{3.6a}$$

$$\frac{\partial \log \mathcal{L}}{\partial y_2} = 2 \cot y_2 \Big(n_{HV} + n_{VV} + n_{0V} - \frac{n_{H0}}{p_{H0}} p_{HV} - \frac{n_{V0}}{p_{V0}} p_{VV} - \frac{n_{00}}{p_{00}} p_{0V} \Big), \tag{3.6b}$$

$$\frac{\partial \log \mathcal{L}}{\partial y_3} = 2\cot y_3 \Big(n_{HH} + n_{HV} + n_{H0} - \frac{n_{0H}}{p_{0H}} p_{HH} - \frac{n_{0V}}{p_{0V}} p_{HV} - \frac{n_{00}}{p_{00}} p_{H0} \Big), \quad (3.6c)$$

$$\frac{\partial \log \mathcal{L}}{\partial y_4} = 2\cot y_4 \Big(n_{VH} + n_{VV} + n_{V0} - \frac{n_{0H}}{p_{0H}} p_{VH} - \frac{n_{0V}}{p_{0V}} p_{VV} - \frac{n_{00}}{p_{00}} p_{V0} \Big). \quad (3.6d)$$

$$\frac{\partial \log \mathcal{L}}{\partial y_1} = 2\cot y_1 \Big(n_{HH} + n_{VH} + n_{DH} + n_{AH} + n_{0H} - \frac{n_{H0}}{p_{H0}} p_{HH} - \frac{n_{V0}}{p_{V0}} p_{VH} - \frac{n_{D0}}{p_{D0}} p_{DH} - \frac{n_{A0}}{p_{A0}} p_{AH} - \frac{n_{00}}{p_{00}} p_{0H} \Big), \quad (3.7a)$$

$$\frac{\partial \log \mathcal{L}}{\partial y_2} = 2\cot y_2 \Big(n_{HV} + n_{VV} + n_{DV} + n_{AV} + n_{0V} - \frac{n_{H0}}{p_{H0}} p_{HV} - \frac{n_{V0}}{p_{V0}} p_{VV} - \frac{n_{D0}}{p_{D0}} p_{DV} - \frac{n_{A0}}{p_{A0}} p_{AV} - \frac{n_{00}}{p_{00}} p_{0V} \Big), \quad (3.7b)$$

$$\frac{\partial \log \mathcal{L}}{\partial y_3} = 2\cot y_3 \Big(n_{HD} + n_{VD} + n_{DD} + n_{AD} + n_{0D} - \frac{n_{H0}}{p_{H0}} p_{HD} - \frac{n_{V0}}{p_{V0}} p_{VD} - \frac{n_{D0}}{p_{D0}} p_{DD} - \frac{n_{A0}}{p_{A0}} p_{AD} - \frac{n_{00}}{p_{00}} p_{0D} \Big), \quad (3.7c)$$

$$\frac{\partial \log \mathcal{L}}{\partial y_4} = 2\cot y_4 \Big(n_{HA} + n_{VA} + n_{DA} + n_{AA} + n_{0A} - \frac{n_{H0}}{p_{H0}} p_{HA} - \frac{n_{V0}}{p_{V0}} p_{VA} - \frac{n_{D0}}{p_{D0}} p_{DA} - \frac{n_{A0}}{p_{A0}} p_{AA} - \frac{n_{00}}{p_{00}} p_{0A} \Big), \quad (3.7d)$$

$$\frac{\partial \log \mathcal{L}}{\partial y_5} = 2\cot y_5 \Big(n_{HH} + n_{HV} + n_{HD} + n_{HA} + n_{H0} - \frac{n_{0H}}{p_{0H}} p_{HH} - \frac{n_{0V}}{p_{0V}} p_{HV} - \frac{n_{0D}}{p_{0D}} p_{HD} - \frac{n_{0A}}{p_{0A}} p_{HA} - \frac{n_{00}}{p_{00}} p_{H0} \Big), \quad (3.7e)$$

$$\frac{\partial \log \mathcal{L}}{\partial y_6} = 2\cot y_6\Big(n_{VH} + n_{VV} + n_{VD} + n_{VA} + n_{V0} - \frac{n_{0H}}{p_{0H}}p_{VH} - \frac{n_{0V}}{p_{0V}}p_{VV} - \frac{n_{0D}}{p_{0D}}p_{VD} - \frac{n_{0A}}{p_{0A}}p_{VA} - \frac{n_{00}}{p_{00}}p_{V0}\Big), \tag{3.7f}$$

$$\frac{\partial \log \mathcal{L}}{\partial y_7} = 2\cot y_7\Big(n_{DH} + n_{DV} + n_{DD} + n_{DA} + n_{D0} - \frac{n_{0H}}{p_{0H}}p_{DH} - \frac{n_{0V}}{p_{0V}}p_{DV} - \frac{n_{0D}}{p_{0D}}p_{DD} - \frac{n_{0A}}{p_{0A}}p_{DA} - \frac{n_{00}}{p_{00}}p_{D0}\Big), \tag{3.7g}$$

$$\frac{\partial \log \mathcal{L}}{\partial y_8} = 2\cot y_8\Big(n_{AH} + n_{AV} + n_{AD} + n_{AA} + n_{A0} - \frac{n_{0H}}{p_{0H}}p_{AH} - \frac{n_{0V}}{p_{0V}}p_{AV} - \frac{n_{0D}}{p_{0D}}p_{AD} - \frac{n_{0A}}{p_{0A}}p_{AA} - \frac{n_{00}}{p_{00}}p_{A0}\Big). \tag{3.7h}$$

The full algorithm is stated as follows:

MLE algorithm using direct gradient method

Start from the maximally mixed state $\rho_0 = 1/4$ and $\boldsymbol{y}_0 = (\pi/4, ..., \pi/4)^T$ with $j = 0$, and a small fixed value of ε,

1. Calculate R_j according to

$$R = \sum_{k=0}^{K-1} \frac{f_k}{\mathrm{tr}(\Pi_k \rho)} \Pi_k, \tag{3.8}$$

   where Π_k is the measurement operator for the k-th outcome which satisfy (1.2). $k = 0$ denote the event where there are no clicks on both side.

2. Break the loop and jump to step 5 if $\mathrm{tr}\{|R_j\rho_j - \rho_j|\} \leq \varepsilon$.

3. Choose a small step ϵ_j and update ρ_j and $\boldsymbol{y}_j$ according to

$$\rho_{j+1} = \frac{\left[1 + \frac{\epsilon_j}{2}(R_j - 1)\right]\rho_j\left[1 + \frac{\epsilon_j}{2}(R_j - 1)\right]}{\mathrm{tr}\left\{\left[1 + \frac{\epsilon_j}{2}(R_j - 1)\right]\rho_j\left[1 + \frac{\epsilon_j}{2}(R_j - 1)\right]\right\}}, \tag{3.9}$$

$$\boldsymbol{y}_{j+1} = \boldsymbol{y}_j + \epsilon_j \nabla_{\boldsymbol{y}} \log \mathcal{L}, \tag{3.10}$$

where $\nabla_{\boldsymbol{y}} \log \mathcal{L} = (\frac{\partial \log \mathcal{L}}{\partial y_1}, ..., \frac{\partial \log \mathcal{L}}{\partial y_k})^T$ and the derivatives of log-likelihood for double crosshair measurement and simplified measurement are stated in (3.7) and (3.6) respectively.

