

Titre: Optimal Pumps for Spontaneous Parametric Down-Conversion in
Title: Waveguides

Auteur: Jiaqi Zhao
Author:

Date: 2024

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Zhao, J. (2024). Optimal Pumps for Spontaneous Parametric Down-Conversion in
Citation: Waveguides [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie.
<https://publications.polymtl.ca/59201/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/59201/>
PolyPublie URL:

**Directeurs de
recherche:** Nicolás Quesada, & Sean Molesky
Advisors:

Programme: Génie physique
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Optimal Pumps for Spontaneous Parametric Down-Conversion in Waveguides

JIAQI ZHAO

Département de génie physique

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

Génie physique

Août 2024

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

Optimal Pumps for Spontaneous Parametric Down-Conversion in Waveguides

présenté par **Jiaqi ZHAO**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

a été dûment accepté par le jury d'examen constitué de :

Yves-Alain PETER, président

Nicolás QUESADA, membre et directeur de recherche

Sean MOLESKY, membre et codirecteur de recherche

Stéphane VIRALLY, membre

DEDICATION

“Let there be light (of many frequencies)!”

ACKNOWLEDGEMENTS

I would like to thank Pr. Nicolás Quesada and Pr. Sean Molesky for their guidance, as well as MoleRing and Polyquantique team for fruitful discussions. And I would like to thank you, dear reader, for taking interest in my work.

RÉSUMÉ

La conversion paramétrique descendante spontanée (CPDS) est un phénomène non-linéaire très populaire pour générer des photons annoncés indiscernables. Quand le nombre de photons fils générés est très faible (le régime perturbé), l'état du photon annonceur et annoncé est toujours intriqué spectralement si la fonction d'accord de phase n'est pas une Gaussienne. À cause des termes additionnels provenant de la mise en ordre temporel présent quand le système est décrit comme étant un état compressé multimode dans le cas où le nombre de photons fils produits n'est plus aussi faible, la physique derrière le phénomène se trouve changée. Cette dissertation se concentre principalement sur présenter une méthode d'optimisation de l'enveloppe spectrale de la pompe dans le but de maximiser la séparabilité en fréquence de l'état résultant de la CPDS quand le régime n'est plus perturbé. L'optimisation sous contraintes utilisant l'algorithme BFGS montre que l'enveloppe optimale est très probablement Gaussienne dans ce régime aussi. Cet ouvrage montre aussi une méthode pour transformer les équations de mouvement gouvernants le CPDS en un problème quadratique de matrices. Des résultats sur la relaxation convexe du problème sont montrés, concluant que seulement dans certains cas, si l'enveloppe est une superposition de termes de la base d'Hermite, il n'est pas possible d'avoir une séparabilité presque parfaite. La méthode provenant de l'optimisation convexe n'est cependant pas concluante, car ces résultats ne peuvent pas être généralisés. Cependant, les résultats de la relaxation convexe apportent de nouvelles pistes de réflexion pour la minimisation de l'intrication spectrale entre les rayons fils créés par la CPDS.

ABSTRACT

Spontaneous parametric down-conversion (SPDC) is a popular way to generate indistinguishable photons through heralding. When the yield (signal and idler photons generated per pump pulse) of SPDC in waveguides is very low (perturbative regime), heralded photons cannot be indistinguishable if the phase-matching function is not Gaussian. Because of the time-ordering correction terms in the description of multimode squeezed states, SPDC at high yield follows different physics. This thesis will be focused on finding the pump's spectral envelope that would maximize the spectral separability between the idler and signal states at higher yield. Numerical constrained optimization results with BFGS algorithm show that such a pump would likely be Gaussian. A scheme to turn the dynamics of the SPDC into a quadratic matrix equations is given. Convex optimization results attempting to show that there can be no pump from a finite superposition of Hermite basis elements that can give perfect spectral separability are shown. However, those results are only valid in some restrictive cases. While inconclusive, the convex optimization results open up reflection on future models to bound the spectral separability of idler and signal states.

TABLE OF CONTENTS

DEDICATION	iii
ACKNOWLEDGEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vi
TABLE OF CONTENTS	vii
LIST OF TABLES	x
LIST OF FIGURES	xii
LIST OF SYMBOLS AND ACRONYMS	xv
LIST OF APPENDICES	xvi
CHAPTER 1 INTRODUCTION	1
1.1 Indistinguishable Photons in Quantum Information	1
1.2 Photon Generation With Spontaneous Parametric Down-Conversion in Per- turbative Regime in Waveguides	2
1.2.1 Hamiltonian of spontaneous parametric down-conversion	2
1.2.2 Joint spectral amplitude	4
1.3 Spectral Entanglement of Photons From Spontaneous Parametric Down-Conversion in Perturbative Regime	5
1.3.1 Fully indistinguishable photons	5
1.3.2 Quantitative measurement of spectral entanglement	8
1.4 SPDC in High-Gain Regime	9
1.5 Inverse Design	10
1.6 Research Objectives	10
1.7 Structure of Memoir	11
CHAPTER 2 METHODS	12
2.1 Dynamics of Annihilation and Creation Operators of Signal and Idler Beams	12
2.1.1 From the Hamiltonian to the equations of dynamics	12

2.1.2	Photon pair number and Schmidt number from propagators	14
2.1.3	Relation with perturbative regime	18
2.2	Objective Function With Decoupled Equations of Motion	19
2.2.1	Decoupling the equations of motion	19
2.2.2	Making the equations of motion dimensionless	20
2.2.3	Propagators of pseudo-quadratures	22
2.2.4	Commutation between pseudo-quadratures and relations between de- coupled propagators	24
2.2.5	Equivalent formulation of Schmidt number	25
2.3	Numerical Optimization	26
2.3.1	Quasi-Newton method	26
2.3.2	Penalty method and regularization	28
2.4	Convex Optimization and Convex Relaxations	30
2.4.1	Convex problem	30
2.4.2	Lagrangian dual of a problem	31
2.5	Quadratically Constrained Quadratic Programs (QCQP)	34
2.5.1	The Lagrangian dual of QCQP	35
2.5.2	Semidefinite relaxation	37
2.6	Spectral Minimization Problem as a Quadratically Constrained Problem . .	39
2.6.1	Equations of motion as quadratic constraints	39
2.6.2	Structure of variables	40
2.6.3	Additional quadratic constraints	41
2.6.4	Affine constraints	42
2.6.5	Real scalar valued constraints	43
2.7	Convexification of The Spectral Minimization Problem	44
2.7.1	Semidefinite relaxation of quadratic constraints	44
2.7.2	Convex objective function	45
CHAPTER 3	RESULTS	46
3.1	Numerical Optimization With JAX	46
3.1.1	Fixing the number of mean photon pairs per pulse	47
3.1.2	Optimization taking into account the variance	47
3.2	Optimization Over Pseudo-Quadratures	50
3.3	Optimizing Over Gaussians	51
3.4	Convex Relaxation	52
3.4.1	Low-gain regime lower bounds with semidefinite relaxation	53

3.4.2 High-gain regime lower bounds with semidefinite relaxation	59
CHAPTER 4 CONCLUSION	65
4.1 Summary of Works	65
4.2 Limitations	65
4.3 Future Research	65
REFERENCES	67
APPENDICES	74

LIST OF TABLES

Table 2.1	Physical parameters for an SPDC scheme within a PPKTP waveguide. These values describe the difference of inverse of group velocities, the length of the waveguide, and the strength of the second-order nonlinear interaction.	14
Table 3.1	Table showing the difference between the optimal pump and the optimal Gaussian pump. Since the value is less than 10^{-3} , it is reasonable to say the numerically optimal pump is always Gaussian.	52
Table 3.2	Lower bounds on the objective function for different values of $ V $ when $\langle n \rangle = 1$ with Hermite basis constraint and without Hermite basis constraint. When $ V \geq 1.0$, the bound is driven to 0. However, there exists a quadratically feasible pump that can give a mean number of photon pairs per pulse of 1 for $ V \geq 1.0$, meaning the semidefinite heuristics used in low-gain are not effective in high-gain. For some values of $ V $, only setting the pump to be a superposition of Hermite basis elements can make the problem infeasible with constraints on the nuclear norm of propagators.	59
Table 3.3	Optimal value of objective function as a function of $ V $ for when finite difference constraints are applied and when they are not. Here, the parameter ζ is between 0 and 0.1. When the finite difference constraints are respected, the optimal value does improve, but it still approaches 0 when $ V $ is close to the amplitude for which the low-gain optimal pump would give $\langle n \rangle = 1$	62
Table 3.4	Value of $\text{Tr} [\mathbf{V}_{\text{opt}} \mathbf{V}_{\text{opt}}^\dagger]$ as a function of $ V $. While compared to Table 3.2, lower values of $\text{Tr} [\mathbf{V}_{\text{opt}} \mathbf{V}_{\text{opt}}^\dagger]$ are often associated with a lower minimum of objective function. But when $\text{Tr} [\mathbf{V}_{\text{opt}} \mathbf{V}_{\text{opt}}^\dagger]$ is closer to 1, the optimal value of objective function is indisputably higher than when $\text{Tr} [\mathbf{V}_{\text{opt}} \mathbf{V}_{\text{opt}}^\dagger]$ is much lesser than 1.	62

Table 3.5	Bounds given by the optimization scheme when fixing the pump to be $\mathbf{V}_{\text{fix}}$, with $\mathbf{V}_{\text{fix}}$ the matrix given by the low-gain optimal pump normalized such that $\text{Tr} [\mathbf{V}_{\text{fix}} \mathbf{V}_{\text{fix}}^\dagger] = 1$. The value of objective function with matrix exponential is given by when $\mathbf{U}_{\pm N_z} = \exp [(\mathbf{K} + V \mathbf{V}_{\text{fix}}) \zeta_f]$ with $\zeta_f = 1$ and the objective function after optimization being the objective function obtained with SDR. It can be seen they are not very far off considering the numerical errors in approximating integration by summation.	63
Table D.1	Error on numerical approximation of the integral on the dynamics. When $\langle n \rangle$ increases, the error does not. Its value is however quite large, so rescaling the frequency range or the waveguide parameters may be needed to obtain better accuracy. But to do so requires an larger positive semidefinite matrix, which requires more computational ressources.	80
Table D.2	Error on numerical approximation of the integral on the dynamics at $\langle n \rangle = 1$ with low-gain optimal pump. When N_z increases, the error decreases. Indeed, rescaling the problem can lead to better accuracy in numerical approximation.	80

LIST OF FIGURES

Figure 1.1	Example of JSA using Gaussian (left) and non-Gaussian PMF (right), both obtained with a Gaussian shaped pump. On the right, the PMF used is a sinc function. The left JSA gives a value of $g^{(2)} = 2$ and the right JSA gives a lower value. The frequency range depends on the dispersion relation of the material.	7
Figure 2.1	A visualization of the frequency distribution of the signal, idler and pump beams. The representation is only there to give an idea and does not represent physical scale. For $J \in \{S, I, P\}$, it is assumed $\bar{\omega}_J$ is the center frequency of the beam J and the amplitude of said beam is within $[\omega_{J0}, \omega_{Jf}]$ and negligible outside of this range. Then, the amplitude of the beam J would be significant only on $\omega + \bar{\omega}_J \in [\omega_{J0}, \omega_{Jf}]$. The energy conservation requirement $\bar{\omega}_S + \bar{\omega}_I = \bar{\omega}_P$ for SPDC to happen is also shown.	13
Figure 2.2	Criteria for convexity for the Schmidt number function. For some range of α , the criteria is respected (≤ 0), but not for all $\alpha \in [0, 1]$, meaning the objective function is a non-convex function.	32
Figure 2.3	Criteria for convexity for the constraint on $\langle n \rangle$. For $\alpha = 0$ and $\alpha = 1$, the constraint is respected ($\langle n \rangle = 1$), but for any other value, it is not. The set generated by the constraint is then a non-convex set.	32
Figure 2.4	A visualization of the problem. The discretization step is of equal distances, meaning the length between 0 and 1 is $\Delta\zeta$, which is also the length between 1 and 2 and so on. The propagators at 0 are $\mathbf{U}_{\pm 0}$, or $\mathbf{I}$, at 1 they are $\mathbf{U}_{\pm 1}$, at 2 $\mathbf{U}_{\pm 2}$ until at N, $\mathbf{U}_{\pm N_z}$. The pulse in blue is the pump beam, the ones in orange and red are signal and idler beams with less energy than the pump pulse. They are often entangled spectrally.	41

Figure 3.1	Real part of the optimal pumps for different values of $\langle n \rangle$ with a Gaussian initialization seed. Note that the higher the gain is, the wider the optimal pump becomes. The real part of the pump is proportional to the imaginary part. It is then always possible to multiply the complex pump by another global phase such that the values of observables remain the same. In this new frame of reference, the imaginary part can be set to zero, which is why only the real part of the optimal pump is shown. The table under the visualization of the optimal pumps contains the Schmidt number obtained for each pump shown in the figure.	48
Figure 3.2	Comparison between the Schmidt number of the optimal pump using as seed the optimal Gaussian at low-gain regime and the Schmidt number of a Gaussian with width of said low-gain optimal pump and amplitude scaled to match $\langle n \rangle$. At low-gain, the Schmidt numbers are the same, but as $\langle n \rangle$ grows, the optimized pump has a lower value of K	49
Figure 3.3	Real part of the optimal pumps obtained with random initialization seeds for different $\langle n \rangle$. Inequality on the variance implied by a regularization is added as well as a penalty on the expected value of the frequency distribution of the pump. Since the imaginary part is still proportional to the real part, we are again in the frame of reference where the imaginary part is negligible, hence why it was not shown. Note that it is very similar to Fig. 3.1.	51
Figure 3.4	Normalized optimal pumps when optimizing over the pseudo-quadratures. Penalty on the frequency distribution is made such that the center frequency of the pump is zero, and regularization on the variance is also added. The results look very similar to the real part obtained in Fig. 3.3.	52
Figure 3.5	Normalized optimal Gaussian pump for different values of $\langle n \rangle$. The width increases with the increase in gain. This may be a consequence of the spectral widening of signal and idler beams observed in experiments [1]. Note that the plot looks very similar to that of Fig. 3.4. The table below shows the Schmidt number evaluated at different values of $\langle n \rangle$ for optimal Gaussians. They are very close to the Schmidt number obtained when every discretization in frequency is allowed to vary freely. Then, we can infer that any discontinuity or negative values from prior optimization schemes may simply be a numerical artifact.	53

Figure 3.6	Eigenvalues of the optimal solution of the semidefinite relaxation in perturbative regime, where λ_i represents the i th eigenvalue. The eigenvalues are classified in increasing order, meaning that $\lambda_i \geq \lambda_j$ if $i \geq j$. The N_ω largest eigenvalues are on a scale of 10^5 , but the rest of the eigenvalues are not negligible, so the semidefinite matrix cannot be said to be of rank N_ω	55
Figure 3.7	Representation of adding more convex constraints to convex relaxation with 2 variables. When more constraints are added, the convex feasible set will approximate that of the nonconvex set. Feasible points from the old relaxation with a suboptimal minimum will then be excluded from the new, tighter convex feasible set.	55
Figure 3.8	Eigenvalues of the semidefinite matrix when affine constraints on the linear part of the pump are applied to the pump in order to fix it to be $\mathbf{V}_{\text{fix}}$. λ_i is the i th eigenvalue. It is noteworthy that the smallest eigenvalues are smaller than those in Fig. 3.6, showing the solution is lower rank.	64

LIST OF SYMBOLS AND ACRONYMS

$v_{I/S/P}$	Group velocity of idler/signal/pump beam
ω	Frequency in units of s^{-1}
Ω	Unitless frequency
$\bar{\omega}_J$	Center frequency of the beam J
$\omega_J(k)$	Dispersion relations of the beam J
N_ω	Number of discretization in frequency
$\Delta\omega$	Size of discretization in frequency
$\Delta\Omega$	Size of discretization in unitless frequency
z	Frequency in units of distance
ζ	Unitless distance
N_z	Number of discretization of the waveguide
Δz	Size of discretization of the waveguide
$\Delta\zeta$	Unitless size of discretization of the waveguide
$\hbar$	Planck's constant
ℓ	Length of the waveguide
SPDC	Spontaneous parametric down-conversion
QMP	Quadratic matrix program
obj. f.	Objective function
QCQP	Quadratically constrained quadratic program
SDR	Semidefinite relaxation
SDP	Semidefinite program
argmin	point that would give minimum
argmax	point that would give maximum
H.c.	Hermitian conjugate

LIST OF APPENDICES

Appendix A	Passage from time dependent Heisenberg equations of motion to frequency modes equations of motion	74
Appendix B	Formulation of Schmidt number from the propagator	76
Appendix C	Commutation relation between the pseudo-quadrature operators . . .	79
Appendix D	Error on the numerical approximation of the equations of motion . .	80
Appendix E	Equivalence in objective function of nuclear norm and Schmidt number	81

CHAPTER 1 INTRODUCTION

1.1 Indistinguishable Photons in Quantum Information

The ability of photons to effect quantum interference is key to many quantum technologies, such as quantum metrology [2], quantum communication [3], quantum computing [4] and quantum sampling [5]. By driving 2 indistinguishable photons towards a beamsplitter, there will be either two photons or none in each of the output modes [6]. This phenomenon is often called the Hong-Ou-Mandel effect [7]. The quality of the interference is highly dependent on the indistinguishability of the photons, hence the reason why a large number of indistinguishable photons needs to be generated at a fast rate in order to scale up interference. Preferably, a perfect photon source for quantum interference should at least

- Generate photons on demand
- Output perfectly pure photon states
- Have a minimalist source design

In the event that on-demand generation is not possible, the probability of photon generation should nevertheless be high. There are various technologies to produce indistinguishable single photons. A non-exhaustive list includes vacancy centers in diamonds, semiconductor-based quantum wells, and heralded sources [8].

Technologies based on vacancy centers in diamonds attempt to engineer the impurities within diamond lattices in order to produce well-controlled states of light. Issues with this platform include the possibility of differences in the properties between photons produced by different centers [9] because of the variation of local stress felt by individual emitters. Another problem with this method is the engineering difficulty of growing the defect at an exact location.

Quantum wells provide another way of generating indistinguishable single photons. However, a problem they face as well is the difficulty of operation. While the single photons generated can be nearly indistinguishable with a good generation rate, the temperature of operation needs to be cryogenic [10]. Also, as with defects in diamonds and other atom-based photon generation sources, for such systems to scale up, every individual emitter needs to be carefully engineered, as the tolerance for fabrication errors can be low.