4. Set $j = j + 1$ and start from step 1 again.

5. $\boldsymbol{\eta}$ is calculated from $\boldsymbol{y}_j$. Return ρ_j as $\hat{\rho}_{\mathrm{MLE}}$ and $\boldsymbol{\eta}$ as $\hat{\boldsymbol{\eta}}_{\mathrm{MLE}}$.

The small step ϵ in step 3 can be chosen to be a fixed value. $\epsilon = 0.6$ works quite well for most cases. Alternatively, one could perform a line search procedure to search for the optimal ϵ such that the increase in $\log \mathcal{L}$ is largest. By adopting the line search method, the number of iterations needed can be decreased. However, each iteration is computationally more expensive and thus it is less efficient overall. Therefore, it is more preferred to choose a fixed value for ϵ.

If the log-likelihood is indeed not concave and has multiple maxima, the algorithm that was just stated can be used to provide evidence for that. For instance, one might start at different starting points and perform the iterative algorithm. Since the direct gradient method will search for the nearest local maxima, it might converge to different local maxima from different starting points if the log-likelihood have multiple maxima.

3.3 Multiple maxima in the likelihood function

For both measurements, we found that in some cases, the direct gradient method converges to different local maxima from different starting points. This shows that the log-likelihood for both measurements might not be concave and might have multiple maxima. When the likelihood function has multiple maxima, it might still be possible to find the global maximum of

the likelihood function using other methods and take the global maximum as the maximum-likelihood estimator $\hat{\rho}_{\mathrm{MLE}}$. However, ambiguity might arise when the likelihood function has multiple maxima with approximately the same height. All the maxima with approximately the same height are equally good as estimator and there is no reason to choose one over another. Besides, the global maximum might change from one maxima to another when there is slight change in the measured data. This means that the estimator is extremely sensitive to changes in the measured data if we naively choose the global maximum as the estimator.

Let us look at an example where there are multiple maxima in the likelihood function for the simplified measurement. The relative frequencies measured are

$$\begin{aligned} f &= (f_{HH}, f_{HV}, f_{H0}, f_{VH}, f_{VV}, f_{V0}, f_{0H}, f_{0V}, f_{00}) \\ &= (0.0782, 0.0553, 0.1461, 0.0346, 0.0245, 0.0655, 0.1679, 0.1151, 0.3128) \end{aligned}$$

In this case, hundreds of different maxima have been observed and we will only list out three of them.
First maximum:[2]

$$\eta_1 = 0.03712, \quad \eta_2 = 0.7991, \quad \eta_1' = 0.9777, \quad \eta_2' = 0.1745,$$

$$\rho = \begin{pmatrix} 0.2152 & * & * & * \\ * & 0.0708 & * & * \\ * & * & 0.5409 & * \\ * & * & * & 0.1731 \end{pmatrix},$$

$$\text{log-likelihood} = -1.938179;$$

[2]Note that we can only deduce the diagonal components of the density operator since this is an incomplete tomography.

second maximum:

$$\eta_1 = 0.5622, \quad \eta_2 = 0.3893, \quad \eta_1' = 0.4344, \quad \eta_2' = 0.3497,$$

$$\rho = \begin{pmatrix} 0.3214 & * & * & * \\ * & 0.3223 & * & * \\ * & * & 0.1780 & * \\ * & * & * & 0.1784 \end{pmatrix},$$

$$\text{log-likelihood} = -1.938189;$$

third maximum:

$$\eta_1 = 0.8638, \quad \eta_2 = 0.2887, \quad \eta_1' = 0.3289, \quad \eta_2' = 0.8314,$$

$$\rho = \begin{pmatrix} 0.2768 & * & * & * \\ * & 0.5734 & * & * \\ * & * & 0.0482 & * \\ * & * & * & 0.1017 \end{pmatrix},$$

$$\text{log-likelihood} = -1.938186.$$

We could also calculate the detection probabilities from these three maxima. From the first maximum:

$$\begin{aligned} p &= (p_{HH}, p_{HV}, p_{H0}, p_{VH}, p_{VV}, p_{V0}, p_{0H}, p_{0V}, p_{00}) \\ &= (0.0781, 0.0553, 0.1462, 0.0350, 0.0241, 0.0654, 0.1675, 0.1155, 0.3128); \end{aligned}$$

from the second maximum:

$$p = (0.0785, 0.0545, 0.1466, 0.0350, 0.0243, 0.0653, 0.1672, 0.1161, 0.3124);$$

from the third maximum:

$$p = (0.0786, 0.0544, 0.1465, 0.0346, 0.0244, 0.0656, 0.1675, 0.1160, 0.3123).$$

The three maxima are equally good as the estimator since they have approximately the same height (log-likelihood). However, the three maxima give us very different information since they are very far away from each other in the space of states and efficiencies. The fidelity of the three states are calculated to be as such:

$$F(\rho_1, \rho_2) = 0.90, \quad F(\rho_1, \rho_3) = 0.74, \quad F(\rho_2, \rho_3) = 0.96.$$

The distance between the estimated efficiencies are calculated to be as such:

$$|\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2| = 0.73, \quad |\boldsymbol{\eta}_1 - \boldsymbol{\eta}_3| = 1.16, \quad |\boldsymbol{\eta}_2 - \boldsymbol{\eta}_3| = 0.59.$$

On the other hand, the detection probabilities calculated from the three maxima are very close to each other and the relative frequencies. The problem arises when we transform from probabilities to state and efficiencies. Small changes in the probabilities might lead to large changes in state and efficiencies.

In the rest of this section, we provide a possible explanation of having multiple maxima which are very far apart and yet have detection probabilities which are so close. Equation (2.7b) can be expressed as

$$\begin{pmatrix} | & | \\ \vec{a} & \vec{b} \\ | & | \end{pmatrix} \begin{pmatrix} \frac{1}{\eta_1} \\ \frac{1}{\eta_2} \end{pmatrix} = \begin{pmatrix} | \\ \vec{c} \\ | \end{pmatrix}. \tag{3.11}$$

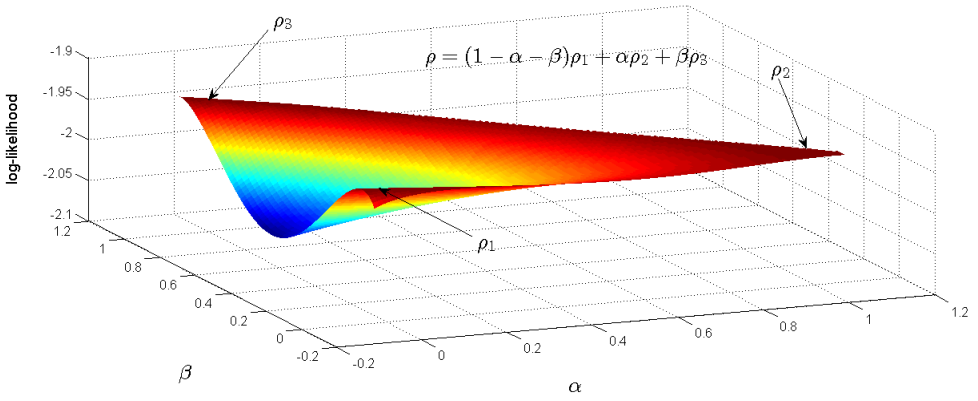


Figure 3.2: ρ_1, ρ_2, and ρ_3 are the three maxima. The graph of log-likelihood of convex combinations of the three maxima is plotted. It is clear from the graph that the log-likelihood function is not a concave function.