Single-photon generation by heralding is based on bulk properties of the material instead of carefully engineered individual photon emitters [11]. The quality of photons produced by this process is dependent on the properties of the material, such as its nonlinear susceptibility and its dispersion relations. The process of heralding is a conditional photon generation method. The working principle of such a source is based on the generation of two photons at the same time. One of these photons will be measured and will herald the generation of the second photon [12]. They are often called signal and idler photons. Such sources can operate at room temperature, and the heralded photons can have high indistinguishability between themselves. However, the photon pair generation is a probabilistic event.

1.2 Photon Generation With Spontaneous Parametric Down-Conversion in Perturbative Regime in Waveguides

1.2.1 Hamiltonian of spontaneous parametric down-conversion

Spontaneous parametric down-conversion (SPDC) is a common nonlinear process used for heralding [12] since it makes use of the second-order nonlinear susceptibility, which is stronger than any of the higher-order ones [13]. While SPDC can happen within microresonators such as microrings, the nonlinear interaction is dependent on the quality factor of the resonator [14]. The slowly varying spectral envelope of the signal and idler beams is often Lorentzian-shaped because of the resonance within the rings, something that does not occur within SPDC in waveguides. Another issue with ring resonators is their sensitivity to thermal fluctuations [15]. The waveguide case is also interesting due to its on-chip scalability and relative ease of fabrication.¹

The description of the SPDC process begins by observing that it is governed by a squeezing Hamiltonian. With the quantization of the field $\vec{D}(\vec{r})$ instead of the typical $\vec{E}(\vec{r})$ [17] for nonlinear interaction, as the use of the field $E(\vec{r})$ can lead to contradictions with existing theory [18,19], the Hamiltonian in the Schödinger picture is given. For second-order nonlinear interaction, it is

$$\hat{H}_{\text{NL}} = -\frac{1}{3\epsilon_0} \int d\vec{r} \Gamma_2^{(ijl)}(\vec{r}) D^{(i)}(\vec{r}) D^{(j)}(\vec{r}) D^{(l)}(\vec{r}), \quad (1.1)$$

where ϵ_0 is the vacuum permittivity, $\Gamma_2^{(ijl)} = \frac{\chi_2^{(ijl)}(\vec{r})}{\epsilon_0 \epsilon_1^3(\vec{r})}$ a term related to the nonlinear suscepti-

¹It is worth noting that there are two common types of SPDC: degenerate and non-degenerate cases. The non-degenerate case allows heralding, as the signal and idler photons can be driven towards different channels. Even though heralding cannot be done with the degenerate case, it is still used in key applications of quantum optics, such as Gaussian boson sampling [16].

bility tensor and the permittivity of the material. The terms i, j and l iterate over the axis in space, namely x, y and z . The term $D^{(i)}(\vec{r})$ is the component in i of the vector $\vec{D}(\vec{r})$. Note the absence of t within the Hamiltonian, since we are working in the Schödinger picture. Quantizing the field $D(\vec{r})$ such as

$$\begin{aligned}\vec{D}(\vec{r}) &\rightarrow \hat{\vec{D}}(\vec{r}) = \int \sum_{J \in \{S, I, P\}} dk \sqrt{\frac{\hbar \omega_J(k)}{4\pi}} \mathbf{d}_{Jk}(x, y) \hat{a}_J(k) \exp[ikz] + \text{H.c.} \\ &\approx \int \sum_{J \in \{S, I, P\}} dk \sqrt{\frac{\hbar \bar{\omega}_J}{4\pi}} \mathbf{d}_J(x, y) \hat{a}_J(k) \exp[ikz] + \text{H.c.},\end{aligned}\quad (1.2)$$

assuming $\omega_J(k) \approx \bar{\omega}_J$ and $\mathbf{d}_{Jk}(x, y) \approx \mathbf{d}_J(x, y)$. Here, z is the propagation direction, $\mathbf{d}_{Jk}(x, y)$, $\omega_J(k)$, $\bar{\omega}_J$, $\mathbf{d}_J(x, y)$ and $\hat{a}_J(k)$ are, respectively, for the beam J , a function of its polarization for different values of k , its dispersion relation, the center frequency, the function of polarization at central wavenumber, and its annihilation operator. For the other variables, $\hbar$ is Plank's constant and H.c. stands for Hermitian conjugate. The wavenumber is represented by k . The values taken by J are $J \in \{S, I, P\}$. Assuming the pump is undepleted, the nonlinear Hamiltonian is [20]

$$\hat{H}_{\text{SPDC}} = -\sqrt{v_I v_S v_P} \int_0^\ell dz \int dk dk' dk'' \zeta(z) \hat{a}_I^\dagger(k) \hat{a}_S^\dagger(k') \alpha(k'') \exp[-i(k + k' - k'')z] + \text{H.c.}, \quad (1.3)$$

$$\zeta(z) = \frac{1}{\epsilon_0} \sqrt{\frac{\hbar^3 \bar{\omega}_I \bar{\omega}_S \bar{\omega}_P}{(4\pi)^3 v_I v_S v_P}} \int dx dy \Gamma_2^{(ijl)}(\vec{r}) \left[d_I^{(i)}(x, y) d_S^{(j)}(x, y) \right]^* d_P^{(l)}(x, y), \quad (1.4)$$

where $\alpha(k)$ is the envelope of the pump as the beam is treated classically, ℓ the length of the waveguide, v_J and $\bar{\omega}_J$ are the group velocity at center frequency and the center frequency of beam J respectively. It is indeed a quadratic function in creation and annihilation operators. Taking the Hamiltonian associated with the linear interaction without zero-point energy, one obtains

$$\hat{H}_L = \sum_{J \in \{S, I, P\}} \int dk \hbar \omega_J(k) \hat{a}_J^\dagger(k) \hat{a}_J(k). \quad (1.5)$$

The total Hamiltonian of the system is then $\hat{H} = \hat{H}_L + \hat{H}_{\text{SPDC}}$.

1.2.2 Joint spectral amplitude

With an assumption on negligible group velocity dispersion, we approximate

$$k_J(\omega) \approx k_J(\bar{\omega}_J) + \frac{\omega - \bar{\omega}_J}{v_J}, \quad (1.6)$$

where $k_J(\omega)$ is the wavenumber at frequency ω . Exiting the Schrödinger picture, one can write the Hamiltonian in the interaction picture and changing from wavenumber to frequency [18]

$$\begin{aligned} \hat{H}_I(t) &= -\zeta \int d\omega d\omega' d\omega'' \int_0^\ell dz \exp[i(\omega + \omega' - \omega'')t] \\ &\quad \exp[-i(k_S(\omega) + k_I(\omega') - k_P(\omega''))z] \hat{a}_S^\dagger(\omega) \hat{a}_I^\dagger(\omega') \alpha(\omega'') + \text{H.c.}, \end{aligned} \quad (1.7)$$

if the nonlinear susceptibility is independent of z . Here, k , k' and k'' are replaced by the terms $k_S(\omega)$, $k_I(\omega')$ and $k_P(\omega'')$, which are the dispersion relations for the signal, idler and pump beams. With $\mathcal{T}$ representing the time-ordering operator, the state at a future time would be given by

$$|\psi(t)\rangle = \mathcal{T} \exp \left[\frac{1}{i\hbar} \int_{t_0}^t dt' \hat{H}_I(t') \right] |\psi_0\rangle \approx \frac{1}{i\hbar} |\psi_0\rangle + \frac{1}{i\hbar} \int_{t_0}^t dt' \hat{H}_I(t') |\psi(t_0)\rangle, \quad (1.8)$$

by removing all higher-order terms. This approximation is valid if the system is in low-gain or perturbative regime. If the state $|\psi_0\rangle$ is the vacuum, then in terms of photon number states,

$$|\psi(t)\rangle \approx \sqrt{1 - \mathcal{N}^2} |\text{vac}\rangle + \mathcal{N} |\text{II}\rangle, \quad (1.9)$$

with $\mathcal{N}^2 \ll 1$ and where we take $\mathcal{N}$ to be real and positive without loss of generality, hence why low-gain. The state $|\text{II}\rangle$ represents the photon pair state (one photon for the idler beam and another one for the signal beam). The integral in Eq. (1.8) is given by

$$\begin{aligned} &\int_{t_0}^t dt' \hat{H}_I(t') \\ &= -2\pi\zeta \int d\omega d\omega' \int_0^\ell dz \exp[-i(k_S(\omega) + k_I(\omega') - k_P(\omega + \omega'))z] \hat{a}_S^\dagger(\omega) \hat{a}_I^\dagger(\omega') \alpha(\omega + \omega') + \text{H.c.} \end{aligned} \quad (1.10)$$

Eq. (1.10) is obtained by keeping in mind that the integral over t is $2\pi\delta(\omega + \omega' - \omega'')$, then the integral over ω'' replaces ω'' by $\omega + \omega'$. The term $\alpha(\omega + \omega')$ would represent the pump's

slowly varying spectral envelope. Eq. (1.10) becomes

$$\begin{aligned} & \int_{t_0}^t dt' \hat{H}_I(t') \\ &= -\pi\zeta\ell \int d\omega d\omega' \alpha(\omega + \omega') \Phi(\omega, \omega') a_S^\dagger(\omega) a_I^\dagger(\omega') \times \\ & \quad \exp \left[i\frac{\ell}{2} \left(\left(\frac{1}{v_S} - \frac{1}{v_P} \right) (\omega - \bar{\omega}_S) + \left(\frac{1}{v_I} - \frac{1}{v_P} \right) (\omega' - \bar{\omega}_I) \right) \right] + \text{H.c.}, \end{aligned} \quad (1.11)$$

after integration over z on the domain from 0 to ℓ , where ℓ is the length of the waveguide, and with the assumption of negligible group velocity dispersion. The term

$$\begin{aligned} \Phi(\omega, \omega') &= \text{sinc} \left[(k_I(\omega) + k_S(\omega') - k_P(\omega + \omega')) \frac{\ell}{2} \right] \\ &= \text{sinc} \left[\left(\left(\frac{1}{v_S} - \frac{1}{v_P} \right) (\omega - \bar{\omega}_S) + \left(\frac{1}{v_I} - \frac{1}{v_P} \right) (\omega' - \bar{\omega}_I) \right) \frac{\ell}{2} \right] \\ &= \text{sinc} \left[\Delta k(\omega, \omega') \frac{\ell}{2} \right] \end{aligned} \quad (1.12)$$

is called the phase-matching function (PMF), assuming negligible group velocity dispersion. The multiplication between the pump's slowly varying envelope and the PMF gives the joint spectral amplitude (JSA), or more exactly

$$\text{JSA}(\omega, \omega') = -\pi\zeta\ell\alpha(\omega + \omega')\Phi(\omega, \omega') \quad (1.13)$$

Here it is taken into account that $k_P(\bar{\omega}_S + \bar{\omega}_I) - k_S(\bar{\omega}_S) - k_I(\bar{\omega}_I) = 0$. The function $\alpha(\omega + \omega')\Phi(\omega, \omega')$ is then the JSA. Again, if $|\psi(t_0)\rangle$ is the vacuum ($|\text{vac}\rangle$),

$$|\psi(t)\rangle \approx |\text{vac}\rangle + \zeta_1 \int d\omega d\omega' \alpha(\omega + \omega') \Phi(\omega, \omega') |\text{II}\rangle, \quad (1.14)$$

where $\zeta_1 \propto \zeta$ and the prefactor multiplying $|\text{vac}\rangle$ is approximated to be 1.

1.3 Spectral Entanglement of Photons From Spontaneous Parametric Down-Conversion in Perturbative Regime

1.3.1 Fully indistinguishable photons

While perfect separability is desired, one must approach such a subject with care. Almost true indistinguishability has been shown to be realizable [21] with a BBO crystal waveguide. The separability comes from the narrow distribution in ω (or ω') of the term $\Phi(\omega, \omega')$ in the

JSA but a broad one in ω' (or ω) by tailoring the group velocities at the center frequencies of the signal, idler and pump beams. Because of the difference in the distribution of the JSA in ω and ω' , the bandwidths of the signal and idler photons differ greatly. This generation scheme may not be optimal for applications where signal and idler bandwidths are important.

From the description of the joint spectral amplitude (JSA) as a result of multiplication between the term conserving the energy ($\alpha(\omega + \omega')$) and the term conserving the momentum ($\Phi(\omega, \omega')$), conditions on the phase-matching function (PMF) and the pump envelope (PE) can enable truly separable twin-photon states, where the idler and signal beams are not spectrally entangled. If the PMF and PE are both Gaussian, the JSA can be written as a multiplication between functions in frequency of the idler and signal beams [22]. In fact, it has been shown that perfect separability can only be obtained if and only if both functions are Gaussians [23]. If the PMF is not Gaussian, then to minimize the entanglement, the pump must be Gaussian.

Spectral separability has also been shown experimentally for different shapes of PMF by measuring $g^{(2)}(0)$, or second-order correlation function at zero delay, sometimes written as simply $g^{(2)}$. For non-Gaussian PMF, the spectral separability is indeed always worse than if the PMF were Gaussian [24]. Although even if the shapes are different, engineering the PMF such that its width is similar to that of the would-be optimal Gaussian PMF can minimize the spectral entanglement between the idler and signal beams. In Fig. 1.1, an example of JSA from a Gaussian PMF and one from a non-gaussian PMF are shown. When the PMF is Gaussian, there is no spectral entanglement between the signal and idler beam and $g^{(2)} = 2$, while when the PMF is non-Gaussian, it is between $1 \leq g^{(2)} < 2$ [25].

To generate a Gaussian PMF, modification to the waveguide is needed. A popular method is to modulate the value of the nonlinear susceptibility between domains of the waveguide [26]. The reasoning for such modification is that PMF can be written as

$$\Phi(\omega_i, \omega_s) = \sqrt{2\pi} \int dz \chi_{\text{mod}}(z) \exp[i\Delta k(\omega_i, \omega_s)z], \quad (1.15)$$

where $\chi_{\text{mod}}(z)$ is the normalized second-order nonlinear susceptibility. For the previous formulation of PMF, it was assumed that the nonlinear susceptibility is constant through the waveguide. If this is not the case, then the PMF needs to take it into account. Modifying $\chi_{\text{mod}}(z)$ results in directly changing $\Phi(\omega_i, \omega_s)$. To modulate $\chi_{\text{mod}}(z)$, a very popular method is ferroelectric domain poling. For different sections along the axis z of the waveguide, the

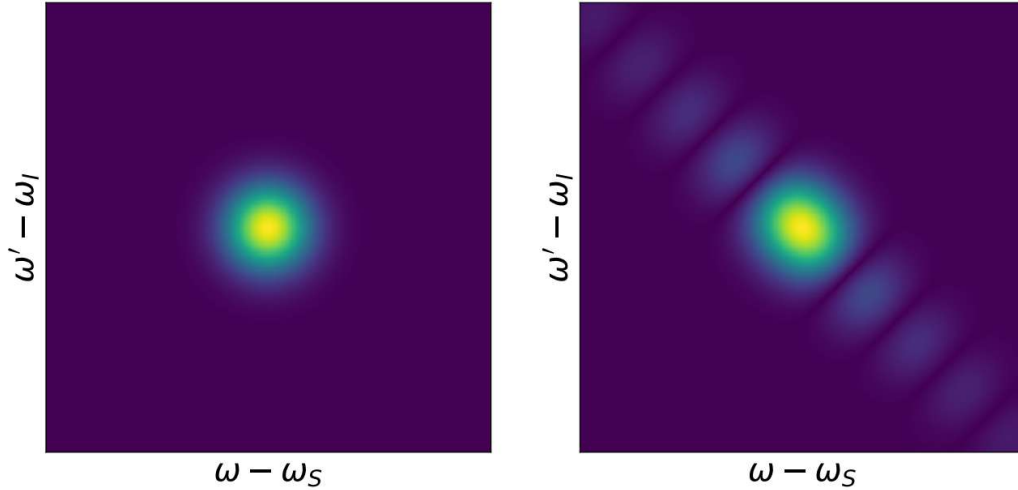


Figure 1.1 Example of JSA using Gaussian (left) and non-Gaussian PMF (right), both obtained with a Gaussian shaped pump. On the right, the PMF used is a sinc function. The left JSA gives a value of $g^{(2)} = 2$ and the right JSA gives a lower value. The frequency range depends on the dispersion relation of the material.

value of $\chi_{\text{mod}}(z)$ is carefully engineered. This change in nonlinear susceptibility of ferroelectric material can be done by applying an electric field at different sections that will be called domains [26]. Then, $\chi_{\text{mod}}(z)$ can be written as a Fourier series

$$\chi_{\text{mod}}(z) = \sum_{-\infty}^{\infty} G_m \exp[iK_m z], \quad (1.16)$$

$$K_m = \frac{2\pi m}{\Lambda}, \quad (1.17)$$

where G_m alternates between 1 and -1 , Λ is the poling period length and m represents the m th domain. By changing Λ for different m so it becomes Λ_m , it is possible to obtain a Gaussian PMF [27]. Other methods of crystal poling exist to give the total PMF a Gaussian shape without changing the structure of phase-matching conditions [28–30]; instead of changing the length of each period, the sign of G_m is chosen wisely with Λ constant. Either way, the waveguide needs to go through additional engineering, and the material used for the waveguide needs to have ferroelectric properties so it can be poled, which can limit the type of materials and can add complications to waveguide manufacturing.

When the crystal is not being poled, it is possible to obtain better separability using a spectral filter on the signal and idler photons [31]. This action, however, decreases the power of both beams and decreases SPDC efficiency. An ideal heralded source would be one that does not

need poling or filtering, yet still gives good spectral separability. In the perturbative regime, if the bandwidth of the signal beam is comparable to that of the idler beam, this source does not exist.

1.3.2 Quantitative measurement of spectral entanglement

It is valuable to find a quantitative measurement of the spectral separability of the signal and idler beams when the PMF is not Gaussian. Experimentally, the indistinguishability of heralded photons can be measured by the probability of photons not interfering with each other. This probability is directly related to the JSA, more precisely to its number of Schmidt modes [32]. In fact, the probability of heralded photons interfering with each other can be obtained with a partial trace on the state $|\text{II}\rangle$ in Eq. (1.14) [33]. The Schmidt number, a value closely related to the number of Schmidt modes, is often a good indication of the entanglement between two states [34–36]. This value is typically used when a state is written as

$$|\phi\rangle = \sum_{n,v} C(n,v) |n\rangle |v\rangle, \quad (1.18)$$

and one wants to know how entangled $|\phi\rangle$ is. From matrix representation of $C(n,v)$ and singular value decomposition, the state can be written as

$$|\phi\rangle = \sum_{n,v} \mathbf{C}^{(n,v)} |n\rangle |v\rangle = \sum_{n,v} \left(\sum_{\alpha} \sqrt{\lambda_{\alpha}} s_{\alpha}^{(n)} w_{\alpha}^{(v)\dagger} \right) |n\rangle |v\rangle, \quad (1.19)$$

$$= \sum_{\alpha} \sqrt{\lambda_{\alpha}} |N_{\alpha}\rangle |M_{\alpha}\rangle, \quad (1.20)$$

where $\mathbf{C}^{(n,v)}$ is the matrix element of $\mathbf{C}$ associated with the state $|n\rangle |v\rangle$, $|N_{\alpha}\rangle = \sum_n s_{\alpha}^{(n)} |n\rangle$ and $|M_{\alpha}\rangle = \sum_v w_{\alpha}^{(v)\dagger} |v\rangle$. The Schmidt number is defined as

$$K = \frac{(\sum_{\alpha} \lambda_{\alpha})^2}{\sum_{\alpha} \lambda_{\alpha}^2}, \quad (1.21)$$

which is $\frac{1}{\sum_{\alpha} \lambda_{\alpha}^2}$ if the state $|\phi\rangle$ is normalized.