Expression of η_1 can be derived as

$$\eta_1 = \frac{(\vec{b} \times \vec{a}) \cdot (\vec{b} \times \vec{a})}{(\vec{b} \times \vec{a}) \cdot (\vec{b} \times \vec{c})}, \tag{3.12}$$

and the fractional uncertainty of η_1 can be derived as

$$\frac{\delta\eta_1}{\eta_1} = \frac{2(\vec{b} \times \vec{a}) \cdot \delta(\vec{b} \times \vec{a})}{(\vec{b} \times \vec{a}) \cdot (\vec{b} \times \vec{a})} - \frac{\delta((\vec{b} \times \vec{a}) \cdot (\vec{b} \times \vec{c}))}{(\vec{b} \times \vec{a}) \cdot (\vec{b} \times \vec{c})}. \tag{3.13}$$

From (3.13), we conclude that even when the variation in probabilities is small, the fractional uncertainty of η_1 can be quite large if $(\vec{b} \times \vec{a})$ is quite small or $\vec{b}$ is almost parallel to $\vec{a}$. For all the cases with multiple maxima in the likelihood, we notice that $\vec{b}$ is indeed almost parallel to $\vec{a}$. The angle between them, $\theta_{\vec{a},\vec{b}}$ is smaller than 0.01 rad when multiple maxima were observed. Typical values of $\theta_{\vec{a},\vec{b}}$ found in the simulations are listed below:

$$\text{first quartile} = 0.034\,\text{rad}, \quad \text{median} = 0.073\,\text{rad}, \quad \text{third quartile} = 0.13\,\text{rad}.$$

The tenth percentile of $\theta_{\vec{a},\vec{b}}$ is 0.013 rad which is about the value when we start to observe multiple maxima. In our simulations, multiple maxima occur in about 5% of the cases.

3.4 Ratios of detector efficiencies are known

The ambiguity we encounter in the previous section suggests that self-calibration scheme for the double crosshair measurement and the simplified measurement might not be feasible in practice although it can be done theoretically. We can still extract some information about the detectors if we take the prior knowledge of the ratios of detector efficiencies into account. In practice, the ratios of detector efficiencies can be measured much more easily than the absolute detector efficiencies. For instance, one could shine a laser on two detectors for the same amount of time, the ratio of efficiencies of the two detectors can be measured by measuring the ratio of counts detected on the two detectors.

When the prior knowledge of the ratios of detector efficiencies is taken into account, the number of unknowns to be determined are reduced significantly. The unknowns to be determined are the state parameters and the largest detector efficiency $\eta_{\max}$. All the detector efficiencies can be deduced from the ratios once we know $\eta_{\max}$. For this new scheme, there are only minor modifications to the algorithms for direct gradient method in section 3.2. Only the largest detector efficiency $\eta_{\max}$ is parametrized by parameter y,

$$\eta_{\max} = \sin^2 y. \tag{3.14}$$

All the other detector efficiencies can be deduced from the ratios of efficiencies $\boldsymbol{\kappa}$ once we know $\eta_{\max}$,

$$\boldsymbol{\eta} = \eta_{\max}\,\boldsymbol{\kappa} = \eta_{\max}\,(\kappa_1, \kappa_2, \kappa_1', \kappa_2')^T, \tag{3.15a}$$

or

$$\boldsymbol{\eta} = \eta_{\max}\,\boldsymbol{\kappa} = \eta_{\max}\,(\kappa_1, ..., \kappa_4, \kappa_1', ..., \kappa_4')^T. \tag{3.15b}$$

The derivatives of log-likelihood with respect to the parameter y in (3.6) and (3.7) need to be modified as well. The new derivatives of log-likelihood are stated in (3.16) and (3.17) for the simplified measurement and the double crosshair measurement respectively.

$$\begin{aligned}\frac{\partial \log \mathcal{L}}{\partial y} = 2\cot y\Big(&2n_{HH} + 2n_{HV} + 2n_{VH} + 2n_{VV} \\ &+ \frac{n_{H0}}{p_{H0}}(p_{H0} - p_{HH} - p_{HV}) + \frac{n_{V0}}{p_{V0}}(p_{V0} - p_{VH} - p_{VV}) \\ &+ \frac{n_{0H}}{p_{0H}}(p_{0H} - p_{HH} - p_{VH}) + \frac{n_{0V}}{p_{0V}}(p_{0V} - p_{HV} - p_{VV}) \\ &- \frac{n_{00}}{p_{00}}(p_{H0} + p_{V0} + p_{0H} + p_{0V})\Big),\end{aligned} \tag{3.16}$$

$$\begin{aligned}\frac{\partial \log \mathcal{L}}{\partial y} = 2\cot y\Big(&2n_{HH} + 2n_{HV} + 2n_{HD} + 2n_{HA} \\ &+ 2n_{VH} + 2n_{VV} + 2n_{VD} + 2n_{VA} \\ &+ 2n_{DH} + 2n_{DV} + 2n_{DD} + 2n_{DA} \\ &+ 2n_{AH} + 2n_{AV} + 2n_{AD} + 2n_{AA} \\ &+ \frac{n_{H0}}{p_{H0}}(p_{H0} - p_{HH} - p_{HV} - p_{HD} - p_{HA}) \\ &+ \frac{n_{V0}}{p_{V0}}(p_{V0} - p_{VH} - p_{VV} - p_{VD} - p_{VA}) \\ &+ \frac{n_{D0}}{p_{D0}}(p_{D0} - p_{DH} - p_{DV} - p_{DD} - p_{DA}) \\ &+ \frac{n_{A0}}{p_{A0}}(p_{A0} - p_{AH} - p_{AV} - p_{AD} - p_{AA}) \\ &+ \frac{n_{0H}}{p_{0H}}(p_{0H} - p_{HH} - p_{VH} - p_{DH} - p_{AH}) \\ &+ \frac{n_{0V}}{p_{0V}}(p_{0V} - p_{HV} - p_{VV} - p_{DV} - p_{AV}) \\ &+ \frac{n_{0D}}{p_{0D}}(p_{0D} - p_{HD} - p_{VD} - p_{DD} - p_{AD}) \\ &+ \frac{n_{0A}}{p_{0A}}(p_{0A} - p_{HA} - p_{VA} - p_{DA} - p_{AA}) \\ &- \frac{n_{00}}{p_{00}}(p_{H0} + p_{V0} + p_{D0} + p_{A0} + p_{0H} + p_{0V} + p_{0D} + p_{0A})\Big).\end{aligned} \tag{3.17}$$

Similar to previous sections, we do not have a mathematical proof that the log-likelihood function is a concave function because the likelihood function for this case is still too complicated. However, in more than ten thousands different simulations, we found that the direct gradient method always converge to the same point no matter where the starting point is. This numerical evidence suggests that either the likelihood function for this new scheme has only one maximum or the case of having multiple maxima is extremely rare.