While the Schmidt number is a good indicator on the separability of the twin-photon state, a more general evaluation is using the second-order correlation function ($g^{(2)}$). Its calculation directly uses the creation and annihilation operators of either the signal beam or the idler beam [37] instead of relying on the singular value decomposition of $\alpha(\omega + \omega')\Phi(\omega, \omega')$. The

$g^{(2)}$ value of the SPDC process is

$$g^{(2)} = 1 + \frac{\sum_k \sinh^4(r_k)}{\left(\sum_k \sinh^2(r_k)\right)^2}, \quad (1.22)$$

where r_k are the singular values of the kernel of the integral within the interaction Hamiltonian. Note that in the perturbative regime, $g^{(2)} \approx 1 + \frac{1}{K}$.

1.4 SPDC in High-Gain Regime

A problem with the high-gain regime is the trade-off between the probability of SPDC producing more photons and the fidelity of the process [38]. Defining the probability of SPDC to produce photons as the probability that a measurement of the heralding state yields a result different from vacuum and the fidelity as the probability of one measurement yielding a single photon, one increases when the other decreases. This is a problem often encountered when measuring with threshold detectors [39, 40] (detectors that click when photons are detected).

Using photon-number-resolving detectors (PNRDs), the single photon fidelity can stay at 1 with the probability that measuring the heralded state will yield a single photon to be $\frac{1}{4}$ [38]. This result is obtained by taking into account that measuring a single photon with the PNRDs is modeled by $\hat{\Pi} = (|1\rangle\langle 1|)_{\text{heralded}}$. While PNRDs typically need to be cooled at temperatures below 1 K, recent advances in photon-resolving detection with superconducting nanowire detectors [41] are allowing PNRDs to be more accessible. High-gain regime SPDC is then becoming a more and more attractive source of single photons with high yield.

Save for the multi-photon event noise, another problem of high-gain SPDC is its more complex theory compared to the perturbative regime. For the non-perturbative SPDC, the JSA is different from the low-gain regime. Additional time-ordering terms of the evolution operator from the interaction Hamiltonian need to be taken into account [42, 43]. The pump shape maximizing the spectral separability is then not guaranteed to be Gaussian anymore. It has also been shown that the spectral width of signal and idler beams increases as the gain increases both experimentally [1] and theoretically [44], another hint that the physics at high-gain differs from that of the perturbative regime.

It has been shown from simulation that with an unpoled waveguide, in the high-gain regime,

if the pump has a Gaussian shape with the same width as the optimal low-gain pump, it is impossible to obtain perfectly indistinguishable heralded photons [45]. Even with domain engineering, with the distortions of the JSA caused by nonnegligible time-ordering terms, the Schmidt number increases with the gain. Inevitably, it leads one to wonder if there are any pump shapes that can counter the spectral broadening caused by the increased pump power.

In other words, what are the spectral characteristics of a pump that minimizes the Schmidt number at different values of the mean photon pair per pulse?

Because poling would require engineering the ferroelectric domains of the waveguide, the unapodized case will be of interest here.

1.5 Inverse Design

When human intuition is not enough, computers can fulfill the need for better design. Such is the power of inverse design; using optimization (either based on data [46] or mathematical optimization techniques [47]) to obtain solutions that may fare better than if obtained with analytical calculations. Inverse design has often been used in photonics, albeit on the structure geometries [48, 49].

For quantum information, various uses of inverse design have seen light, such as quantum circuit modeling with wavefront-matching algorithm [50] and optimization of quantum dot structure [51]. Attempts at applying inverse design to SPDC while keeping non-perturbative phenomena in mind have been made, but still the focus was on the 3D structure of the nonlinear crystal [52]. The principle of inverse design is simple yet powerful; assume a set of observables Σ and physically feasible parameters X where $\Sigma = f(X)$, the solution to the inverse design problem is X^* minimizing $||\Sigma - \tilde{\Sigma}||$ where $\tilde{\Sigma}$ is a targeted set of observables and $||\cdot||$ is any pertinent norm. The rest of the manuscript will be focused on applying this idea to the case of high-gain SPDC within waveguides with negligible group velocity dispersion in the unpoled case.

1.6 Research Objectives

This project's goal includes:

1. Evaluate the spectral content of the pump beam that would maximize the spectral

separability between the idler and signal beam through numerical optimization in high-gain regime.

2. Write the dynamics of SPDC and the constraints on the mean number of photon pairs as quadratic equations and find a convex relaxation for the spectral separability problem.
3. Prove that there can be no pump beam that can give perfect spectrally separable idler and signal states.

1.7 Structure of Memoir

The next sections of the memoir will go into details about the optimization models used for this project, as well as how they are used within the context of SPDC spectral separability. Both numerical and convex optimization models will be explained. A description of the theory of high-gain regime SPDC, followed by decoupling of Heisenberg's equations of motion, will be made. An overview of the theory of numerical and convex optimization will be shown. Finally, the convex relaxation of the spectral separability within the framework of this project will be explained. The results of both numerical and convex optimization will then be shown.

CHAPTER 2 METHODS

2.1 Dynamics of Annihilation and Creation Operators of Signal and Idler Beams

2.1.1 From the Hamiltonian to the equations of dynamics

Rephrasing the linear Hamiltonian in Eq. (1.5) with negligible group velocity dispersion, changing the space of representation by defining

$$\hat{\psi}_J(z) = \frac{1}{\sqrt{2\pi}} \int dk \hat{b}_J(k) e^{i(k-k_J)z}, \quad (2.1)$$

it can be written as

$$\begin{aligned} \hat{H}_L &= \sum_J \int dk \left(\hbar\omega(k_J) \hat{b}_J^\dagger(k) \hat{b}_J(k) + \hbar v_J [k - k_J] \hat{b}_J^\dagger(k) \hat{b}_J(k) \right) \\ &= \sum_J \int dz \left(\hbar\omega(k_J) \hat{\psi}_J^\dagger(z) \hat{\psi}_J(z) + \frac{i}{2} \hbar v_J \left([\partial_z \hat{\psi}_J^\dagger(z)] \hat{\psi}_J(z) - \hat{\psi}_J^\dagger(z) [\partial_z \hat{\psi}_J(z)] \right) \right). \end{aligned} \quad (2.2)$$

The nonlinear Hamiltonian of SPDC can be written as

$$\hat{H}_{NL} = -\hbar \int dz \xi(z) \hat{\psi}_S^\dagger(z) \hat{\psi}_I^\dagger(z) \hat{\psi}_P(z) + \text{H.c.}, \quad (2.3)$$

where H.c. stands for Hermitian conjugate and $\xi(z)$ is associated with the polarization of the signal, idler and pump beams as well as with the $\chi_2(z)$ strength and the permittivity of the material. With the assumption that the pump acts as a classical field $\hat{\psi}_P(z, t) = \langle \hat{\psi}_P(z, t) \rangle$, using Heisenberg equations of motion, the time-evolution of the operators $\hat{\psi}_J(z)$ can be seen as

$$\partial_t \hat{\psi}_S(z, t) = [\hat{\psi}_S(z, t), \hat{H}_L + \hat{H}_{NL}] = i\xi(z) \langle \hat{\psi}_P(z, t) \rangle \hat{\psi}_I^\dagger(z, t) - v_S \partial_z \hat{\psi}_S(z, t) - i\bar{\omega}_S \hat{\psi}_S(z, t), \quad (2.4)$$

$$\partial_t \hat{\psi}_I^\dagger(z, t) = [\hat{\psi}_I^\dagger(z, t), \hat{H}_L + \hat{H}_{NL}] = -i\xi^*(z) \langle \hat{\psi}_P^\dagger(z, t) \rangle \hat{\psi}_S(z, t) - v_I \partial_z \hat{\psi}_I^\dagger(z, t) + i\bar{\omega}_I \hat{\psi}_I^\dagger(z, t), \quad (2.5)$$

$$(\partial_t + v_P \partial_z + i\bar{\omega}_P) \langle \hat{\psi}_P(z, t) \rangle = 0. \quad (2.6)$$

The pump's field can be given by the solution of (2.6), being $\langle \hat{\psi}_P(z, t) \rangle = \psi_0(z - v_P t) e^{-i\bar{\omega}_P t}$, so we can say it is undepleted. An idea of the frequency distribution of signal, idler and pump beams can be shown in Fig. 2.1.

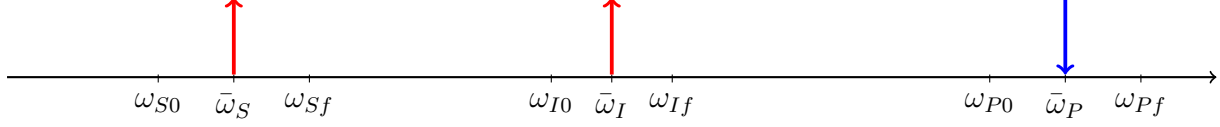


Figure 2.1 A visualization of the frequency distribution of the signal, idler and pump beams. The representation is only there to give an idea and does not represent physical scale. For $J \in \{S, I, P\}$, it is assumed $\bar{\omega}_J$ is the center frequency of the beam J and the amplitude of said beam is within $[\omega_{J0}, \omega_{Jf}]$ and negligible outside of this range. Then, the amplitude of the beam J would be significant only on $\omega + \bar{\omega}_J \in [\omega_{J0}, \omega_{Jf}]$. The energy conservation requirement $\bar{\omega}_S + \bar{\omega}_I = \bar{\omega}_P$ for SPDC to happen is also shown.

We can change to a moving frame

$$\hat{\psi}_J(z, t) = \frac{1}{\sqrt{2\pi v_J}} \int d\omega e^{-i(\omega + \bar{\omega}_J)t} e^{i\omega \frac{z}{v_P}} \hat{a}_J(z, \omega). \quad (2.7)$$

The different terms appearing in Eq. (2.4) can be written as

$$\partial_t \hat{\psi}_S(z, t) = \frac{1}{\sqrt{2\pi v_S}} \int d\omega [-i(\omega + \bar{\omega}_S)] e^{-i(\omega + \bar{\omega}_S)t} e^{i\omega \frac{z}{v_P}}, \quad (2.8a)$$

$$\begin{aligned} i\xi(z) \langle \hat{\psi}_P(z, t) \rangle \hat{\psi}_I^\dagger(z, t) &= \frac{i\xi(z)}{2\pi \sqrt{v_P v_I}} \int d\omega' d\omega \exp[-i(\omega + \omega' + \bar{\omega}_S + \bar{\omega}_I)t] \exp\left[i(\omega + \omega') \frac{z}{v_P}\right] \\ &\quad \exp[i(\omega' + \omega_I)t] \exp\left[-i\omega' \frac{z}{v_P}\right] \langle \hat{a}_P(z, \omega + \omega')t \rangle \hat{a}_I^\dagger(z, \omega'), \end{aligned} \quad (2.8b)$$

$$v_S \partial_z \hat{\psi}_S(z, t) = \sqrt{\frac{v_S}{2\pi}} \int d\omega \exp[-i(\omega + \bar{\omega}_S)t] \exp\left[i\frac{z}{v_P}\right] \left(\frac{i\omega}{v_P} + \partial_z \hat{a}_S(z, \omega)\right). \quad (2.8c)$$

Putting them together, the dynamics for $\hat{a}_S(z, \omega)$ is

$$\partial_z \hat{a}_S(z, \omega) = \frac{i}{\Delta v_S} \omega \hat{a}_S(z, \omega) + \frac{i\xi(z)}{\sqrt{2\pi}} \int d\omega' \beta(\omega + \omega') \hat{a}_I^\dagger(z, \omega'), \quad (2.9)$$

where $\frac{1}{\Delta v_{S/I}} = \frac{1}{v_{S/I}} - \frac{1}{v_P}$ and $\beta(\omega + \omega') = \langle \hat{a}_P(\omega + \omega') \rangle$. The $\bar{\omega}_S$ disappears from the equation. Appendix A shows the details of the calculation. For $\hat{a}_I^\dagger(z, \omega)$, with the same process, its

equation of motion [53]

$$\partial_z \hat{a}_I^\dagger(z, \omega) = -\frac{i}{\Delta v_I} \omega \hat{a}_I^\dagger(z, \omega) - \frac{i\xi^*(z)}{\sqrt{2\pi}} \int d\omega' \beta^*(\omega + \omega') \hat{a}_S(z, \omega') \quad (2.10)$$

can be obtained. For the pump, the differential equation is

$$\partial_z \langle \hat{a}_P(z, \omega) \rangle = 0, \quad (2.11)$$

showing its spectral distribution does not change throughout the waveguide. Typical parameter values, for an SPDC scheme in PPKTP waveguide, are shown in Table 2.1 [54].

Table 2.1: Physical parameters for an SPDC scheme within a PPKTP waveguide. These values describe the difference of inverse of group velocities, the length of the waveguide, and the strength of the second-order nonlinear interaction.

parameter	value
$\frac{1}{v_S} - \frac{1}{v_P}$	$1.70 \times 10^{-10} \text{ s} \cdot \text{m}^{-1}$
$\frac{1}{v_P} - \frac{1}{v_I}$	$1.02 \times 10^{-10} \text{ s} \cdot \text{m}^{-1}$
$ \xi(z) $	$2.80 \times 10^1 \text{ W}^{-\frac{1}{2}} \text{m}^{-1}$
ℓ	$8 \times 10^{-3} \text{ m}$

2.1.2 Photon pair number and Schmidt number from propagators

The joint-spectral amplitude (JSA) of the high-gain SPDC can be obtained from the solutions of the coupled differential equations [55] from Eq. (2.9) and (2.10). These solutions are in the form of

$$\hat{a}_S(z_1, \omega) = \int U_{SS}(z_1 - z_0, \omega, \omega') \hat{a}_S(z_0, \omega') d\omega' + \int U_{SI}(z_1 - z_0, \omega, \omega') \hat{a}_I^\dagger(z_0, \omega') d\omega', \quad (2.12a)$$

$$\hat{a}_I(z_1, \omega) = \int U_{IS}^*(z_1 - z_0, \omega, \omega') \hat{a}_S(z_0, \omega') d\omega' + \int U_{II}^*(z_1 - z_0, \omega, \omega') \hat{a}_I^\dagger(z_0, \omega') d\omega', \quad (2.12b)$$

with $z_1 - z_0 = \ell \geq 0$, where z_0 is the starting position of the waveguide and z_1 the end. The terms $U_{SS}(z_1 - z_0, \omega, \omega')$, $U_{SI}(z_1 - z_0, \omega, \omega')$, $U_{IS}^*(z_1 - z_0, \omega, \omega')$ and $U_{II}^*(z_1 - z_0, \omega, \omega')$ are integral kernels that are related to the parameters of the problem once they are given matrix form. The integral can be discretized as a linear combination of the annihilation and creation operators at different frequencies. For example, $\omega_0 = -\frac{N_\omega-1}{2}\Delta\omega$ and $\omega_{N_\omega} = \frac{N_\omega-1}{2}\Delta\omega$, where the number of discretization steps in frequency is N_ω , which is always an odd number so that

$\omega \frac{N_\omega - 1}{2} = 0$ and the step of discretization is $\Delta\omega$. Let the kernels in Eq. (2.12b) and (2.12a) be represented as

$$U_{SS}(z_1 - z_0, \omega, \omega') d\omega \equiv U_{SS}(z_1 - z_0, \omega_i, \omega_j) \Delta\omega = (\mathbf{U}_{SS}(z_1, z_0))_{ij}, \quad (2.13)$$

$$U_{SI}(z_1 - z_0, \omega, \omega') d\omega \equiv U_{SI}(z_1 - z_0, \omega_i, \omega_j) \Delta\omega = (\mathbf{U}_{SI}(z_1, z_0))_{ij}, \quad (2.14)$$

$$U_{IS}^*(z_1 - z_0, \omega, \omega') d\omega \equiv U_{IS}^*(z_1 - z_0, \omega_i, \omega_j) \Delta\omega = (\mathbf{U}_{IS}^*(z_1, z_0))_{ij}, \quad (2.15)$$

$$U_{II}^*(z_1 - z_0, \omega, \omega') d\omega \equiv U_{II}^*(z_1 - z_0, \omega_i, \omega_j) \Delta\omega = (\mathbf{U}_{II}^*(z_1, z_0))_{ij}, \quad (2.16)$$

the matrix

$$\mathbf{U}(z_1, z_0) = \begin{bmatrix} \mathbf{U}_{SS}(z_1, z_0) & \mathbf{U}_{SI}(z_1, z_0) \\ \mathbf{U}_{IS}^*(z_1, z_0) & \mathbf{U}_{II}^*(z_1, z_0) \end{bmatrix} \quad (2.17)$$

is the propagator of the annihilation and creation operators. The blocks of the propagator follow relations shown in Eqs. (2.18), which in turn are based on commutation relations respected by $\hat{a}_S(z, \omega)$ and $\hat{a}_I(z, \omega)$

$$\mathbf{U}_{SS}(z_1, z_0) \mathbf{U}_{SS}(z_1, z_0)^\dagger - \mathbf{U}_{SI}(z_1, z_0) \mathbf{U}_{SI}(z_1, z_0)^\dagger = \mathbf{I}, \quad (2.18a)$$

$$\mathbf{U}_{II}(z_1, z_0) \mathbf{U}_{II}(z_1, z_0)^\dagger - \mathbf{U}_{IS}(z_1, z_0) \mathbf{U}_{IS}(z_1, z_0)^\dagger = \mathbf{I}, \quad (2.18b)$$

$$\mathbf{U}_{SS}(z_1, z_0) \mathbf{U}_{IS}(z_1, z_0)^\dagger - \mathbf{U}_{SI}(z_1, z_0) \mathbf{U}_{II}(z_1, z_0)^T = \mathbf{0}, \quad (2.18c)$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{0}$ the element-wise 0 matrix [55]. The commutation respected by $\hat{a}_S(z, \omega)$ and $\hat{a}_S(z, \omega)$ are

$$[\hat{a}_{S/I}(z, \omega), \hat{a}_{S/I}^\dagger(z, \omega')] = \delta(\omega - \omega'), \quad (2.19)$$

$$[\hat{a}_{S/I}(z, \omega), \hat{a}_{I/S}^\dagger(z, \omega')] = [\hat{a}_{S/I}(z, \omega), \hat{a}_{I/S}(z, \omega')] = 0, \quad (2.20)$$

where $\delta(\omega - \omega')$ is Dirac's delta. In the same way, the creation and annihilation operators can be vectorized as

$$(\hat{\mathbf{a}}_S(z))_i = \hat{a}_S(z, \omega_i), \quad (2.21)$$

$$(\hat{\mathbf{a}}_I^\dagger(z))_i = \hat{a}_I^\dagger(z, \omega_i). \quad (2.22)$$

Linear superposition of operators propagating from the start of the nonlinearity to its end

can be written as

$$\begin{bmatrix} \hat{\mathbf{a}}_S(z_1) \\ \hat{\mathbf{a}}_I^\dagger(z_1) \end{bmatrix} = \begin{bmatrix} \mathbf{U}_{SS}(z_1, z_0) & \mathbf{U}_{SI}(z_1, z_0) \\ \mathbf{U}_{IS}^*(z_1, z_0) & \mathbf{U}_{II}^*(z_1, z_0) \end{bmatrix} \begin{bmatrix} \hat{\mathbf{a}}_S(z_0) \\ \hat{\mathbf{a}}_I^\dagger(z_0) \end{bmatrix}. \quad (2.23)$$

Then, the relation between the matrix in Eq. (2.23) and the dynamics can be seen as

$$\begin{bmatrix} \mathbf{U}_{SS}(z_1, z_0) & \mathbf{U}_{SI}(z_1, z_0) \\ \mathbf{U}_{IS}^*(z_1, z_0) & \mathbf{U}_{II}^*(z_1, z_0) \end{bmatrix} = \exp \left(\begin{bmatrix} \frac{i\omega}{\Delta v_S} & \frac{i\xi(z)}{\sqrt{2\pi}} \Delta\omega \boldsymbol{\beta} \\ -\frac{i\xi^*(z)}{\sqrt{2\pi}} \Delta\omega \boldsymbol{\beta}^* & -\frac{i\omega}{\Delta v_I} \end{bmatrix} (z_1 - z_0) \right), \quad (2.24)$$

where $\boldsymbol{\omega}$ is a diagonal matrix with the discretized frequency domain on its diagonal and $\boldsymbol{\beta}$ is a matrix such that $\beta_{ij} = \beta(\omega_i + \omega_j)$.

The Schmidt number K can be written in terms of photon number operators as

$$\frac{\langle \int d\omega d\omega' \hat{N}(z, \omega) \hat{N}(z, \omega') \rangle - \langle \int d\omega \hat{N}(z, \omega) \rangle^2}{\langle \int d\omega \hat{N}(z, \omega) \rangle^2} = 1 + \frac{1}{K}. \quad (2.25)$$

The details of the derivation are in Appendix B. The operator $\hat{N}(\omega, \omega')$ can be written as

$$\hat{N}(z, \omega) \equiv \hat{N}_S(z, \omega) = \hat{a}_S^\dagger(z, \omega) \hat{a}_S(z, \omega). \quad (2.26)$$

Note the values of $\langle \int d\omega \hat{N}(z, \omega) \rangle$ and $\langle \int d\omega d\omega' \hat{N}(z, \omega) \hat{N}(z, \omega') \rangle$ would not have changed whether $\hat{N}(z, \omega)$ is defined with the signal or idler beam, as will be shown later. Discretizing this operator in frequencies, Eq. (2.26) can be expressed in terms of annihilation and creation operators at the start of the waveguide

$$\begin{aligned} & \hat{a}_S(z_1, \omega_i)^\dagger \hat{a}_S(z_1, \omega_i) \\ &= \left(\sum_j (\mathbf{U}_{SS}^*(z_1, z_0))_{ij} (\hat{\mathbf{a}}_S^\dagger(z_0))_j + \sum_j (\mathbf{U}_{SI}^*(z_1, z_0))_{ij} (\hat{\mathbf{a}}_I(z_0))_j \right) \\ & \quad \left(\sum_j (\mathbf{U}_{SS}(z_1, z_0))_{ij} (\hat{\mathbf{a}}_S(z_0))_j + \sum_j (\mathbf{U}_{SI}(z_1, z_0))_{ij} (\hat{\mathbf{a}}_I^\dagger(z_0))_j \right). \end{aligned} \quad (2.27)$$

Taking the initial state to be vacuum ($|\text{vac}\rangle$),

$$\left\langle \int d\omega \hat{N}(z_1, \omega) \right\rangle = \langle \text{vac} | \int d\omega \hat{a}_S^\dagger(z_1, \omega) \hat{a}_S(z_1, \omega) | \text{vac} \rangle, \quad (2.28)$$

$$\left\langle \int d\omega d\omega' \hat{N}(z_1, \omega) \hat{N}(z_1, \omega') \right\rangle = \langle \text{vac} | \int d\omega d\omega' \hat{a}_S^\dagger(z_1, \omega') \hat{a}_S(z_1, \omega') \hat{a}_S^\dagger(z_1, \omega) \hat{a}_S(z_1, \omega) | \text{vac} \rangle, \quad (2.29)$$

We can then write

$$\begin{aligned} & \left\langle \int d\omega d\omega' \hat{N}(z, \omega) \hat{N}(z, \omega') \right\rangle \\ &= \left\langle \int d\omega \hat{N}(z, \omega) \right\rangle + \left(\text{Tr} \left(\mathbf{U}_{SI}^*(z_1, z_0) \mathbf{U}_{SI}^T(z_1, z_0) \right) \right)^2 + \text{Tr} \left(\left(\mathbf{U}_{SI}^*(z_1, z_0) \mathbf{U}_{SI}^T(z_1, z_0) \right)^2 \right), \end{aligned} \quad (2.30)$$

and state the objective function as

$$f(\beta(\omega)) = K - 1 = \frac{\left(\text{Tr} \left(\mathbf{U}_{SI}^*(z_1, z_0) \mathbf{U}_{SI}^T(z_1, z_0) \right) \right)^2}{\text{Tr} \left(\left(\mathbf{U}_{SI}^*(z_1, z_0) \mathbf{U}_{SI}^T(z_1, z_0) \right)^2 \right)} - 1, \quad (2.31)$$

as the smallest value of K is theoretically 1. The mean number of photons generated per pulse can be written as

$$\langle n \rangle = \left\langle \int d\omega \hat{N}(z_1, \omega) \right\rangle = \text{Tr} \left(\mathbf{U}_{SI}^*(z_1, z_0) \mathbf{U}_{SI}^T(z_1, z_0) \right) = \text{Tr} \left(\mathbf{U}_{SI}(z_1, z_0) \mathbf{U}_{SI}^\dagger(z_1, z_0) \right). \quad (2.32)$$

If the idler beam was taken instead, the objective function and mean photon number would be

$$f(\beta(\omega)) = \frac{\left(\text{Tr} \left(\mathbf{U}_{IS}^T(z_1, z_0) \mathbf{U}_{IS}^*(z_1, z_0) \right) \right)^2}{\text{Tr} \left(\left(\mathbf{U}_{IS}^T(z_1, z_0) \mathbf{U}_{IS}^*(z_1, z_0) \right)^2 \right)}, \quad (2.33)$$

$$\langle n \rangle = \text{Tr} \left(\mathbf{U}_{IS}^T(z_1, z_0) \mathbf{U}_{IS}^*(z_1, z_0) \right). \quad (2.34)$$

Writing the singular value decomposition [55] of $\mathbf{U}_{SI}(z_1, z_0)$ and $\mathbf{U}_{IS}(z_1, z_0)$ as

$$\mathbf{U}_{SI}(z_1, z_0) = \sum_l \sigma_l v_l^{(S)} w_l^{(I)\dagger}, \quad (2.35)$$

$$\mathbf{U}_{IS}(z_1, z_0) = \sum_l \sigma_l v_l^{(I)} w_l^{(S)\dagger}, \quad (2.36)$$

where $v_l^{(S/I)}$ and $w_l^{(S/I)}$ are vectors representing the singular modes of the propagators, then it would not have mattered whether the signal or the idler beam was chosen for the definition

of $\hat{N}(z, \omega)$ since only the singular values matter. To minimize the objective function while outputting $\langle n \rangle$ photon pairs on average per pulse, the optimization problem becomes

$$\min_{\beta} \frac{\left(\text{Tr} \left(\mathbf{U}_{SI}^*(z_1, z_0) \mathbf{U}_{SI}^T(z_1, z_0) \right) \right)^2}{\text{Tr} \left(\left(\mathbf{U}_{SI}^*(z_1, z_0) \mathbf{U}_{SI}^T(z_1, z_0) \right)^2 \right)} - 1, \quad (2.37)$$

$$\text{s.t. } \text{Tr} \left(\mathbf{U}_{SI}^*(z_1, z_0) \mathbf{U}_{SI}^T(z_1, z_0) \right) - \langle n \rangle = 0, \quad (2.38)$$

where s.t. stands for such that. For simplicity, instead of writing propagators with (z, z_0) for $z \leq z_1$, z_0 will be implicitly zero without loss of generality. For example, $\mathbf{U}_{SS}(z, z_0)$ will be written as $\mathbf{U}_{SS}(z)$ instead.

2.1.3 Relation with perturbative regime

Within the perturbative regime, with a change of reference frame

$$\hat{c}_j(z, \omega) = \exp \left[i\omega z \left(\frac{1}{v_j} - \frac{1}{v_P} \right) \right] \hat{a}_j(z, \omega), \quad (2.39)$$

where $j \in \{S, I\}$, the equations of motion can be written as

$$\partial_z \hat{c}_S(z, \omega) = i\gamma \int d\omega' \beta(\omega + \omega') \hat{c}_I^\dagger(z, \omega') \exp \left[-iz \left(\frac{\omega}{v_S} - \frac{\omega}{v_P} + \frac{\omega'}{v_I} - \frac{\omega'}{v_P} \right) \right], \quad (2.40)$$

$$\partial_z \hat{c}_I^\dagger(z, \omega) = -i\gamma^* \int d\omega' \beta^*(\omega + \omega') \hat{c}_S(z, \omega') \exp \left[iz \left(\frac{\omega}{v_S} - \frac{\omega}{v_P} + \frac{\omega'}{v_I} - \frac{\omega'}{v_P} \right) \right], \quad (2.41)$$

where $\gamma = \frac{\xi(z)}{\sqrt{2\pi}}$ if assuming the nonlinear susceptibility does not change along the waveguide. Approximating

$$\hat{c}_I^\dagger(\ell, \omega) \approx \hat{c}_I^\dagger(0, \omega), \quad (2.42)$$

where ℓ is the length of the waveguide, the operator associated with the signal beam is

$$\hat{c}_S(\ell, \omega) \approx \hat{c}_S(0, \omega) + i\gamma \int_0^\ell dz' \int d\omega' \beta(\omega + \omega') \hat{c}_I^\dagger(0, \omega') \exp \left[-iz' \left(\frac{\omega}{v_S} - \frac{\omega}{v_P} + \frac{\omega'}{v_I} - \frac{\omega'}{v_P} \right) \right] \quad (2.43)$$

$$\begin{aligned} &= \hat{c}_S(0, \omega) + i\gamma \frac{\ell}{2} \int d\omega' \beta(\omega + \omega') \text{sinc} \left(\frac{\ell}{2} \left(\frac{\omega}{v_S} - \frac{\omega}{v_P} + \frac{\omega'}{v_I} - \frac{\omega'}{v_P} \right) \right) \\ &\quad \cdot \exp \left[i\frac{\ell}{2} \left(\frac{\omega}{v_S} - \frac{\omega}{v_P} + \frac{\omega'}{v_I} - \frac{\omega'}{v_P} \right) \right] \hat{c}_I^\dagger(0, \omega'). \end{aligned} \quad (2.44)$$

In the symmetric group velocity matching regime ($\frac{1}{v_S} - \frac{1}{v_P} = \frac{1}{v_P} - \frac{1}{v_I}$), the second equation can be written as

$$\hat{c}_S(\ell, \omega) \approx \hat{c}_S(0, \omega) + i\gamma \frac{\ell}{2} \int d\omega' \beta(\omega + \omega') \text{sinc} \left(\frac{\ell}{2} \left(\frac{\omega - \omega'}{\Delta v} \right) \right) \exp \left[i \frac{\ell}{2} \left(\frac{\omega - \omega'}{\Delta v} \right) \right] \hat{c}_I^\dagger(0, \omega'), \quad (2.45)$$

where $\frac{1}{\Delta v} = \frac{1}{v_S} - \frac{1}{v_P}$. When projecting onto discretized frequency space, the JSA can be approximated by the moment $\mathbf{U}_{SS}^\dagger(z_1, z_0) \mathbf{U}_{SI}(z_1, z_0)$ since the block relating $\hat{c}_S(0, \omega)$ and $\hat{c}_S(l, \omega')$ is the identity, the JSA can be written as

$$JSA(\omega, \omega') \approx i\gamma \frac{\ell}{2} \beta(\omega + \omega') \text{sinc} \left(\frac{\ell}{2} \left(\frac{\omega - \omega'}{\Delta v} \right) \right) \exp \left[i \frac{\ell}{2} \left(\frac{\omega - \omega'}{\Delta v} \right) \right], \quad (2.46)$$

which is the same as the JSA in perturbative regime as shown in Eq. (1.13) save for a phase.

In the low-gain regime, the probability density of generating signal and idler photons can be given by $|JSA(\omega, \omega')|^2$ [56] such that the mean number of photon pairs generated per pump pulse is $\langle n \rangle = \int d\omega d\omega' |JSA(\omega, \omega')|^2$. Then, with the values given by Table 2.1 and a Gaussian pump described by $\beta(\omega + \omega') = \sqrt{\frac{E_p \tau}{\sqrt{2\pi}}} \exp \left[-(\omega + \omega')^2 \tau^2 \right]$ for some fixed width τ , we have

$$\begin{aligned} \langle n \rangle &= \frac{|\xi(z)|^2}{2\pi} E_p \tau \sqrt{\frac{2}{\pi}} \frac{\ell^2}{4} \int d\omega d\omega' \exp \left[-2(\omega + \omega')^2 \tau^2 \right] \text{sinc} \left[\frac{\ell}{2} \frac{(\omega - \omega')}{\Delta v} \right]^2 \\ &= \frac{|\xi(z)|^2}{2\pi} E_p \tau \sqrt{\frac{2}{\pi}} \frac{\ell^2}{4} \frac{1}{2} \int du dq \exp \left[-2u^2 \tau^2 \right] \text{sinc} \left[\frac{\ell}{2\Delta v} q \right]^2 \\ &= \frac{|\xi(z)|^2 E_p \ell \Delta v}{8} \\ &\approx 5.76 \times 10^9 (E_p) / \text{Joule}. \end{aligned} \quad (2.47)$$

Assuming E_p (the energy of the pump) is on the order of picojoules, then if $E_p = 10^{-12}$ J, $\langle n \rangle \approx 5.76 \times 10^{-3}$. For the calculations, $\frac{1}{\Delta v}$ is approximated by $\frac{1}{\Delta v} = \frac{1}{2} \left(\frac{1}{\Delta v_S} + \frac{1}{\Delta v_I} \right)$.

2.2 Objective Function With Decoupled Equations of Motion

2.2.1 Decoupling the equations of motion

While Sec. 2.1 gives a description of physics in high-gain regime for SPDC in waveguide, the equations of motion are coupled. For a numerically easier representation, it is worth to

decouple them. However, the optimization problem from coupled dynamics will be solved to show the considerations for the equations to be decoupled are not too restrictive. Under symmetric group velocity matching, the spectral distribution of the signal and idler beams is identical [23]. Paired with the assumption that the pump is Fourier-limited with a global constant phase that can be neglected, the equations

$$\frac{1}{v_S} - \frac{1}{v_P} = \frac{1}{v_P} - \frac{1}{v_I} = \frac{1}{\Delta v}, \quad (2.48a)$$

$$\beta(\omega) = \beta^*(\omega) \quad (2.48b)$$

will hold. Then, the SPDC equations of motion can be decoupled. Defining the pseudo-quadratures as

$$\hat{X}_+(z, \omega) = \frac{1}{\sqrt{2}} \left(\hat{a}_S(z, \omega) + i \frac{\gamma}{|\gamma|} \hat{a}_I^\dagger(z, \omega) \right) = \frac{1}{\sqrt{2}} \left(\hat{a}_S(z, \omega) + \hat{a}_I^\dagger(z, \omega) \right), \quad (2.49)$$

$$\hat{X}_-(z, \omega) = \frac{1}{\sqrt{2}} \left(\hat{a}_S(z, \omega) - i \frac{\gamma}{|\gamma|} \hat{a}_I^\dagger(z, \omega) \right) = \frac{1}{\sqrt{2}} \left(\hat{a}_S(z, \omega) - \hat{a}_I^\dagger(z, \omega) \right), \quad (2.50)$$

if $\gamma = -i|\gamma|$. Then, the decoupled equations of motion are written as

$$\partial_z \hat{X}_+(z, \omega) = \frac{i\omega}{\Delta v} \hat{X}_+(z, \omega) + |\gamma| \int \beta(\omega + \omega') \hat{X}_+(z, \omega') d\omega', \quad (2.51)$$

$$\partial_z \hat{X}_-(z, \omega) = \frac{i\omega}{\Delta v} \hat{X}_-(z, \omega) - |\gamma| \int \beta(\omega + \omega') \hat{X}_-(z, \omega') d\omega'. \quad (2.52)$$

2.2.2 Making the equations of motion dimensionless

Changes of variables can make the problem unitless. Assume ω and z are in the interval

$$-\omega_f \leq \omega \leq \omega_f, \quad (2.53)$$

$$0 \leq z \leq \ell, \quad (2.54)$$

the dynamics can be rewritten in terms of rescaled variables that would be defined as

$$0 \leq \zeta \leq 1, \quad (2.55)$$

$$-\frac{\omega_f \ell}{\Delta v} \leq \Omega \leq \frac{\omega_f \ell}{\Delta v}, \quad (2.56)$$

where $\zeta = \frac{z}{\ell}$ and $\Omega = \omega \frac{\ell}{\Delta v}$. The equations of motion can be written as

$$\begin{aligned} \partial_{\ell\zeta} \hat{X}_{\pm}(\ell\zeta, \frac{\Delta v}{\ell}\Omega) &= \frac{i\Delta v\Omega}{\ell\Delta v} \hat{X}_{\pm}(\ell\zeta, \frac{\Delta v}{\ell}\Omega) \\ &\pm |\gamma| \int \frac{\Delta v}{\ell} d\Omega' \beta \left(\frac{\Delta v}{\ell} (\Omega + \Omega') \right) \hat{X}_{\pm}(\ell\zeta, \frac{\Delta v}{\ell}\Omega'). \end{aligned} \quad (2.57)$$