After taking the prior information of the ratios of the detector efficiencies into account, the problem of having multiple maxima in the likelihood func-

tion disappears and we can always find the unique estimator $\hat{\rho}_{\mathrm{MLE}}$ and $\hat{\boldsymbol{\eta}}_{\mathrm{MLE}}$. From the same data in the example with multiple maxima in section 3.3, an unique estimator $\hat{\rho}_{\mathrm{MLE}}$ can be found after considering the prior information of the ratios of the detector efficiencies. Moreover, $\hat{\rho}_{\mathrm{MLE}}$ is very close to the true state ρ_{True}. In particular, the fidelity of the two states $F(\hat{\rho}_{\mathrm{MLE}}, \rho_{\mathrm{True}})$ is found to be 0.999995 which is very close to one.

3.5 Unknown total number of photons

So far, we have been assuming that we know the total number of photon pairs emitted by the source and thus the number of undetected photon pairs, n_{00} is known. This assumption is however not realistic. A more realistic assumption is that the total number of photons emitted by the source is unknown but it follows Poissonian distribution with mean ν and the mean number of photons ν is known.

In this case, the measured data D for the double crosshair measurement consists of twenty-four numbers instead of twenty-five numbers since we do not know the number of undetected photon pairs, n_{00} now. For the simplified measurement, D consists of eight number instead of nine. We now define a new likelihood function where we sum over each cases with different number

of photon pairs.

$$\begin{aligned}
\mathcal{L}(D|\rho,\boldsymbol{\eta},\nu) &= \sum_{n=N}^{\infty} \frac{\nu^n}{n!} e^{-\nu} \mathcal{L}(n_0, D|\rho,\boldsymbol{\eta}), \qquad \text{where } N = \sum_{k=1}^{K-1} n_k \\
&= \sum_{n=N}^{\infty} \frac{\nu^n}{n!} e^{-\nu} \frac{n!}{N!(n-N)!} {p_0}^{n-N} \prod_{k=1}^{K-1} {p_k}^{n_k} \\
&= \frac{e^{-\nu}}{N!} \sum_{n=N}^{\infty} \frac{\nu^n}{(n-N)!} {p_0}^{n-N} \prod_{k=1}^{K-1} {p_k}^{n_k} \\
&= \frac{e^{-\nu}\nu^N}{N!} \sum_{m=0}^{\infty} \frac{(\nu p_0)^m}{m!} \prod_{k=1}^{K-1} p_k^{n_k} \\
&= \frac{e^{-\nu}\nu^N}{N!} e^{\nu p_0} \prod_{k=1}^{K-1} {p_k}^{n_k}. \qquad (3.18)
\end{aligned}$$

Discarding an unimportant constant factor, we get

$$\mathcal{L}(D|\rho,\boldsymbol{\eta},\nu) = e^{\nu p_0} \prod_{k=1}^{K-1} {p_k}^{n_k}. \qquad (3.19)$$

The multinomial factor $\frac{n!}{N!(n-N)!}$ has to be included in second line of (3.18) in order to take into account all possible sequences of having N detected photon pairs out of a total of n photon pairs emitted by the source.

Similarly, the maximum-likelihood estimator is obtained by maximizing the new log-likelihood function

$$\log \mathcal{L}(D|\rho,\boldsymbol{\eta},\nu) = \nu p_0 + \sum_{k=1}^{K-1} n_k \log p_k. \qquad (3.20)$$

This new log-likelihood is slightly different from (1.9). Thus, there are also minor modifications to the algorithms for the direct gradient method. The modifications needed are

1. Define a new R matrix

$$R = \sum_{k=1}^{K-1} \frac{n_k}{\mathrm{tr}(\Pi_k \rho)} \Pi_k + \nu \Pi_0. \tag{3.21}$$

2. Modify the loop terminating criterion

$$\mathrm{tr}\left\{ \left| \frac{R_j \rho_j}{\mathrm{tr}(R_j \rho_j)} - \rho_j \right| \right\} \leq \varepsilon. \tag{3.22}$$

3. Modify the change in ρ for each step

$$\rho_{j+1} = \frac{\left[1 + \frac{\epsilon}{2}\left(R_j - \mathrm{tr}(R_j \rho_j)\right)\right] \rho_j \left[1 + \frac{\epsilon}{2}\left(R_j - \mathrm{tr}(R_j \rho_j)\right)\right]}{\mathrm{tr}\left\{ \left[1 + \frac{\epsilon}{2}\left(R_j - \mathrm{tr}(R_j \rho_j)\right)\right] \rho_j \left[1 + \frac{\epsilon}{2}\left(R_j - \mathrm{tr}(R_j \rho_j)\right)\right] \right\}}. \tag{3.23}$$

4. Modify the derivatives of log-likelihood in (3.16) and (3.17) by replacing the factor $\frac{n_{00}}{p_{00}}$ with ν.

We have thus shown a way to get rid of the unrealistic assumption that the total number of photon pairs emitted is known. However, we will continue to assume that the total number of photon pairs emitted is known when constructing error regions in the next chapter. Only slight modifications are needed if one choose to work with the more realistic case where the total number of photons pairs emitted follows Poissonian distribution.

Chapter 4

Constructing error regions

4.1 Monte Carlo integration

To report the error regions, it remains to calculate the multi-dimensional integrals in (1.17) and (1.18). Both integrals are hard to be computed directly as the boundary in probability space can be very complicated due to the nontrivial term $w_{\mathrm{qu}}(p)$. Among all the numerical integration methods, Monte Carlo integration suggests itself as its performance is determined by the number of sample points taken and not the dimension of the problem. Random samples are generated according to the distribution $w_0(p)$ and $w_D(p)$ respectively, and the size $s_{\mathcal{R}}$ and credibility $c_{\mathcal{R}}$ are computed by calculating the ratios of number of points lies within $\mathcal{R}$ to the number of points lies within the reconstruction space $\mathcal{R}_0$.

The problem of computing the multi-dimensional integrals has thus been reduced to generating random sample points according to a density distribution $w(p)$. This is however not trivial. In [14], Shang *et al.* discussed the methods of importance sampling, rejection sampling, and Metropolis-Hastings Monte Carlo sampling. Although the three methods are easy to implement, they suffer from the problems of being costly in CPU time or require a large number of points to get a good sample. In [12], Seah *et al.*

discussed the Hamiltonian Monte Carlo (HMC) method [9] which is relatively harder to implement but takes much shorter time. In this thesis, we adopt the Hamiltonian Monte Carlo method.

4.2 Hamiltonian Monte Carlo

The main idea of Hamiltonian Monte Carlo (HMC) is to apply Hamiltonian dynamics by identifying the variables of interest as position variables $\mathrm{q} = (\mathrm{q}_1, \mathrm{q}_2, ..., \mathrm{q}_K)$ and introducing fictitious momentum variables $\mathrm{p} = (\mathrm{p}_1, \mathrm{p}_2, .., \mathrm{p}_K)$. The Hamiltonian is defined in terms of the target distribution $w(\mathrm{q})$,

$$H(\mathrm{q}, \mathrm{p}) = \frac{1}{2} \sum_j \mathrm{p}_j^2 + U(\mathrm{q}). \tag{4.1}$$

The potential energy $U(\mathrm{q})$ is given by

$$U(\mathrm{q}) = -\log w(\mathrm{q}). \tag{4.2}$$

By performing the HMC algorithm, we will obtain a set of sample points q which follows the target distribution $w(\mathrm{q})$. The HMC algorithm is stated as follows:

HMC algorithm

1. Choose an arbitrary starting point $\mathrm{q}^{(1)}$ and set $j = 1$.

2. Randomly generate $\mathrm{p}^{(j)}$ from multivariate Gaussian distribution with mean zero and variance one.

3. Evolve $(\mathrm{q}^{(j)}, \mathrm{p}^{(j)})$ according to Hamiltonian equations of motion

$$\frac{\mathrm{d}\mathrm{q}_i}{\mathrm{d}t} = \frac{\partial H}{\partial \mathrm{p}_i}, \quad \frac{\mathrm{d}\mathrm{p}_i}{\mathrm{d}t} = -\frac{\partial H}{\partial \mathrm{q}_i} \tag{4.3}$$

for a duration of T to get $(\mathrm{q}^{(j)}(T), \mathrm{p}^{(j)}(T))$ and set $(\mathrm{q}^*, \mathrm{p}^*) = (\mathrm{q}^{(j)}(T), -\mathrm{p}^{(j)}(T))$.

4. Calculate the acceptance ratio[1]

$$a = \min[1, e^{H(\mathrm{q}^{(j)}, \mathrm{p}^{(j)}) - H(\mathrm{q}^*, \mathrm{p}^*)}]. \tag{4.4}$$

5. Draw a random number b from uniform distribution $U(0, 1)$. If $b < a$, set $\mathrm{q}^{(j+1)} = \mathrm{q}^*$; otherwise, set $\mathrm{q}^{(j+1)} = \mathrm{q}^{(j)}$.

6. Set $j = j + 1$. If j equals to the desired number of samples, escape the loop; otherwise, return to step 2.

For computer implementation of step 3 in HMC algorithm, the differential equations in (4.3) must be discretized. This is done by the leapfrog method. The total time duration T is divided into L sub-intervals $\tau = T/L$. For each of the sub-intervals τ, $(\mathrm{q}(t_0 + \tau), \mathrm{p}(t_0 + \tau))$ are obtained from $(\mathrm{q}(t_0), \mathrm{p}(t_0))$ by the leapfrog method which is stated as follows:

Leapfrog method

1. Let the potential energy govern (4.3) for a duration of $\frac{\tau}{2}$,

$$\mathrm{p}\left(t_0 + \frac{\tau}{2}\right) = \mathrm{p}(t_0) + \frac{\tau}{2} u(\mathrm{q}(t_0)), \tag{4.5}$$

where $u(\mathrm{q}) = -\nabla H = w(\theta)^{-1} \nabla w(\theta)$ is the force.

2. Let the kinetic energy govern (4.3) for a duration of τ,

$$\mathrm{q}(t_0 + \tau) = \mathrm{q}(t_0) + \tau \mathrm{p}\left(t_0 + \frac{\tau}{2}\right). \tag{4.6}$$

[1]The acceptance ratio is always equals to one if the evolution in step 3 is exact due to conservation of energy.

3. Let the potential energy govern (4.3) for a duration of $\frac{\tau}{2}$,

$$\mathrm{p}(t_0+\tau) = \mathrm{p}\left(t_0+\frac{\tau}{2}\right) + \frac{\tau}{2}u(\mathrm{q}(t_0+\tau)). \tag{4.7}$$

4.3 State parametrization

All the sampling methods in [14] are done by parametrizing the state by probabilities as in (1.13). This will requires a physicality check for the constraints in (1.12) which is computationally expensive. As a result, all those sampling methods are costly in CPU time. In [12], the HMC sampling are done by parametrizing all the physical states in the state space or reconstruction space directly. By doing this, the physicality check which is computationally expensive can be avoided since all the constraints in (1.12) are automatically satisfied.

For informationally complete POM, a $d \times d$ density operator can be parametrized by $d^2 - 1$ real parameters. This can be done by any upper-triangular matrix A with real diagonal entries such that

$$\rho = A^\dagger A, \quad \text{and} \quad \mathrm{tr}\{A^\dagger A\} = 1. \tag{4.8}$$

For the case when $d = 4$, A is given by

$$A = \begin{pmatrix} C_1 & C_2E_{10} & C_3E_{11} & C_4E_{12} \\ 0 & C_5 & C_6E_{13} & C_7E_{14} \\ 0 & 0 & C_8 & C_9E_{15} \\ 0 & 0 & 0 & S_9 \end{pmatrix}, \tag{4.9}$$

where

$$C_1 = \cos\theta_1, \quad S_1 = \sin\theta_1; \quad C_k = S_{k-1}\cos\theta_k, \quad S_k = S_{k-1}\sin\theta_k, \tag{4.10}$$

and

$$E_k = e^{-\mathrm{i}\theta_k}. \tag{4.11}$$

4.3.1 Simplified measurement

In the simplified measurement, the three state parameters : $\langle 1 \otimes \sigma_z \rangle$, $\langle \sigma_z \otimes 1 \rangle$, and $\langle \sigma_z \otimes \sigma_z \rangle$ fully characterize the diagonal entries of the state ρ. Therefore, the reconstruction space can be chosen to be the set of all the states with vanishing non-diagonal components. The parametrization of A is given by

$$A = \begin{pmatrix} \cos\theta_1 & 0 & 0 & 0 \\ 0 & \cos\theta_2 \sin\theta_1 & 0 & 0 \\ 0 & 0 & \cos\theta_3 \sin\theta_2 \sin\theta_1 & 0 \\ 0 & 0 & 0 & \sin\theta_3 \sin\theta_2 \sin\theta_1 \end{pmatrix}, \tag{4.12}$$

or

$$A = \begin{pmatrix} C_1 & 0 & 0 & 0 \\ 0 & C_2 & 0 & 0 \\ 0 & 0 & C_3 & 0 \\ 0 & 0 & 0 & S_3 \end{pmatrix}, \tag{4.13}$$

where C_k and S_k are given in (4.10). We are now left with the last parameter corresponding to $\eta_{\max}$. This can be done in the similar way as in section 3.4,

$$\eta_{\max} = \sin^2(\theta_4). \tag{4.14}$$

After reparametrization, the prior or posterior density in terms of θ are given by

$$w(\theta) = w(p) \left| \frac{\partial p}{\partial \theta} \right|, \tag{4.15}$$

where $\left|\frac{\partial p}{\partial \theta}\right|$ is the determinant of the Jacobian. With the choice of primitive

prior (1.15), the prior density is given by

$$w_0(\theta) = \left|\frac{\partial p}{\partial \theta}\right|, \tag{4.16}$$

whereas the posterior density is given by

$$w_D(\theta) = \mathcal{L}(D|\rho, \boldsymbol{\eta}) \left|\frac{\partial p}{\partial \theta}\right|. \tag{4.17}$$