The equations can be simplified as

$$\begin{aligned} \partial_{\zeta} \hat{X}_{\pm}(\ell\zeta, \frac{\Delta v}{\ell}\Omega) &= i\Omega \hat{X}_{\pm}(\ell\zeta, \frac{\Delta v}{\ell}\Omega) \\ &\pm |\gamma| \int \Delta v d\Omega' \beta \left(\frac{\Delta v}{\ell} (\Omega + \Omega') \right) \hat{X}_{\pm}(\ell\zeta, \frac{\Delta v}{\ell}\Omega'). \end{aligned} \quad (2.58)$$

Replacing the variables ω and z by Ω and ζ , the pseudo-quadratures become

$$\hat{X}_{\pm}(z, \omega) \rightarrow \hat{X}_{\pm}(\zeta, \Omega), \quad (2.59)$$

since $\hat{X}_{\pm}(\zeta, \Omega) = \hat{X}_{\pm}(f(z), g(\omega))$, the equations of motion become

$$\begin{aligned} \partial_{\zeta} \hat{X}_{\pm}(\zeta, \Omega) &= i\Omega \hat{X}_{\pm}(\zeta, \Omega) \\ &\pm |\gamma| \int \Delta v d\Omega' \beta \left(\frac{\Delta v}{l} (\Omega + \Omega') \right) \hat{X}_{\pm}(\zeta, \Omega'). \end{aligned} \quad (2.60)$$

The units of $|\gamma|$ can be given by dimensional analysis of the original equations of dynamics. Let $[\cdot]$ signify the dimension of $\cdot$,

$$[|\gamma|] [|\beta|] = [z]^{-1} [\omega]^{-1} = [\Delta v]^{-1}. \quad (2.61)$$

Then, a change of variable

$$|\gamma| \Delta v \beta \left(\frac{\Delta v}{\ell} (\omega + \omega') \right) \rightarrow \beta (\Omega + \Omega'), \quad (2.62)$$

makes $\beta(\Omega + \Omega')$, ζ and Ω all unitless. In Eq. (2.61), the amplitude of γ , Δv and that of the original β are all absorbed into the amplitude of the new β , allowing the equations of motion to be written as

$$\begin{aligned} \partial_{\zeta} \hat{X}_{\pm}(\zeta, \Omega) &= i\Omega \hat{X}_{\pm}(\zeta, \Omega) \\ &\pm \int d\Omega' \beta (\Omega + \Omega') \hat{X}_{\pm}(\zeta, \Omega'). \end{aligned} \quad (2.63)$$

Through a similar argument, it is also possible to remove the units from the equations of motion for the annihilation and creation operators. The pseudo-quadratures can then be written as

$$\hat{X}_+(\zeta, \Omega) = \frac{1}{\sqrt{2}} \left(\hat{a}_S(\zeta, \Omega) + \hat{a}_I^\dagger(\zeta, \Omega) \right), \quad (2.64a)$$

$$\hat{X}_-(\zeta, \Omega) = \frac{1}{\sqrt{2}} \left(\hat{a}_S(\zeta, \Omega) - \hat{a}_I^\dagger(\zeta, \Omega) \right), \quad (2.64b)$$

2.2.3 Propagators of pseudo-quadratures

With the definition of pseudo-quadratures given by Eq. (2.64), the annihilation and creation operators are given by

$$\hat{a}_S(\zeta, \Omega) = \frac{1}{\sqrt{2}} (\hat{X}_+(\zeta, \Omega) + \hat{X}_-(\zeta, \Omega)), \quad (2.65)$$

$$\hat{a}_I^\dagger(\zeta, \Omega) = \frac{1}{\sqrt{2}} (\hat{X}_+(\zeta, \Omega) - \hat{X}_-(\zeta, \Omega)). \quad (2.66)$$

By calling

$$\hat{X}_+(\zeta, \Omega) = \int U_+(\zeta, \Omega, \Omega') \hat{X}_+(0, \Omega') d\Omega', \quad (2.67)$$

$$\hat{X}_-(\zeta, \Omega) = \int U_-(\zeta, \Omega, \Omega') \hat{X}_-(0, \Omega') d\Omega', \quad (2.68)$$

we can show that

$$\begin{aligned} & \frac{1}{\sqrt{2}} \left(\hat{a}_S(\zeta, \Omega) + \hat{a}_I^\dagger(\zeta, \Omega) \right) \\ &= \frac{1}{\sqrt{2}} \int U_+(\zeta, \Omega, \Omega') \left[\hat{a}_S(0, \Omega') + \hat{a}_I^\dagger(0, \Omega') \right] d\Omega' \\ &= \frac{1}{\sqrt{2}} \int [U_{SS}(\zeta, \Omega, \Omega') + U_{IS}^*(\zeta, \Omega, \Omega')] \hat{a}_s(0, \Omega') d\Omega' \\ &+ \frac{1}{\sqrt{2}} \int [U_{SI}(\zeta, \Omega, \Omega') + U_{II}^*(\zeta, \Omega, \Omega')] \hat{a}_I^\dagger(0, \Omega') d\Omega', \end{aligned} \quad (2.69)$$

for $\hat{X}_+(\zeta, \Omega)$ and

$$\begin{aligned}
& \frac{1}{\sqrt{2}} \left(\hat{a}_S(\zeta, \Omega) - \hat{a}_I^\dagger(\zeta, \Omega) \right) \\
&= \frac{1}{\sqrt{2}} \int U_-(\zeta, \Omega, \Omega') \left[\hat{a}_S(0, \Omega') - \hat{a}_I^\dagger(0, \Omega') \right] d\Omega' \\
&= \frac{1}{\sqrt{2}} \int [U_{SS}(\zeta, \Omega, \Omega') - U_{IS}^*(\zeta, \Omega, \Omega')] \hat{a}_S(0, \Omega') d\Omega' \\
&\quad - \frac{1}{\sqrt{2}} \int [U_{II}^*(\zeta, \Omega, \Omega') - U_{SI}(\zeta, \Omega, \Omega')] \hat{a}_I^\dagger(0, \Omega') d\Omega', \tag{2.70}
\end{aligned}$$

for $\hat{X}_-(\zeta, \Omega)$. From these equations,

$$\begin{aligned}
U_+(\zeta, \Omega, \Omega') &= U_{SS}(\zeta, \Omega, \Omega') + U_{IS}^*(\zeta, \Omega, \Omega') \\
&= U_{II}^*(\zeta, \Omega, \Omega') + U_{SI}(\zeta, \Omega, \Omega'), \tag{2.71}
\end{aligned}$$

$$\begin{aligned}
U_-(\zeta, \Omega, \Omega') &= U_{SS}(\zeta, \Omega, \Omega') - U_{IS}^*(\zeta, \Omega, \Omega') \\
&= U_{II}^*(\zeta, \Omega, \Omega') - U_{SI}(\zeta, \Omega, \Omega'). \tag{2.72}
\end{aligned}$$

In matrix representation, using the same method as in Sec. 2.1.2, the propagators relations become

$$\mathbf{U}_+(\zeta) = \mathbf{U}_{SS}(\zeta) + \mathbf{U}_{IS}^*(\zeta) = \mathbf{U}_{II}^*(\zeta) + \mathbf{U}_{SI}(\zeta), \tag{2.73}$$

$$\mathbf{U}_-(\zeta) = \mathbf{U}_{SS}(\zeta) - \mathbf{U}_{IS}^*(\zeta) = \mathbf{U}_{II}^*(\zeta) - \mathbf{U}_{SI}(\zeta). \tag{2.74}$$

The relation between the blocks of the propagator of annihilation and creation operators and of the newly defined propagators is

$$\mathbf{U}_{SS}(\zeta) = \mathbf{U}_{II}^*(\zeta) = \frac{1}{2} [\mathbf{U}_+(\zeta) + \mathbf{U}_-(\zeta)], \tag{2.75a}$$

$$\mathbf{U}_{IS}^*(\zeta) = \mathbf{U}_{SI}(\zeta) = \frac{1}{2} [\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta)]. \tag{2.75b}$$

With the same logic as in Sec. 2.1.2, the solution of the equations of motion is given by

$$\mathbf{U}_+(\zeta) = \exp [(\mathbf{i}\mathbf{\Omega} + \Delta\mathbf{\Omega}\mathbf{\beta}) \zeta], \tag{2.76}$$

$$\mathbf{U}_-(\zeta) = \exp [(\mathbf{i}\mathbf{\Omega} - \Delta\mathbf{\Omega}\mathbf{\beta}) \zeta], \tag{2.77}$$

where this time $\mathbf{\Omega}$ is a diagonal matrix with elements that of the discretized domain of Ω and $\mathbf{\beta}$ a matrix such that $\beta_{ij} = \beta(\Omega_i + \Omega_j)$ in the new unitless domain.

2.2.4 Commutation between pseudo-quadratures and relations between decoupled propagators

Using the canonical commutation relations

$$[\hat{a}_j(\zeta, \Omega), \hat{a}_k^\dagger(\zeta, \Omega)] = \delta_{jk} \delta(\Omega - \Omega'), \quad (2.78)$$

$$[\hat{a}_j(\zeta, \Omega), \hat{a}_k(\zeta, \Omega)] = 0, \quad (2.79)$$

with δ_{jk} being the Kronecker's Delta, $j, k \in \{S, I\}$ and $\delta(\Omega - \Omega')$ being Dirac's delta, the commutation relations between the newly defined operators will follow specific commutation relations set by Eq. (2.78) and (2.79) such as

$$[\hat{X}_+(\zeta, \Omega_i), \hat{X}_+(\zeta, \Omega_j)] = 0, \quad (2.80a)$$

$$[\hat{X}_+(\zeta, \Omega_i), \hat{X}_+^\dagger(\zeta, \Omega_j)] = 0, \quad (2.80b)$$

$$[\hat{X}_+(\zeta, \Omega_i), \hat{X}_-(\zeta, \Omega_j)] = 0, \quad (2.80c)$$

$$[\hat{X}_+(\zeta, \Omega_i), \hat{X}_-^\dagger(\zeta, \Omega_j)] = \delta_{ij}, \quad (2.80d)$$

$$[\hat{X}_-(\zeta, \Omega_i), \hat{X}_-^\dagger(\zeta, \Omega_j)] = 0, \quad (2.80e)$$

$$[\hat{X}_-(\zeta, \Omega_i), \hat{X}_-(\zeta, \Omega_j)] = 0. \quad (2.80f)$$

Details of calculations are shown in Appendix C. Because of Eqns. (2.80) and more particularly Eq. (2.80e), the relation

$$\mathbf{U}_+(\zeta) \mathbf{U}_-^\dagger(\zeta) = \mathbf{I} \quad (2.81)$$

holds true. When looking at the propagators as matrix exponentials, it can be seen that

$$\mathbf{U}_+(\zeta) \mathbf{U}_-^\dagger(\zeta) = \exp [(\mathbf{K} + \mathbf{V}) \zeta] \exp [-(\mathbf{K} + \mathbf{V}) \zeta] = \mathbf{I}, \quad (2.82)$$

where $\mathbf{K} = i\Omega$ and $\mathbf{V} = \Delta\Omega\beta$. It can be seen further that

$$\mathbf{U}_-(\zeta)\mathbf{U}_+^\dagger(\zeta) = \exp[(\mathbf{K} - \mathbf{V})\zeta] \exp[-(\mathbf{K} - \mathbf{V})\zeta] = \mathbf{I}, \quad (2.83a)$$

$$\begin{aligned} \mathbf{U}_+(\zeta)\mathbf{U}_-^\dagger(\zeta') &= \exp[(\mathbf{K} + \mathbf{V})\zeta] \exp[-(\mathbf{K} + \mathbf{V})\zeta'] \\ &= \exp[(\mathbf{K} + \mathbf{V})(\zeta - \zeta')] = \mathbf{U}_+(\zeta - \zeta'), \end{aligned} \quad (2.83b)$$

$$\begin{aligned} \mathbf{U}_-(\zeta)\mathbf{U}_+^\dagger(\zeta') &= \exp[(\mathbf{K} - \mathbf{V})\zeta] \exp[-(\mathbf{K} - \mathbf{V})\zeta'] \\ &= \exp[(\mathbf{K} - \mathbf{V})(\zeta - \zeta')] = \mathbf{U}_-(\zeta - \zeta'), \end{aligned} \quad (2.83c)$$

$$[\mathbf{U}_\pm(\zeta), \mathbf{U}_\mp^\dagger(\zeta)] = \mathbf{0}. \quad (2.83d)$$

The relation $\mathbf{U}_-(\zeta) = \mathbf{U}_+^{-1}(\zeta)$ can then be established, thus

$$\mathbf{U}_+(\zeta) = \mathbf{S}\mathbf{D}_+\mathbf{W}^\dagger, \quad (2.84)$$

$$\mathbf{U}_-(\zeta) = \mathbf{S}\mathbf{D}_-^{-1}\mathbf{W}^\dagger, \quad (2.85)$$

where the singular modes encoded in matrix form ($\mathbf{S}$ and $\mathbf{W}$) of both propagators are the same. The singular values are also inverse of each other, as $\mathbf{D}_+ = \mathbf{D}_-^{-1}$.

2.2.5 Equivalent formulation of Schmidt number

A reformulation of the Schmidt number from the blocks of the propagator of the annihilation and creation operators in Eq. (2.31) in the formalism of propagators of the pseudo-quadratures gives

$$\begin{aligned} \min_{\beta} \frac{\text{Tr} [\mathbf{U}_{SI}^*(\zeta)\mathbf{U}_{SI}^T(\zeta)]^2}{\text{Tr} [(\mathbf{U}_{SI}^*(\zeta)\mathbf{U}_{SI}^T(\zeta))^2]} - 1 &= \min_{\beta} \frac{\text{Tr} [\mathbf{U}_{SI}(\zeta)\mathbf{U}_{SI}^\dagger(\zeta)]^2}{\text{Tr} [(\mathbf{U}_{SI}(\zeta)\mathbf{U}_{SI}^\dagger(\zeta))^2]} - 1 \\ &= \min_{\beta} \frac{\text{Tr} [(\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta))(\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta))^\dagger]^2}{\text{Tr} [((\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta))(\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta))^\dagger)^2]} - 1 \\ &\equiv \min_{\beta} \frac{\text{Tr} [(\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta))(\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta))^\dagger]^2}{\text{Tr} [((\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta))(\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta))^\dagger)^2]}. \end{aligned} \quad (2.86)$$

Minimization of Schmidt number can be seen as being an alternative to minimization of the rank of $\mathbf{U}_{+N} - \mathbf{U}_{-N}$. An equivalent objective function would be

$$\begin{aligned} \min_{\beta} \quad & \frac{|\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta)|_{\text{Nuc}}}{2\sqrt{\langle n \rangle}}, \\ \text{s.t.} \quad & \frac{1}{4} \text{Tr} [(\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta))(\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta))^\dagger] - \langle n \rangle = 0, \end{aligned} \quad (2.87)$$

where $|\cdot|_{\text{Nuc}}$ is the nuclear norm defined as $|\mathbf{A}|_{\text{Nuc}} = \text{Tr} [\sqrt{\mathbf{A}\mathbf{A}^\dagger}]$. The advantage of this formulation of objective function is that it is a norm multiplied by a constant. In fact, the nuclear norm of a matrix is a convex envelope for its rank [57]. However, this norm is akin to adding together the absolute values of the square root of the eigenvalues of $\mathbf{A}\mathbf{A}^\dagger$, meaning it is not differentiable everywhere. For an optimization scheme that requires the gradient of the objective function, the Schmidt number formulation will still be used. The nuclear norm can be used however in the context of convex optimization, as will be explained later on. The equivalence between the nuclear norm and the Schmidt number mainly comes from the argmin and argmax (the values of $\mathbf{U}_+(\zeta)$ and $\mathbf{U}_-(\zeta)$ at minimum and maximum, respectively), as well as the minimum and maximum value of problem in Eq. (2.87) being the same as that of Eq. (2.86). Details are shown in Appendix E. The nuclear norm objective function will be used in the context of convex optimization and the Schmidt number for numerical optimization. The optimization methods used in both numerical and convex optimization are all iterative, with the index k representing the k th iteration of the scheme.

2.3 Numerical Optimization

Because of the smoothness of the matrix exponential, an intuitive search for the minimum of the constrained problem would be with a gradient-based method. A presentation of the theory behind the numerical optimization method used to solve the constrained problem will be made here.

2.3.1 Quasi-Newton method

Steepest descent is a popular gradient-based method to minimize an objective function, partly due to its simplicity and to the mild assumptions needed on the function to minimize. Gradient-based methods are often iterative. The index k in this section represents the k th iteration unless specified otherwise. The intuition behind the algorithm is as follows: it is assumed that an objective function mapping from real vector space ($\mathbb{R}^n$) to the real scalar

space can be approximated by

$$f(x_{k+1}) - f(x_k) \approx \nabla f(x_k)^T (x_{k+1} - x_k) + \mathcal{O}\left((x_{k+1} - x_k)^2\right), \quad (2.88)$$

where $x_{k+1}, x_k \in \mathbb{R}^n$ are two points on the domain and $\mathcal{O}\left((x_{k+1} - x_k)^2\right)$ is the second-order error of the Taylor approximation. Assuming

$$\left| \nabla f(x_k)^T (x_{k+1} - x_k) \right| \gg \mathcal{O}\left((x_{k+1} - x_k)^2\right), \quad (2.89)$$

let $x_{k+1} = x_k - \alpha_k d_k$, if $\alpha_k \geq 0$ and $\nabla f(x_k)^T d_k \leq 0$, then $\nabla f(x_k)^T (x_{k+1} - x_k) \leq 0$ so that $f(x_{k+1}) - f(x_k) \leq 0$ by Eq. (2.88). The case where $d_k = \nabla f(x_k)$ gives the direction where the first order term in the Taylor expansion has the lowest value, hence why this d_k is called the steepest descent direction. A condition for minimum that is often a stopping criteria for algorithms is $f(x_{k+1}) - f(x_k) = \epsilon \rightarrow 0$. The process of finding a good α_k is often called a line search. A good line search can help $f(x_{k+1}) - f(x_k)$ converge faster to 0 [58]. For the rate of convergence of steepest descent using the exact line search with a quadratic objective function, it is known that it at least converges linearly [59]. An exact line search is the problem of finding the value of α_k that solves $\min_{\alpha_k} f(x_k - \alpha_k d_k)$.