After the force components,

$$u_s(\theta) = \frac{\partial}{\partial \theta_s} \log w(\theta) \tag{4.18}$$

are calculated, the HMC algorithm can now be performed to sample the prior or posterior.

Note that $\frac{\partial p}{\partial \theta}$ must be a square matrix, otherwise its determinant cannot even be defined. Not all of the probabilities are independent to each other and the dimension of the probability space must be equal to the number of parameters θ. Thus, a set of independent probabilities must be chosen to compute $\frac{\partial p}{\partial \theta}$. For the case of perfect detectors, it is obvious that p_{HH}, p_{HV}, and p_{VH} form such a set as they must be summed to one with p_{VV}. For the case of imperfect detectors with known ratios of efficiencies, the expressions of probabilities are given by (2.2) (with efficiencies given by (3.15a)). With the extra factors of κ and $\eta_{\max}$, it might not be obvious which probabilities are independent to each other. The set of independent probabilities is found by finding a minimal set of probabilities that is sufficient to solve for all the state parameters and $\eta_{\max}$. One such choice is the set $\{p_{HH}, p_{HV}, p_{VH}, p_{VV}\}$. For instance, one could use the first two expression of $\langle 1 \otimes \sigma_z \rangle$ in (2.6a) to first solve for $\eta_{\max}$. All the other state parameters in (2.6) can then be solved easily.

In figure 4.1, we show the size and credibility as a function of λ for the data $D = \{3, 4, 3, 6, 2, 1, 10, 4, 2\}$. The sample size used for Monte Carlo

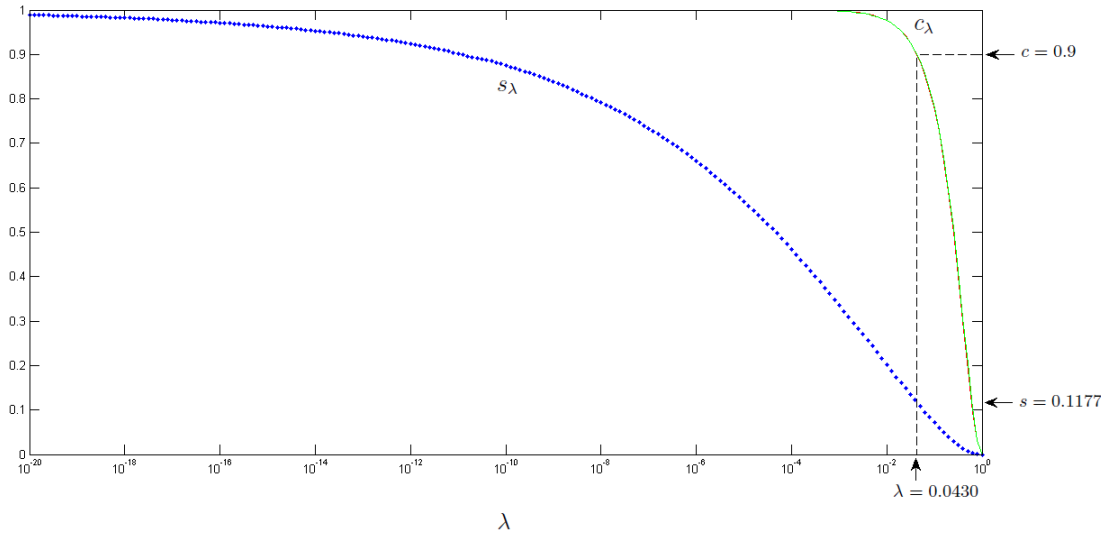


Figure 4.1: Size and credibility as a function of λ. Blue dots is s_λ from prior sampling, green line is c_λ from posterior sampling, red line is c_λ from computing (1.19)

integration is 50000 points. The credibility was calculated by two different methods, the first being sampling from posterior and compute (1.18), the second is by computing (1.19) after we have obtained size as a function of λ. The credibility obtained from both methods agree with each other, as they should. This also indicates that the sample of 50000 points is sufficiently large to get an accurate result in this case.

The value of λ that corresponds to credibility $c = 0.9$ is found to be 0.0430. Thus, the SCR with credibility $c = 0.9$ is the BLR with $\lambda = 0.0430$ and its size s is 0.1177. The true state is contained in the SCR with credibility $c = 0.9$.

4.3.2 Double crosshair measurement

In this scenario, there is no easy way to parametrize the reconstruction space. The eight state parameters: $\langle 1 \otimes \sigma_x \rangle$, $\langle 1 \otimes \sigma_z \rangle$, $\langle \sigma_x \otimes 1 \rangle$, $\langle \sigma_x \otimes \sigma_x \rangle$, $\langle \sigma_x \otimes \sigma_z \rangle$, $\langle \sigma_z \otimes 1 \rangle$, $\langle \sigma_z \otimes \sigma_x \rangle$, and $\langle \sigma_z \otimes \sigma_z \rangle$ plus an unknown parameter

$Q = \langle \sigma_y \otimes \sigma_y \rangle$ fully characterize the real part of ρ.

$$\rho = \frac{1}{4} \begin{pmatrix} 1 + S_{1z} + S_{z1} + S_{zz} & S_{1x} + S_{zx} & S_{x1} + S_{xz} & S_{xx} - Q \\ S_{1x} + S_{zx} & 1 - S_{1z} + S_{z1} - S_{zz} & S_{xx} + Q & S_{x1} - S_{xz} \\ S_{x1} + S_{xz} & S_{xx} + Q & 1 + S_{1z} - S_{z1} - S_{zz} & S_{1x} - S_{zx} \\ S_{xx} - Q & S_{x1} - S_{xz} & S_{1x} - S_{zx} & 1 - S_{1z} - S_{z1} + S_{zz} \end{pmatrix} \tag{4.19}$$

where $S_{ij} = \langle \sigma_i \otimes \sigma_j \rangle$ and $\sigma_1 = 1$. For the state to be positive semi-definite, there is a range of permissible values of Q,

$$-1 \leq Q_{\min}(\rho) \leq Q \leq Q_{\max}(\rho) \leq 1. \tag{4.20}$$

We could introduce extra parameter associated with Q, and thus have the real part of ρ as the reconstruction space. The real part of ρ can be parametrized using (4.9) with $E_k = 0$. The HMC sampling can be performed on this higher dimensional space, and the extra parameter is then marginalized to arrive at the proper sample. However, to marginalize the extra parameters due to Q, we need to have a closed formula of $Q_{\max}(\rho) - Q_{\min}(\rho)$ which we do not have. Instead, all the sample points that we obtained from HMC sampling are supplied with a weight of $\frac{1}{Q_{\max}(\rho) - Q_{\min}(\rho)}$ to account for the extra parameter introduced during the sampling.

As stated in previous section, a set of independent probabilities has to be chosen so that $\frac{\partial p}{\partial \theta}$ is a square matrix. The number of parameters θ is ten. However, one of them is the additional parameter corresponds to $Q = \langle \sigma_y \otimes \sigma_y \rangle$. We can take Q as one of the probabilities and find the remaining nine independent probabilities.

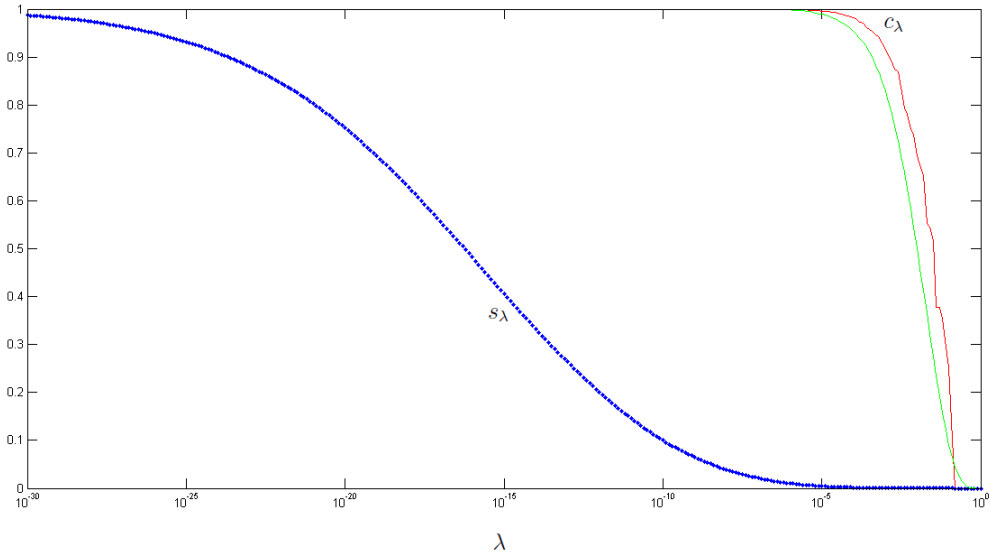
For the case of perfect detectors, it is not hard to see from the expressions of probabilities ((2.1) with all η equal to one) that $\{p_{HH}, p_{HV}, p_{HD}, p_{VH}, p_{VD}, p_{DH}, p_{DV}, p_{DD}\}$ is a set of independent probabilities. For the case of imperfect detectors with known ratios of efficiencies, the expressions of probabilities are given by (2.1) (with efficiencies given by (3.15b)). With the extra factor of κ and $\eta_{\max}$, it is again not obvious which probabilities

are independent to each other. As in previous part, we obtain the set of independent probabilities by finding a minimal set of probabilities which is sufficient to solve for all the state parameters and $\eta_{\max}$. One such choice is the set $\{p_{HH}, p_{HV}, p_{HD}, p_{VH}, p_{VV}, p_{VD}, p_{DH}, p_{DV}, p_{DD}\}$. For instance, one could use the first and third expression of $\langle 1 \otimes \sigma_z \rangle$ in (2.4a) to first solve for $\eta_{\max}$. All the other state parameters in (2.4) can then be solved easily.

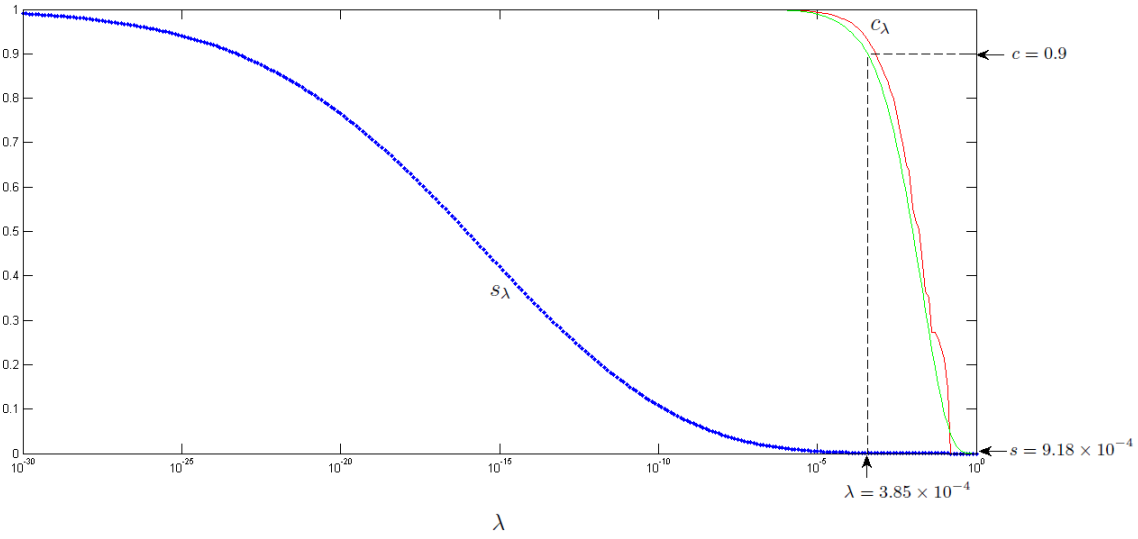
With the set of nine independent probabilities and Q, $\frac{\partial p}{\partial \theta}$ can be calculated and so are (4.16), (4.17), and (4.18). We now have everything needed to performed the HMC algorithm. However, each sample point that we obtained from the HMC algorithm must be attached with a weight of $\frac{1}{Q_{\max}(\rho) - Q_{\min}(\rho)}$ to account for the fact that Q is the external parameter we introduced.

In figure 4.2, we show the size and credibility as a function of λ for the data $D = \{1, 2, 0, 0, 6, 3, 3, 3, 2, 3, 2, 3, 4, 1, 6, 0, 0, 1, 0, 0, 1, 10, 10, 1, 13\}$. The sample size used for Monte Carlo integration is 250000 points and 750000 points. The credibility was also calculated by the two different methods. We notice difference in the credibility from the two methods. However, the credibility from the two methods become closer as the sample size increase. The true value of s_λ near $\lambda = 1$ is very small and thus s_λ obtained from Monte Carlo simulation can have high discrepancies from the true value if the sample size is not large enough. This causes the calculated s_λ to be not smooth near $\lambda = 1$. This explains why the red curve (which is obtained from integrating s_λ) is not smooth and has a sharp decrease to zero near $\lambda = 1$. In this case, the sample of 750000 points (which is very large in other situations) is not good enough to have a smooth c_λ from calculating (1.19) whereas the c_λ from posterior sampling is smooth even with a much smaller sample size. Thus, the calculation of c_λ from posterior sampling is preferred in this case.