Issues with the steepest descent algorithm include how it may converge very slowly, or even worse, it may not converge to a minimum. A saddle point is an example of such a case. At such a point, the gradient is 0, and as such, the assumption in Eq. (2.89) does not hold true any longer. Higher order elements of the Taylor expansion need then to be taken into account, such as using the Hessian for the second-order term. Benefits of integrating the Hessian of the objective function into the search algorithm include possibly obtaining a better rate of convergence of the program. Taking the second-order error to be

$$\mathcal{O}\left((x_{k+1} - x_k)^2\right) = \frac{1}{2} (x_{k+1} - x_k)^T \nabla^2 f(x_k) (x_{k+1} - x_k) + \mathcal{O}\left((x_{k+1} - x_k)^3\right), \quad (2.90)$$

the Taylor approximation can be given as

$$\begin{aligned} f(x_{k+1}) - f(x_k) \\ \approx \nabla f(x_k)^T (x_{k+1} - x_k) + \frac{1}{2} (x_{k+1} - x_k)^T \nabla^2 f(x_k) (x_{k+1} - x_k) + \mathcal{O}\left((x_{k+1} - x_k)^3\right). \end{aligned} \quad (2.91)$$

For a descent direction such as

$$x_{k+1} = x_k - \left(\nabla^2 f(x_k) \right)^{-1} \nabla f(x_k), \quad (2.92)$$

the Taylor approximation becomes

$$f(x_{k+1} - x_k) - f(x_k) \approx -\frac{1}{2} \nabla f(x_k)^T \left(\nabla^2 f(x_k) \right)^{-1} \nabla f(x_k) + \mathcal{O} \left((x_{k+1} - x_k)^3 \right). \quad (2.93)$$

The matrix $\nabla^2 f(x_k)$ is the Hessian of the objective function evaluated at x_k . The step defined in Eq. (2.92) gives rise to Newton's method in optimization [60]. A key component for convergence using this method is the positive definiteness of the Hessian. For a more complicated objective function, where the curvature can change, the inverse of the Hessian may not be well-defined, or can even be negative definite.

A family of optimization schemes closely related to Newton's method is that of quasi-Newton methods. Instead of taking the inverse of the Hessian, the inverse of a specifically chosen positive definite matrix B_k is taken. Taking the direction $d_k = -B_k^{-1} \nabla f(x_k)$, after performing a line search to find the step $x_{k+1} = x_k + \alpha_k d_k$, the terms $s_k = \alpha_k d_k$ and $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ are defined. To find the term B_{k+1} , the following relation is used [61]

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k^T}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k}. \quad (2.94)$$

Then, one would repeat the process, obtaining $d_{k+1} = -B_{k+1}^{-1} \nabla f(x_{k+1})$ and $x_{k+2} = x_{k+1} + \alpha_{k+1} d_{k+1}$ until $\|y_n\| \rightarrow 0$, or the norm of y_n tends to 0.

2.3.2 Penalty method and regularization

For enforcing equality constraints, such as a fixed number of photon pairs, a penalty method can be used. While it can also be used for inequalities, a discontinuity would have to be included in the formulation [62], so regularization is used instead. The main idea behind both methods is to turn a constrained problem into an unconstrained one.

For equality constraints, such as a fixed number of idler and signal photon pairs per pump pulse, let a constraint $f_1(x) = 0$ and an objective function $f_0(x)$ such that the problem is $\min_x \{f_0(x) | f_1(x) = 0\}$, a penalized objective function $f_\rho^{(k)}(x) = f_0(x) + \frac{1}{2} \rho^{(k)} (f_1(x))^2$, where $\rho^{(k)}$ is a positive real value, can be used. The superscript k keeps track of the iterations of the penalty method so that $\rho^{(k)} < \rho^{(k+1)}$. With k starting at 0, after w iterations, if w is big enough, then $\rho^{(w)} \gg \rho^{(0)}$. Assume $\tilde{x}$ is within the feasible set such that $f_1(\tilde{x}) = 0$, minimizing $f_\rho^{(k)}(\tilde{x})$ is the same as minimizing $f_0(\tilde{x})$. It has been shown that if the gradients of the constraints are linearly independent, then when $\rho^{(k)} \rightarrow \infty$, the local minimum of $f_\rho^{(k)}(x)$

tends to the minimum of the constrained problem [63].

Regularization is a process that turns an objective function under inequality constraints into an unconstrained objective function with a fixed penalty. A justification on the validity of the method can be made with an argument using a mathematical optimization framework. Let a problem be

$$\begin{aligned} \min_x \quad & f(x), \\ \text{s.t.} \quad & g(x) \leq c, \end{aligned} \tag{2.95}$$

with a non-negative $g(x)$, the regularized objective function can be written as

$$\min_x \quad f_\lambda(x) = f(x) + \lambda g(x). \tag{2.96}$$

The function $g(x)$ is often taken as the absolute value or the Euclidean norm of x [64]. Solving the problem of minimizing $f_\lambda(x)$ is equivalent to solving the problem in Eq. (2.95) where the value of c will depend on the value of λ . One way of seeing why this formulation is equivalent is by defining

$$L(\lambda, x) = f(x) + \lambda (g(x) - c), \tag{2.97}$$

where $L(\lambda, x)$ is the Lagrangian of the problem in Eq. (2.95). The Karush–Kuhn–Tucker (KKT) conditions [65] state that necessary conditions for optimality for constrained problems are

$$\nabla f(x^*) + \lambda \nabla g(x^*) = 0, \tag{2.98a}$$

$$\lambda g(x^*) = 0, \tag{2.98b}$$

$$\lambda \geq 0, \tag{2.98c}$$

$$g(x^*) - c \leq 0, \tag{2.98d}$$

where x^* is the argmin to the constrained problem. For a non-negative λ , the gradient of the regularized objective function is $\nabla f_\lambda(x) = \nabla f(x) + \lambda \nabla g(x)$. At optimality, this gradient is close to zero. If $g(x^*) = c$ holds true, then solving the problem of Eq. (2.95) is equivalent to solving the problem in Eq. (2.97).

2.4 Convex Optimization and Convex Relaxations

The complex matrix exponentiation map is highly non-convex. Therefore, there may exist many local minima [66]. A relaxation of the original problem is then needed to bound the global minimum of the constrained problem. The difference between a local minimum and a global minimum is that a local minimum is the smallest value within a certain interval of domain, while a global minimum is the smallest value over all of the domain of definition. The next section is dedicated to an overview of methods of convex optimization used to obtain the results.

2.4.1 Convex problem

A convex problem is of form

$$\begin{aligned} \max_x / \min_x \quad & f_0(x), \\ \text{s.t.} \quad & f_i(x) \leq 0, \\ & f_l(x) = 0, \end{aligned} \tag{2.99}$$

for all $i \in \mathfrak{I}$, $l \in \mathfrak{L}$ where $\mathfrak{I} \cup \mathfrak{L}$ is the interval of positive integers from 1 up to n such that there are n constraints to the problem. Any x that satisfies the constraints is said to be a feasible point and the set that contains all feasible points is sometimes called the feasible region. The functions $f_0(x)$ and $f_i(x)$ can be nonlinear functions where $f : \Omega \mapsto \mathbb{R}$, with Ω being the space of complex or real matrices or vectors and $f_l(x)$ are normally linear.

For the problem in Eq. (2.99) to be convex, if it is one of minimization, the function $f_0(x)$ needs to be convex, which is to say [66]

$$f_0(\alpha x_1 + (1 - \alpha)x_2) - (\alpha f_0(x_1) + (1 - \alpha)f_0(x_2)) \leq 0 \tag{2.100}$$

is true for any $0 < \alpha < 1$ and for any feasible x_1, x_2 . If minimization is preferred, $f_0(x)$ needs to be convex, which is the same definition as in Eq. (2.100) but with $\leq$ instead of $\geq$. Another requirement for convexity is that the constraints $f_i(x) \geq 0$ and $f_l(x) = 0$ need to generate a convex set no matter if the problem is a minimization or maximization, meaning the convex feasible region is a set that includes any element y where

$$y = \alpha x_1 + (1 - \alpha)x_2, \tag{2.101}$$

if x_1 and x_2 are feasible and $0 > \alpha > 1$. An interpretation of the definition is that any point lying on the line between two points within the set is also an element of said set.

Convex problems have the property that any local optima are also global optima. In this sense, convex problems are “easy” to solve [66], as any searching algorithm, if it does converge, will lead to a global optimum. For example, Fig. 2.2 shows the application of convex function criteria from Eq. (2.100) on the objective function defined in Eq. (2.87) with x_1 and x_2 given by the first two terms of the Hermite basis. If the mean number of photon pairs per pulse is given by $\langle n \rangle = 1$ for both x_1 and x_2 , then the criteria in Eq. (2.101) can be evaluated for both pumps as shown in Fig. 2.3. For some range of α , the criteria are not respected, then the objective function is non-convex and the feasible set is a non-convex set, making the problem a non-convex problem. A local minimum of a non-convex problem is not necessarily a global one, but their global optimal value can be bounded by solving their Lagrangian dual, which is a convex problem whose definition will be given below.

2.4.2 Lagrangian dual of a problem

The relaxation of a problem is another easier-to-solve problem where solving it will yield a result that bounds the solution of the original problem. If the bounding is a convex problem, it is called a convex relaxation, meaning its feasible set is convex and the objective function of the relaxation is also convex and is always inferior (superior) or equal to the original objective function for a minimization (maximization) problem. The Lagrangian dual is such a relaxation.

Rewriting the problem in Eq. (2.99) as a general possibly non-convex minimization problem with an objective function $f_0(X) : \mathbb{C}^{n \times m} \rightarrow \mathbb{R}$, $f_l(X) : \mathbb{C}^{n \times m} \rightarrow \mathbb{R}$ and $f_i(X) : \mathbb{C}^{n \times m} \rightarrow \mathbb{R}$, such that the optimization problem is in the form of

$$\begin{aligned} \min_X \quad & f_0(X), \\ \text{s.t.} \quad & f_i(X) \leq 0, \\ & f_l(X) = 0, \end{aligned} \tag{2.102}$$

The Lagrangian of the problem can be written as

$$L(X, \lambda, \nu) = f_0(X) + \sum_{i \in \mathcal{I}} \lambda_i f_i(X) + \sum_{l \in \mathcal{L}} \nu_l f_l(X), \tag{2.103}$$

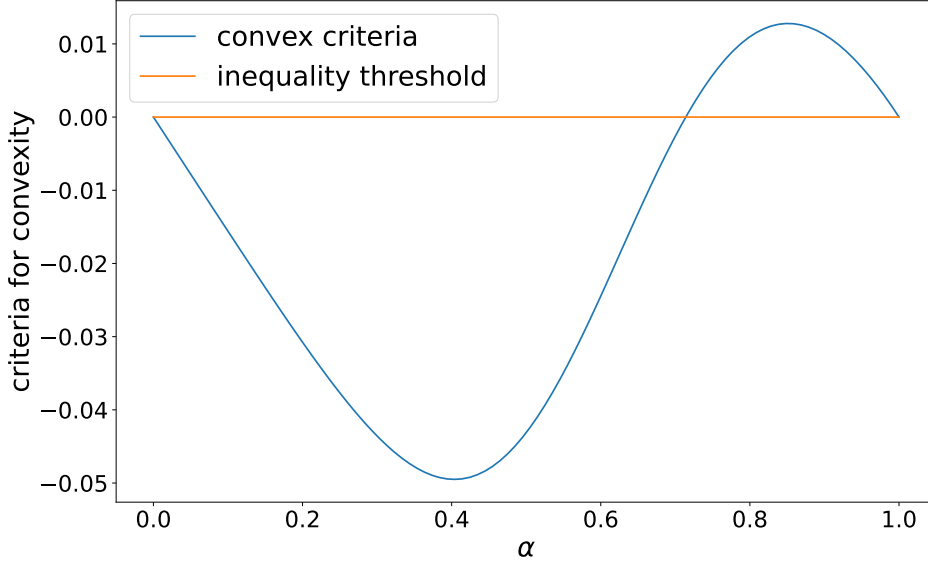


Figure 2.2 Criteria for convexity for the Schmidt number function. For some range of α , the criteria is respected (≤ 0), but not for all $\alpha \in [0, 1]$, meaning the objective function is a non-convex function.

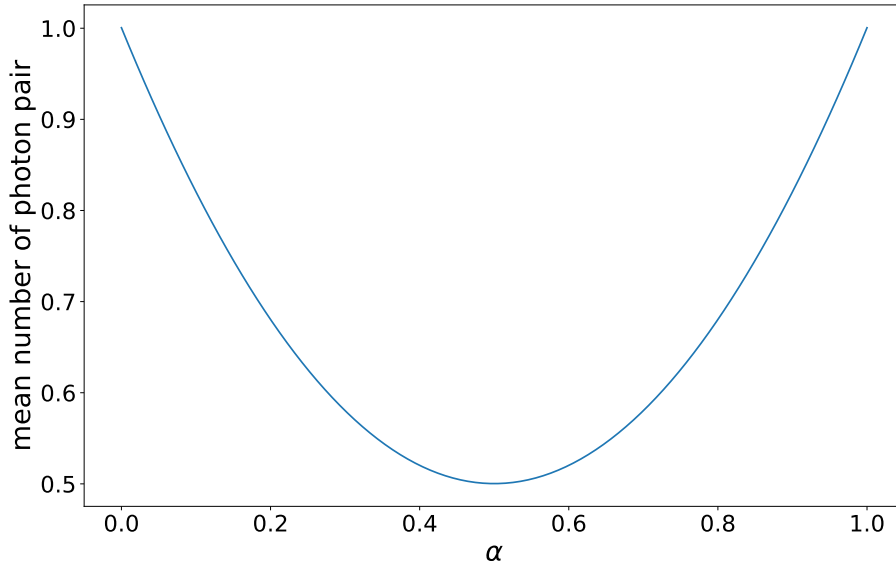


Figure 2.3 Criteria for convexity for the constraint on $\langle n \rangle$. For $\alpha = 0$ and $\alpha = 1$, the constraint is respected ($\langle n \rangle = 1$), but for any other value, it is not. The set generated by the constraint is then a non-convex set.

where λ_i and ν_l are the Lagrange multipliers for the inequality and equality constraints, respectively. It can be shown that [66]

$$\left(\min_{\tilde{X}} L(X, \lambda, \nu) \right) \leq L(\tilde{X}, \lambda, \nu) \leq f_0(\tilde{X}), \quad (2.104)$$

if $\lambda_i \geq 0$, where $\tilde{X}$ is any feasible point, meaning that $f_i(\tilde{X}) \leq 0$ and $f_l(\tilde{X}) = 0$ for all i and l . Eq. (2.104) means that for any values of λ_i and ν_l , the minimum value of the Lagrangian is always lower or equal to the objective function evaluated at any feasible point if the problem is bounded from below and if there exists a feasible set to the problem. In the case where the problem has no lower bounds, we can say that

$$\min_{\tilde{X}} L(X, \lambda, \nu) = -\infty, \quad (2.105)$$

or in the case where it does not have a feasible point,

$$\min_{\tilde{X}} L(X, \lambda, \nu) = \infty \quad (2.106)$$

holds true. The function

$$\min_{\tilde{X}} L(X, \lambda, \nu) = D(\lambda, \nu) \quad (2.107)$$

is often called the dual of the problem. Then, if scenarios in Eq. (2.105) and Eq. (2.106) do not hold, it can be seen that

$$\left(\min_{\tilde{X}} L(X, \lambda, \nu) \quad \forall \quad \lambda, \nu \right) \leq \left(\max_{\lambda, \nu} \min_{\tilde{X}} L(X, \lambda, \nu) \right) \leq \left(\min_{\tilde{X}} f_0(\tilde{X}) \right). \quad (2.108)$$

While Eq. (2.104) gives a lower bound on the original problem, Eq. (2.108) gives the best lower bound that the dual can offer. If the second inequality in Eq. (2.108) is equality, then the solution to the original problem in Eq. (2.102) is obtained through the dual and the problem is strongly dual to the original problem [66].

2.5 Quadratically Constrained Quadratic Programs (QCQP)

A QCQP is a problem where the objective function, as well as the constraints, are quadratic

$$\begin{aligned} \min_x \quad & f_0(x) = x^T Q_0 x + P_0^T x + c_0, \\ \text{s.t.} \quad & f_i(x) = x^T Q_i x + P_i^T x + c_i = 0, \\ & g_l(x) = x^T Q_l x + P_l^T x + c_l \leq 0, \end{aligned} \tag{2.109}$$

where $x \in \mathbb{R}^n$. Again, $i \in \mathfrak{I}$ and $l \in \mathfrak{L}$ where $\mathfrak{I} \cup \mathfrak{L}$ is the interval of positive integers up to a value n such that there are n constraints. The problem can easily be reformulated by taking into account complex valued vectors too. Design problems in electromagnetism can often be modeled as a quadratically constrained quadratic problem (QCQP) stemming from the existence of a Fredholm integral form to the differential equations governing the electromagnetic field [47]. In quantum mechanics, such an integral form exists as the Dyson series [67]. It would be natural to think of QCQPs when dealing with systems under linear differential equation constraints.

QCQPs can be written with matrices instead of vectors, whether complex or real valued. Assume for the space of complex valued matrices where $Y, X \in \mathbb{C}^{N \times M}$, let $\langle Y, X \rangle = \text{Tr} [Y^\dagger X]$ represent the inner product between elements of the said spaces, the problem, often called quadratic matrix problem or quadratic matrix program, can be stated as

$$\begin{aligned} \min_X \quad & f_0(X) = \langle X, Q_0 X \rangle + \frac{1}{2} (\langle P_0, X \rangle + \langle X, P_0 \rangle) + c_0, \\ \text{s.t.} \quad & f_i(X) = \langle X, Q_i X \rangle + \frac{1}{2} (\langle P_i, X \rangle + \langle X, P_i \rangle) + c_i = 0, \\ & g_l(X) = \langle X, Q_l X \rangle + \frac{1}{2} (\langle P_l, X \rangle + \langle X, P_l \rangle) + c_l \leq 0, \end{aligned} \tag{2.110}$$

for Hermitian Q_k , where $k \in \mathfrak{I} \cup \mathfrak{L} \cup \{0\}$. Not only is the problem non-convex, but it is also NP-hard in general [68]. Here is where the Lagrangian dual is useful to bound the solution of the quadratic matrix problem.