The value of λ that corresponds to credibility $c = 0.9$ is found to be 3.85×10^{-4}. Thus, the SCR with credibility $c = 0.9$ is the BLR with $\lambda = 3.85 \times 10^{-4}$ and its size s is 9.18×10^{-4}. The true state is contained in the SCR with credibility $c = 0.9$.



(a)



(b)

Figure 4.2: Size and credibility as a function of λ. Blue dots is s_λ from prior sampling, green line is c_λ from posterior sampling, red line is c_λ from computing (1.19). Plot (a) is obtained from sample of 250000 points, plot (b) is obtained from sample of 750000 points.

Chapter 5

Conclusion

In chapter one, we give a brief introduction to quantum state estimation. We first introduce two ways to find the point estimators i.e. linear inversion and maximum-likelihood estimation. We then introduce self-calibrating quantum state estimation for reconstructing the state and certain properties of measurement devices from the same data. Smallest credible regions (SCR) is introduced as the optimal error regions. The fact that SCR is a bounded-likelihood region (BLR) provides a simple way to report it. We end this chapter by discussing the notions of size and credibility of a region.

In chapter two, we proceed to study the double crosshair measurement and a simplified version of it. We prove that the state parameters and the detector efficiencies can be uniquely determined if the exact detection probabilities are known. This implies that self-calibration scheme for both measurement is feasible theoretically.

In chapter three, we first discuss the use of linear inversion and maximum-likelihood estimation to find the point estimator. We conclude that maximum-likelihood estimation is more preferred as linear inversion might give us unphysical states quite often. However, we face the problem of having multiple maxima in the likelihood function which have approximately the same height. This shows that our self-calibration scheme is not feasible in practice. Par-

tial self-calibration can still be done if we take the prior knowledge of the ratios of detector efficiencies into account. After taken the prior knowledge of the ratios of detector efficiencies into account, the problem of having multiple maxima in the likelihood function disappears. We end chapter three by showing a way to get rid of the unrealistic assumption that the total number of photon pairs emitted is known.

Finally, we move on to construct an error region which will be attached to the maximum-likelihood estimator to express the uncertainties associated with it. To report error regions, the multidimensional integrals for size and credibility have to be computed. Samples with prior and posterior density have to generated to calculate the multidimensional integrals with Monte Carlo integration. We introduce Hamiltonian Monte Carlo (HMC) sampling to generate the sample. Although HMC is the most efficient way to do the sampling, it requires us to parametrize the reconstruction space which is not easy. With the sample in hand, the size and credibility can now be calculated as a function of λ and the SCR can be reported concisely.

Future works

In the typical scenario of quantum key distribution, the two sets of detectors on both sides are separated by a great distance. It might not be practical to measure the ratios of efficiencies of all the detectors. A more practical way to carry out the self-calibration scheme is to measure the ratios of efficiencies of all detectors for each side. Hence, one possible future research direction is to study this more practical scheme.

Appendix A

Least square solution

For an overdetermined system

$$Ax = b, \tag{A.1}$$

define the error by

$$e = |Ax - b|. \tag{A.2}$$

The square of the error is given by

$$e^2 = x^T A^T A x - x^T A^T b - b^T A x + b^T b. \tag{A.3}$$

We want to obtain a vector x such that it minimizes e^2. The variation of e^2 under infinitesimal change of x is given by

$$\begin{aligned} \delta e^2 &= \delta x^T (A^T A x - A^T b) + (x^T A^T A - b^T A)\delta x \\ &= 2\delta x^T (A^T A x - A^T b). \end{aligned} \tag{A.4}$$

By demanding $\delta e^2 = 0$, we obtain

$$A^T A x = A^T b. \tag{A.5}$$

As long as A has full column rank, A^TA will be invertible [1]. Thus, the least square solution of the overdetermined system is given by

$$x = (A^TA)^{-1}A^Tb. \tag{A.6}$$

Bibliography

[1] H. Anton and C. Rorres. *Elementary Linear Algebra*. John Wiley and Sons, Inc, 2010.

[2] K. M. R. Audenaert and S. Scheel. Quantum tomographic reconstruction with error bars: a kalman filter approach. *New J. Phys.*, 11(2):023028, 2009.

[3] C. H. Bennett and G. Brassard. Quantum cryptography: public key distribution and coin tossing. In *IEEE Conf. Computers, Systems, and Signal Processing*, volume 175, New York, 1984.

[4] R. Blume-Kohout. Optimal, reliable estimation of quantum states. *New J. Phys.*, 12(4):043034, 2010.

[5] A. M. Braczyk, D. H. Mahler, L. A. Rozema, A. Darabi, A. M. Steinberg, and D. F. V. James. Self-calibrating quantum state tomography. *New J. Phys.*, 14(8):085003, 2012.

[6] Z. Hradil. Quantum-state estimation. *Phys. Rev. A*, 55:R1561, 1997.

[7] D. Mogilevtsev. Calibration of single-photon detectors using quantum statistics. *Phys. Rev. A*, 82:021807, 2010.

[8] D. Mogilevtsev, J. Řeháček, and Z. Hradil. Relative tomography of an unknown quantum state. *Phys. Rev. A*, 79:020101, 2009.

[9] R. M. Neal. MCMC using hamiltonian dynamics. Published as Chapter 5 of the Handbook of Markov Chain Monte Carlo, 2011.

[10] M. Paris and J. Řeháček. *Quantum State Estimation*, volume 649 of *Lecture Notes in Physics*. Springer, Berlin, Heidelberg, 2004.

[11] N. Quesada, A. M. Brańczyk, and D. F. V. James. Self-calibrating tomography for multidimensional systems. *Phys. Rev. A*, 87:062118, 2013.

[12] Y.-L. Seah, J. Shang, H. K. Ng, D. J. Nott, and B.-G. Englert. Monte carlo sampling from the quantum state space. II, 2014.

[13] J. Shang, H. K. Ng, A. Sehrawat, X. Li, and B.-G. Englert. Optimal error regions for quantum state estimation. *New J. Phys.*, 15(12):123026, 2013.

[14] J. Shang, Y.-L. Seah, H. K. Ng, D. J. Nott, and B.-G. Englert. Monte carlo sampling from the quantum state space. I, 2014.

[15] T. Sugiyama, P. S. Turner, and M. Murao. Precision-guaranteed quantum tomography. *Phys. Rev. Lett.*, 111:160406, 2013.

[16] Y. S. Teo. *Numerical Estimation Schemes for Quantum Tomography*. PhD thesis, 2013.

[17] Y. S. Teo, H. Zhu, B.-G. Englert, J. Řeháček, and Z. Hradil. Quantum-state reconstruction by maximizing likelihood and entropy. *Phys. Rev. Lett.*, 107:020404, 2011.