2.5.1 The Lagrangian dual of QCQP

The Lagrangian of the problem in Eq. (2.110) can be written as

$$\begin{aligned}
L(X, \lambda, \nu) &= \langle X, Q_0 X \rangle + \frac{1}{2} (\langle P_0, X \rangle + \langle X, P_0 \rangle) + c_0 \\
&\quad + \sum_i \lambda_i \left(\langle X, Q_i X \rangle + \frac{1}{2} (\langle P_i, X \rangle + \langle X, P_i \rangle) + c_i \right) \\
&\quad + \sum_l \nu_l \left(\langle X, Q_l X \rangle + \frac{1}{2} (\langle P_l, X \rangle + \langle X, P_l \rangle) + c_l \right) \\
&= \langle X, (Q_0 + \sum_i \lambda_i Q_i + \sum_l \nu_l Q_l) X \rangle \\
&\quad + \frac{1}{2} \left(\left\langle \left(P_0 + \sum_i \lambda_i P_i + \sum_l \nu_l P_l \right), X \right\rangle + \left\langle X, \left(P_0 + \sum_i \lambda_i P_i + \sum_l \nu_l P_l \right) \right\rangle \right) \\
&\quad + c_0 + \sum_i \lambda_i c_i + \sum_l \nu_l c_l. \tag{2.111}
\end{aligned}$$

Let

$$Q(\lambda, \nu) = Q_0 + \sum_i \lambda_i Q_i + \sum_l \nu_l Q_l$$

be positive semidefinite, meaning that no eigenvalue of $Q(\lambda, \nu)$ is negative [68]. The argmin of Lagrangian with respect to X is

$$X^* = -\frac{1}{2} Q^{-\dagger}(\lambda, \nu) P(\lambda, \nu), \tag{2.112}$$

where $-\dagger$ is the Hermitian conjugate of the general pseudo-inverse. If $Q(\lambda, \nu)$ is invertible, then the general pseudo-inverse is simply the standard matrix inverse. Otherwise, the matrix $P(\lambda, \nu)$ needs to be within the column space (range) of $Q(\lambda, \nu)$ [68]. Note that an argmin of a function $f(X)$ is the value of X where $f(X)$ attains its minimum. Since $Q(\lambda, \nu)$ is Hermitian, $Q^{-\dagger}(\lambda, \nu) = Q^{-1}(\lambda, \nu)$. Taking $P(\lambda, \nu) = P_0 + \sum_i \lambda_i P_i + \sum_l \nu_l P_l$, the dual then can be written as

$$\begin{aligned}
D(\lambda, \nu) &= \langle X^*, Q(\lambda, \nu) X^* \rangle \\
&\quad + \frac{1}{2} \left(\langle P(\lambda, \nu), X^* \rangle + \langle \hat{X}, P(\lambda, \nu) \rangle \right) \\
&\quad + c_0 + \sum_i \lambda_i c_i + \sum_l \nu_l c_l. \tag{2.113}
\end{aligned}$$

By inserting the term in Eq. (2.112) into Eq. (2.113), the dual problem can then be stated as

$$\begin{aligned} \max_{\lambda, \nu} \quad & -\frac{1}{4} \left\langle P(\lambda, \nu), Q(\lambda, \nu)^{-1} P(\lambda, \nu) \right\rangle + c_0 + \sum \lambda_i c_i + \sum \nu_l c_l, \\ \text{s.t.} \quad & Q(\lambda, \nu) \succcurlyeq 0, \\ & \nu_l \geq 0, \end{aligned} \tag{2.114}$$

where $A \succcurlyeq 0$ means matrix A is positive semidefinite. The constraint $Q(\lambda, \nu) \succcurlyeq 0$ is needed for the minimum of the Lagrangian over X to not be $-\infty$. A property of the dual problem is that it is a convex problem, as the objective function is concave and the constraint set is convex as it should be. The objective in Eq. (2.114) is sometimes called the matrix fractional function.

Another way to obtain a lower bound on the quadratic matrix problem is with semidefinite relaxation (SDR) of the original problem. The SDR is closely related to the dual. Defining the operator $\text{vec}(\cdot)$ which, when applied to a matrix, vectorizes it:

$$\text{vec}(A) = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_N \end{bmatrix}, \tag{2.115}$$

the element A_i is the i th column of the matrix A . Let $p(\lambda, \nu) = \text{vec}(P(\lambda, \nu))$ be the vectorized matrix $P(\lambda, \nu)$ and $q(\lambda, \nu) = \mathbb{I} \otimes Q(\lambda, \nu)$ where $\mathbb{I}$ is the identity matrix with the same dimensions as $Q(\lambda, \nu)$, the objective function of the dual can be written as

$$\begin{aligned} \max \quad & \gamma, \\ \text{s.t.} \quad & -\frac{1}{4} p(\lambda, \nu)^\dagger q(\lambda, \nu)^{-1} p(\lambda, \nu) + c_0 + \sum \lambda_i c_i + \sum \nu_l c_l \geq \gamma. \end{aligned} \tag{2.116}$$

Defining Schur's complement X/C of a Hermitian matrix X as $A - BC^{-1}B^\dagger$ and X/A as $C - B^\dagger A^{-1}B$ where

$$X = \begin{bmatrix} A & B \\ B^\dagger & C \end{bmatrix}, \tag{2.117}$$

the matrix X is positive semidefinite if and only if C is positive definite, and X/C is positive semidefinite when A is invertible. It is also positive semidefinite if and only if A is positive

definite and X/A is positive semidefinite when C is invertible [69]. Then, assuming $q(\lambda, \nu)$ is invertible, the problem in Eq. (2.116) is equivalent to [70]

$$\begin{aligned} \max \quad & \gamma, \\ \text{s.t.} \quad & - \begin{bmatrix} q(\lambda, \nu) & \frac{1}{2}p(\lambda, \nu) \\ \frac{1}{2}p(\lambda, \nu)^\dagger & \gamma - c_0 - \sum \lambda_i c_i - \sum \nu_l c_l \end{bmatrix} \succcurlyeq 0. \end{aligned} \quad (2.118)$$

The problem in Eq. (2.118) is a semidefinite program, which is a class of problems where an affine function is being optimized and where the constraint set is made of linear equalities and/or inequalities. The “semidefinite” in semidefinite program comes from the addition of a constraint that makes some linear superposition of matrices positive semidefinite [70]. While the specific semidefinite program in Eq. (2.118) is different from the SDR, we showcase that the dual is related to the semidefinite relaxation as both are semidefinite programs. A reason to work with SDR specifically is because a low-rank solution to the SDR gives directly the solution of the original problem, as will be shown in the next section.

2.5.2 Semidefinite relaxation

The quadratic problem in Eq. (2.110) can be written as

$$\begin{aligned} \min_X \quad & f_0(X) = \text{Tr}[Q_0 Y] + \frac{1}{2} (\langle P_0, X \rangle + \langle X, P_0 \rangle) + c_0, \\ \text{s.t.} \quad & f_i(X) = \text{Tr}[Q_i Y] + \frac{1}{2} (\langle P_i, X \rangle + \langle X, P_i \rangle) + c_i = 0, \\ & g_l(X) = \text{Tr}[Q_l Y] + \frac{1}{2} (\langle P_l, X \rangle + \langle X, P_l \rangle) + c_l \leq 0, \\ & Y = X X^\dagger. \end{aligned} \quad (2.119)$$

The constraint $Y = X X^\dagger$ is non-convex. Using the constraint

$$Y - X X^\dagger \succcurlyeq 0 \quad (2.120)$$

instead, any X and Y satisfying $Y = X X^\dagger$ will also be feasible for Eq. (2.120). Semidefinite relaxation’s fundamental idea comes from how this inequality becomes equality when the rank of Y is low enough. This new constraint is convex, and the problem stemming from using the semidefinite constraint instead of the equality constraint is the semidefinite relaxation of the original problem in Eq. (2.119). Any constraints acting on the quadratic parts of X will be acting on Y instead as affine constraints (equations of the form $\text{Tr}[\Gamma Y] - d = 0$ for some

constant matrix Γ and real scalar d). The problem in Eq. (2.119) can then be written as

$$\begin{aligned} \min_{Y, X} \quad & f_0(X, Y) = \text{Tr}[Q_0 Y] + \frac{1}{2} (\langle P_0, X \rangle + \langle X, P_0 \rangle) + c_0, \\ \text{s.t.} \quad & f_i(X, Y) = \text{Tr}[Q_i Y] + \frac{1}{2} (\langle P_i, X \rangle + \langle X, P_i \rangle) + c_i = 0, \\ & g_l(X, Y) = \text{Tr}[Q_l Y] + \frac{1}{2} (\langle P_l, X \rangle + \langle X, P_l \rangle) + c_l \leq 0, \\ & Y - XX^\dagger \succcurlyeq 0. \end{aligned} \quad (2.121)$$

If Y^* and X^* are the optimal solutions, then if the eigenvalues of $Y^* - X^* (X^\dagger)^*$ are close enough to zero, meaning that first, we take into account that M is the smaller dimension of X^* (X^* being a $N \times M$ matrix with $M \leq N$). Then, we can take a matrix $\tilde{Y}$ made from the M largest eigenvalues of Y^* and from the associated eigenvectors. If $f_0(X^*, \tilde{Y}) \approx f_0(X^*, Y^*)$ with $f_i(X^*, \tilde{Y}) \approx 0$ and $g_l(X^*, \tilde{Y}) \leq 0$, for all i, l within some numerical error, then we can give the approximation $Y^* \approx X^* (X^\dagger)^*$. Then X^* will be a global argmin of problem in Eq. (2.110). Otherwise, the solution of Eq. (2.121) will only offer a lower bound to the problem of Eq. (2.110). Through Schur's complement, Eq. (2.120) needs to be true for a matrix

$$Z \approx \begin{bmatrix} Y & X \\ X^\dagger & \mathbb{I} \end{bmatrix} \quad (2.122)$$

to be positive semidefinite. Let n be the dimension of X ($X \in \mathbb{C}^{n \times m}$) where $n \leq m$, if Z is of rank n , then it can be written as

$$Z = \begin{bmatrix} XX^\dagger & X \\ X^\dagger & \mathbb{I} \end{bmatrix} \quad (2.123)$$

and the eigenvalues of $Y - XX^\dagger$ will be close to zero. X can then be extracted from Z . If Z is of higher rank than n , then X cannot be extracted. The terms Q_k , P_k and c_k for all $k \in \mathfrak{J} \cup \mathfrak{L} \cup 0$ will be modified to match the size of Z .

$$A_k = \begin{bmatrix} Q_k & \frac{1}{2}P_k \\ \frac{1}{2}P_k^\dagger & \frac{c_k}{N}\mathbb{I} \end{bmatrix} \quad (2.124)$$

can be assigned to the objective function as well as to each of the constraints. Formally, the semidefinite relaxation is given as

$$\begin{aligned}
\min_Z \quad & f_0(Z) = \text{Tr}[A_0 Z], \\
\text{s.t.} \quad & f_i(Z) = \text{Tr}[A_i Z] = 0, \\
& g_l(Z) = \text{Tr}[A_l Z] \leq 0, \\
& \text{Tr} \left[\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & E_{bj} + E_{jb} \end{bmatrix} Z \right] = 2\delta_{bj}, \\
& Z \succcurlyeq \mathbf{0},
\end{aligned} \tag{2.125}$$

where $E_{bj} \in \mathbb{R}^{n \times n}$ (n being a dimension of X) is a matrix whose bj th index is 1 and 0 for any other indices (so a matrix with a single non-zero element) and δ_{bj} is Kronecker's delta. It has been shown that such relaxation of the quadratic matrix problem gives the same solution as the initial problem if $\mathfrak{J} \cup \mathfrak{L}$ has n elements [71]. The program shown in Eq. (2.125) is slightly different to the commonly seen SDR of the QCQP since it is a relaxation of a quadratic matrix program rather than a vector-based problem [70].

2.6 Spectral Minimization Problem as a Quadratically Constrained Problem

2.6.1 Equations of motion as quadratic constraints

By changing the frame of reference such that

$$\tilde{U}_{\pm}(\zeta) = \exp[-\mathbf{K}\zeta] \mathbf{U}_{\pm}(\zeta), \tag{2.126}$$

$$\tilde{\mathbf{V}} = \exp[-\mathbf{K}\zeta] \mathbf{V} \exp[\mathbf{K}\zeta], \tag{2.127}$$

the equations of motion for the propagators become

$$\partial_{\zeta} \tilde{U}_{\pm}(\zeta) = \pm \tilde{\mathbf{V}} \tilde{U}_{\pm}(\zeta). \tag{2.128}$$

They can be written as Fredholm integral equations by keeping in mind that $\tilde{U}_{\pm}(z=0) = \mathbf{I}$

$$\tilde{U}_{\pm}(\zeta) = \mathbf{I} \pm \int_0^{\zeta} d\zeta' \tilde{\mathbf{V}} \tilde{U}_{\pm}(\zeta'), \tag{2.129}$$

which results in equations

$$\mathbf{U}_{\pm}(\zeta) = \exp[\mathbf{K}\zeta] \pm \int_0^{\zeta} d\zeta' \exp[\mathbf{K}(\zeta - \zeta')] \mathbf{V} \mathbf{U}_{\pm}(\zeta'), \quad (2.130)$$

when changed back to the initial frame of reference. Eq. (2.130) is a constraint that must be followed by any feasible propagators. It also turns out it can be written in a quadratic form, as will be shown later.

2.6.2 Structure of variables

In the last section, the integral form of the equations of motion were laid out. To approximate it numerically, the waveguide is discretized with $\zeta \approx j\Delta\zeta = \zeta_j$, Eqs. (2.130) can be written as

$$\mathbf{U}_{\pm}(\zeta_j) = \mathbf{U}_{\pm j} \approx \exp[\mathbf{K}\zeta_j] \pm \frac{\Delta\zeta}{2} (\mathbf{V} \mathbf{U}_{\pm j} + \exp[\mathbf{K}\zeta_j] \mathbf{V} \mathbf{U}_{\pm 0}) \pm \Delta\zeta \sum_{k=1}^{j-1} \exp[\mathbf{K}\zeta_{j-k}] \mathbf{V} \mathbf{U}_{\pm k}. \quad (2.131)$$

The error on the approximation will depend on the value of $\Delta\zeta$ (size of discretization) as well as on the norm of $\mathbf{V}$. Since the constraint is over all propagators throughout the waveguide, the variables need to take them into account. They also need to include the pump $\mathbf{V}$. The structure of the input to the quadratic matrix program would then be

$$\mathbf{X} = \begin{bmatrix} \mathbf{U}_{+1} \\ \mathbf{U}_{+2} \\ \vdots \\ \mathbf{U}_{+N_z} \\ \mathbf{U}_{-1} \\ \mathbf{U}_{-2} \\ \vdots \\ \mathbf{U}_{-N_z} \\ \mathbf{V} \end{bmatrix}, \quad (2.132)$$

where $\mathbf{U}_{\pm 0}$ is omitted since $\mathbf{U}_{\pm 0} = \mathbf{I}$ is the identity. It is implied that there are N_z steps in ζ such that the waveguide is of length $N_z\Delta\zeta$. A schematic representation of the problem is shown in Fig. 2.4.

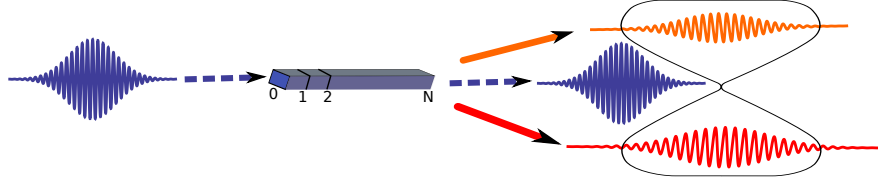


Figure 2.4 A visualization of the problem. The discretization step is of equal distances, meaning the length between 0 and 1 is $\Delta\zeta$, which is also the length between 1 and 2 and so on. The propagators at 0 are $\mathbf{U}_{\pm 0}$, or $\mathbf{I}$, at 1 they are $\mathbf{U}_{\pm 1}$, at 2 $\mathbf{U}_{\pm 2}$ until at N, $\mathbf{U}_{\pm N_z}$. The pulse in blue is the pump beam, the ones in orange and red are signal and idler beams with less energy than the pump pulse. They are often entangled spectrally.

2.6.3 Additional quadratic constraints

Because the constraints on the dynamics are non-convex, a convex relaxation is needed. In order for the convex program to obtain better bounds, it is very useful to add constraints that would be redundant to the original non-convex problem [72]. Adding them would reduce the feasible set of the relaxed problem and would allow the convex feasible set to be closer to the non-convex one.

Constraints from the commutation relations of pseudo-quadratures can be included. These are consequences of the matrix exponential nature of the propagators. Taking into account the equations

$$\mathbf{U}_{+j}\mathbf{U}_{-k}^\dagger = \mathbf{U}_{+(j-k)}, \quad (2.133a)$$

$$\mathbf{U}_{-j}\mathbf{U}_{+k}^\dagger = \mathbf{U}_{-(j-k)}, \quad (2.133b)$$

additional relations are created between the quadratic and linear terms.

For the relaxed problem to be compact and well-defined, it is useful to add constraints on the elements $\mathbf{U}_{\pm j}\mathbf{U}_{\pm j}^\dagger$ as well as on $\mathbf{V}\mathbf{V}^\dagger$. For the constraints on $\langle n \rangle$, which is the mean number of photon pairs per pulse at the end of the waveguide, it is assumed that $\langle n \rangle(\zeta_N) \geq \langle n \rangle(\zeta_{k+1}) \geq \langle n \rangle(\zeta_k)$, that is, the number of mean photon pairs is a non-decreasing

function of the waveguide length. Then, additional constraints to the problem can include

$$\frac{1}{4} \text{Tr} [\mathbf{U}_{+j}^\dagger \mathbf{U}_{+j} + \mathbf{U}_{-j}^\dagger \mathbf{U}_{-j} - (\mathbf{U}_{+(j-1)}^\dagger \mathbf{U}_{+(j-1)} + \mathbf{U}_{-(j-1)}^\dagger \mathbf{U}_{-(j-1)})] \geq 0, \quad (2.134)$$

$$\frac{1}{4} \text{Tr} [\mathbf{U}_{+N_z}^\dagger \mathbf{U}_{+N_z} + \mathbf{U}_{-N_z}^\dagger \mathbf{U}_{-N_z} - 2\mathbf{I}] - \langle n \rangle = 0, \quad (2.135)$$

for any $1 \leq j \leq N_z$, since Eq. (2.38) written in terms of new propagators as

$$\text{Tr} [\mathbf{U}_{SI}(\zeta_{N_z})^* \mathbf{U}_{SI}^T(\zeta_{N_z})] - \langle n \rangle = 0, \quad (2.136)$$

$$\frac{1}{4} \text{Tr} [(\mathbf{U}_{+N_z} - \mathbf{U}_{-N_z})(\mathbf{U}_{+N_z} - \mathbf{U}_{-N_z})^\dagger] - \langle n \rangle = 0. \quad (2.137)$$

Without loss of generality, the matrix representing the pump can always be constrained such that

$$\text{Tr} [\mathbf{V}\mathbf{V}^\dagger] - 1 = \text{Tr} [\mathbf{V}^2] - 1 = 0. \quad (2.138)$$

Note that this constraint is not the same as saying

$$\int d\Omega \beta^2(\Omega) = 1, \quad (2.139)$$

but rather

$$\int d\Omega d\Omega' \beta^2(\Omega + \Omega') = 1. \quad (2.140)$$

The change brought by Eq. (2.138) leads to a modification of the dynamics. Instead of using Eq. (2.131), the constraints giving the discretization of the integral form would be

$$\mathbf{U}_{\pm j} = \exp[\mathbf{K}\zeta_j] \pm \frac{|V|\Delta\zeta}{2} (\mathbf{V}\mathbf{U}_{\pm j} + \exp[\mathbf{K}\zeta_j] \mathbf{V}\mathbf{U}_{\pm 0}) \pm |V|\Delta\zeta \sum_{k=1}^{j-1} \exp[\mathbf{K}\zeta_{j-k}] \mathbf{V}\mathbf{U}_{\pm k}, \quad (2.141)$$

with $|V|$ being akin to a power of the pump that can be modified.

2.6.4 Affine constraints

In order to respect the physics, the matrix $\mathbf{V}$ needs to be of Hankel form ($\mathbf{V}_{ij} = \mathbf{V}_{i+k, j-k} \forall k \in [0, j-i]$ assuming $i \leq j$) and needs to be real-valued under the initial assumptions of equations

of motion decoupling. These constraints lead to adding

$$\text{Im}\{\mathbf{V}\} = \mathbf{0}, \quad (2.142)$$

$$\mathbf{V} = \sum_{i=0}^{2N_\omega-2} c_i \mathbf{H}_i \quad (2.143)$$

to the program, with $\mathbf{H}_i$ being matrices such that

$$(\mathbf{H}_i)_{jk} = \begin{cases} 1 & \text{if } j+k=i, \\ 0 & \text{else.} \end{cases} \quad (2.144)$$

For example, if $\mathbf{H}_i \in \mathbb{R}^{3 \times 3}$, then if $i=3$,

$$\mathbf{H}_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \quad (2.145)$$

2.6.5 Real scalar valued constraints

Typically, constraints stem from functions mapping from a vector space to the reals. For matrix equalities, element-wise constraints will be given by the trace of a multiplication between said matrix with elements of a basis. For example, a very simple constraint with complex valued matrices can be written from $\mathbf{XY} = \mathbf{Z}$ to $\text{Tr}[\mathbf{R}_i^\dagger \mathbf{XY} - \mathbf{R}_i^\dagger \mathbf{Z}] = 0$ where $\{\mathbf{R}_i\}$ forms a complete matrix basis such that $\sum_i c_i \mathbf{R}_i = \mathbf{A}$ for any $\mathbf{A} \in \mathbb{R}^{K \times M}$, where K and M are the dimensions of $\mathbf{Z}$. To turn the equality from mapping into complex scalars to mapping into real values, adding the trace of complex conjugate would suffice, such as

$$\text{Tr}[\mathbf{R}_i^\dagger \mathbf{XY} - \mathbf{R}_i^\dagger \mathbf{Z}] + \text{Tr}[\mathbf{R}_i (\mathbf{XY})^\dagger - \mathbf{R}_i \mathbf{Z}^\dagger] = 0. \quad (2.146)$$

The previous equation is a constraint stating

$$\Re\{\mathbf{XY} - \mathbf{Z}\} = \mathbf{0}, \quad (2.147)$$

or the real part of the matrix constraint needs to be satisfied. For Eq. (2.133), the scalar constraints for $\mathbf{U}_{\pm j} \mathbf{U}_{\mp k}^\dagger$ are shown in Eq. (2.148a) and the ones for the dynamics (Eq. (2.131)) are shown in Eq. (2.148b).

$$\text{Tr}[\mathbf{R}_i^\dagger \mathbf{U}_{\pm j} \mathbf{U}_{\mp k}^\dagger - \mathbf{R}_i^\dagger \mathbf{U}_{\pm(j-k)}] + \text{Tr}[\mathbf{R}_i \mathbf{U}_{\mp k} \mathbf{U}_{\pm j}^\dagger - \mathbf{R}_i \mathbf{U}_{\pm(j-k)}] = 0, \quad (2.148a)$$

$$\begin{aligned}
& \text{Tr} \left[-\mathbf{R}_i \mathbf{U}_{\pm j} + \mathbf{R}_i \exp[\mathbf{K} \zeta_j] \pm \frac{|V| \Delta \zeta}{2} \mathbf{R}_i \mathbf{U}_{\pm j} \mathbf{V}^\dagger + \exp[\mathbf{K} \zeta_j] \mathbf{R}_i \mathbf{U}_{\pm 0} \mathbf{V}^\dagger \right] \\
& + \text{Tr} \left[\pm |V| \Delta \zeta \sum_{k=1}^{j-1} \exp[\mathbf{K} \zeta_{j-k}] \mathbf{R}_i \mathbf{U}_{\pm k} \mathbf{V}^\dagger - \mathbf{R}_i^\dagger \mathbf{U}_{\pm j}^\dagger + \mathbf{R}_i^\dagger \exp[-\mathbf{K} \zeta_j] \right] \\
& \pm \text{Tr} \left[\frac{|V| \Delta \zeta}{2} \mathbf{R}_i^\dagger (\mathbf{V} \mathbf{U}_{\pm j}^\dagger + \exp[-\mathbf{K} \zeta_j] \mathbf{V} \mathbf{U}_{\pm 0}^\dagger) \pm |V| \Delta \zeta \sum_{k=1}^{j-1} \mathbf{R}_i^\dagger \exp[-\mathbf{K} \zeta_{j-k}] \mathbf{V} \mathbf{U}_{\pm k}^\dagger \right] = 0,
\end{aligned} \tag{2.148b}$$

for all i . The imaginary part of the constraints can be obtained by changing $\mathbf{R}_i \rightarrow i\mathbf{R}_i$ and keeping the same equations as in Eqs. (2.148). Because Eq. (2.143) paired with Eq. (2.142) imply there are different constraints on the real and imaginary parts of V , they become, as real scalar-valued functions,

$$\text{Tr} [-i\mathbf{R}_i^T \mathbf{V} + i\mathbf{R}_i \mathbf{V}^\dagger] = 0, \tag{2.149}$$

$$\text{Tr} [\mathbf{R}_i^T \mathbf{V} - \sum c_j \mathbf{R}_i^T \mathbf{H}_j + \mathbf{R}_i \mathbf{V}^\dagger - \sum c_j \mathbf{R}_i \mathbf{H}_j^\dagger] = 0, \tag{2.150}$$

where there is no need for multiplication of $\mathbf{R}_i$ by i . The constraints (2.138), (2.135) and (2.134) do not require $\mathbf{R}_i$ projections, as they are already real scalar-valued constraints.

2.7 Convexification of The Spectral Minimization Problem

With the materials from previous sections, the convex relaxation of spectral separability of SPDC can be stated.

2.7.1 Semidefinite relaxation of quadratic constraints

The quadratic constraints governing the dynamics and the commutation relations do not form a convex set. The relaxation in Sec. 2.5.2 will be used to approximate these constraints with convex ones by creating a semidefinite matrix $\mathbf{Y}$. If at optimum, it is of rank N_ω or lower, then for $\mathbf{X} \in \mathbb{C}^{(2N_z-1)N_\omega \times N_\omega}$,

$$\mathbf{Y}^* \approx \begin{bmatrix} \mathbf{X}^* (\mathbf{X}^\dagger)^* & \mathbf{X}^* \\ (\mathbf{X}^\dagger)^* & \mathbf{I} \end{bmatrix}, \tag{2.151}$$

where $\mathbf{X}$ is defined in Sec. 2.6.2. Then the pump $\mathbf{V}$ and the propagators $\mathbf{U}_{\pm N}$ can be extracted by taking into account that the matrix would be

$$\mathbf{Y}^* = \left[\begin{array}{ccc|ccc|c|c} \mathbf{U}_{+1}\mathbf{U}_{+1}^\dagger & \mathbf{U}_{+1}\mathbf{U}_{+2}^\dagger & \cdots & \mathbf{U}_{+1}\mathbf{U}_{-1}^\dagger & \mathbf{U}_{+1}\mathbf{U}_{-2}^\dagger & \cdots & \mathbf{U}_{+1}\mathbf{V}^\dagger & \mathbf{U}_{+1} \\ \mathbf{U}_{+2}\mathbf{U}_{+1}^\dagger & \mathbf{U}_{+2}\mathbf{U}_{+2}^\dagger & \cdots & \mathbf{U}_{+2}\mathbf{U}_{-1}^\dagger & \mathbf{U}_{+2}\mathbf{U}_{-2}^\dagger & \cdots & \mathbf{U}_{+2}\mathbf{V}^\dagger & \mathbf{U}_{+2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \hline \mathbf{U}_{-1}\mathbf{U}_{+1}^\dagger & \mathbf{U}_{-1}\mathbf{U}_{+2}^\dagger & \cdots & \mathbf{U}_{-1}\mathbf{U}_{-1}^\dagger & \mathbf{U}_{-1}\mathbf{U}_{-2}^\dagger & \cdots & \mathbf{U}_{-1}\mathbf{V}^\dagger & \mathbf{U}_{-1} \\ \mathbf{U}_{-2}\mathbf{U}_{+1}^\dagger & \mathbf{U}_{-2}\mathbf{U}_{+2}^\dagger & \cdots & \mathbf{U}_{-2}\mathbf{U}_{-1}^\dagger & \mathbf{U}_{-2}\mathbf{U}_{-2}^\dagger & \cdots & \mathbf{U}_{-2}\mathbf{V}^\dagger & \mathbf{U}_{-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \hline \mathbf{V}\mathbf{U}_{+1}^\dagger & \mathbf{V}\mathbf{U}_{+2}^\dagger & \cdots & \mathbf{V}\mathbf{U}_{-1}^\dagger & \mathbf{V}\mathbf{U}_{-2}^\dagger & \cdots & \mathbf{V}\mathbf{V}^\dagger & \mathbf{V} \\ \hline \mathbf{U}_{+1}^\dagger & \mathbf{U}_{+2}^\dagger & \cdots & \mathbf{U}_{-1}^\dagger & \mathbf{U}_{-2}^\dagger & \cdots & \mathbf{V}^\dagger & \mathbf{I} \end{array} \right]. \quad (2.152)$$

The lines separating different blocks are there to give a better idea as to what each block should ideally represent. The blocks representing $\mathbf{U}_{\pm i}\mathbf{U}_{\pm j}^\dagger$, $\mathbf{U}_{\mp i}\mathbf{U}_{\pm j}^\dagger$, $\mathbf{V}\mathbf{V}^\dagger$, $\mathbf{V}\mathbf{U}_{\pm k}^\dagger$ and $\mathbf{U}_{\pm k}\mathbf{V}^\dagger$ with $i, j, k \in N_z$ will often be called the quadratic terms and the blocks representing $\mathbf{U}_{\pm i}$ and $\mathbf{V}$ the linear terms through the rest of the manuscript. The quadratic constraints on $\mathbf{X}$ will become affine on $\mathbf{Y}$. For example, the constraint associated with Eq. (2.148a) would be

$$\text{Tr} \left[\left(\mathbf{R}_i^\dagger \right)_{+j,-k} \mathbf{Y} - \left(\mathbf{R}_i^\dagger \right)_{+(j-k)} \mathbf{Y} + (\mathbf{R}_i)_{-k,+j} \mathbf{Y} - (\mathbf{R}_i)_{+(j-k)} \mathbf{Y} \right] = 0, \quad (2.153)$$

where $(\mathbf{R}_i)_{+j,-k}$ is the matrix of the same dimension as $\mathbf{Y}$, but has the block $\mathbf{R}_i$ where if the operation $\text{Tr} \left[(\mathbf{R}_i)_{+j,-k} \mathbf{Y}^* \right]$ is done, if $\mathbf{Y}^*$ is low-rank, would give the same value as $\text{Tr} \left[\mathbf{R}_i \mathbf{U}_{+j} \mathbf{U}_{-k}^\dagger \right]$. With $(\mathbf{R}_{+(j-k)})$, it would give $\text{Tr} \left[\mathbf{R}_{+(j-k)} \mathbf{Y}^* \right] = \text{Tr} \left[\mathbf{R}_i \mathbf{U}_{+(j-k)} \right]$. The matrix $\mathbf{Y}$ will sometimes be called the semidefinite matrix for the rest of the manuscript.

2.7.2 Convex objective function

The objective function from Eq. (2.87) is a convex function, as nuclear norm is a convex function. Paired with the constraints in the previous sections of Sec. 2.7, the convex relaxation of the problem of minimizing spectral entanglement (maximizing the spectral separability) of SPDC can be formulated. The optimal value of the convex problem will sometimes be called the solution to the SDP or to the SDR for the rest of the manuscript.

CHAPTER 3 RESULTS

3.1 Numerical Optimization With JAX

In the perturbative regime, with the phase-matching function being a sinc-like function, the slowly-varying envelope of the pump has to be Gaussian and has to have a specific width in order to maximize the spectral separability of idler and signal states when in the symmetric group velocity matching regime. The same pump shape will not be optimal for high-gain regime, as will be shown later. In this section, numerical optimization results using the BFGS (Broyden–Fletcher–Goldfarb–Shanno) algorithm [73], an optimization method from the quasi-Newton family, over different degrees of freedom of the pump as well as over different initial seeds are shown. The implementation comes from a module called Jaxopt [74], which is an optimization API built upon JAX [75].

The results come from setting ζ_f , or the length of the waveguide, to 1, $\Omega_i = -28.5$, $\Omega_f = 28.5$ and $N_\omega = 201$, where N_ω is the size of the frequency window, or the size of the propagators. Using the definition of $\Omega_f = \frac{\omega_f \ell}{\Delta v}$, if taking the physical parameters as given in [54] such that $\frac{1}{\Delta v} \approx 1.36 \times 10^{-10} \text{ s} \cdot \text{m}^{-1}$ and $\ell \approx 8 \times 10^{-3} \text{ m}$, then a range of Ω from -28.5 to 28.5 gives an approximate range of frequencies of $2.62 \times 10^{13} \text{ s}^{-1}$ around the center frequency. If the group velocity of the pump is approximately $1.72 \times 10^8 \text{ m} \cdot \text{s}^{-1}$, for a pump centered at $7.83 \times 10^{-7} \text{ m}$, the center frequency is given by $\bar{\omega}_P \approx 2.19 \times 10^{14} \text{ s}^{-1}$. The frequency range in wavelength can be given by $v_P \left(\frac{1}{2.19 \times 10^{14} - 2.63 \times 10^{13}} - \frac{1}{2.19 \times 10^{14} + 2.63 \times 10^{13}} \right) \approx 1.91 \times 10^{-7} \text{ m}$, or 191 nm. The results on pumps centered in frequency are however shown with frequency ranging from $\Omega_i = -14.25$ to $\Omega_f = 14.25$, to better compare the shape of the optimal pump between different values of $\langle n \rangle$. For the formulation of the problem in coupled dynamics, symmetric group velocity matching will be assumed.

The results from present section are shown in order:

1. Optimization over each pixel of the pump is done for different values of number of mean photon pairs per pulse ($\langle n \rangle$) with a uniformly distributed random initial pump varying between zero and 0.01 when in the symmetric group velocity matching regime. The pump's amplitude is dimensionless as shown in Sec. 2.2.2.
2. The same scheme is run, only this time the initial seed is replaced by a Gaussian with width that minimizes the Schmidt number (K) at low-gain regime, but scaled to give

$\langle n \rangle$.

3. Optimization results with a uniformly distributed random seed by modifying each pixel with regularization on the variance of the pump as well as penalty on the center frequency, again at different $\langle n \rangle$, are also shown.
4. Results with optimization using pseudo-quadrature propagators are then presented, as well as optimization over only Gaussians instead of all degrees of freedom of the pump.

3.1.1 Fixing the number of mean photon pairs per pulse

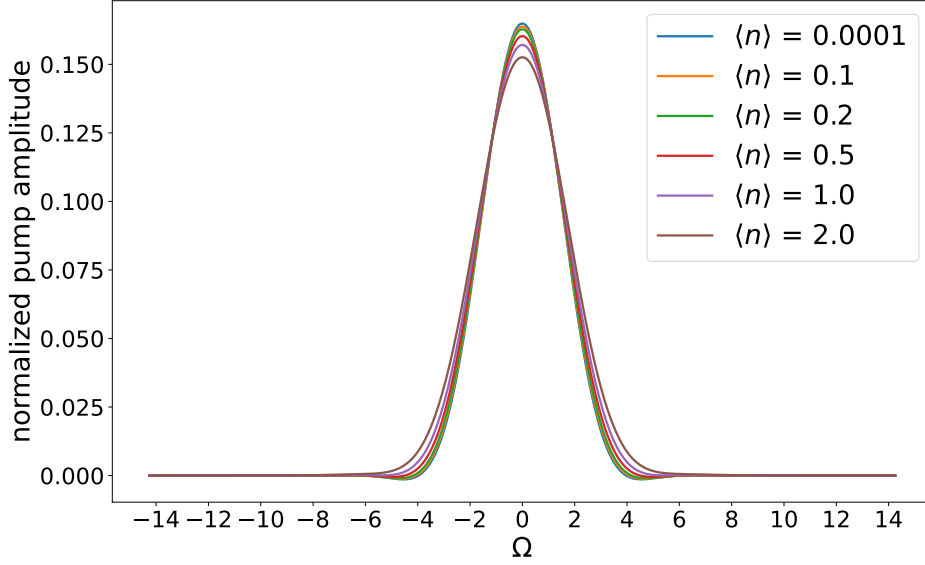
Taking the spectral purity as defined in Eq. (2.31), if the mean number of photon pairs, which is represented by $\text{Tr} [\mathbf{U}_{SI} \mathbf{U}_{SI}^\dagger]$, is at a fixed $\langle n \rangle$, the pump minimizing the Schmidt number is a Delta-like function at the boundary of the frequency domain for different values of $\langle n \rangle$. In other words, they are represented by a spike at the last pixel of the frequency range.

The Schmidt number for different values of $\langle n \rangle$ has value of 1 for this optimization scheme. An issue with these results is that, while the spectral separability is perfect, the nonzero amplitude wanders too far from the central frequency ($\Omega = 0$), going against the initial assumption of negligible group velocity dispersion. From Sec. 2.2.2, when the pump has high amplitude for Ω far from the center, for the same waveguide, it would mean the amplitude for ω_f far from the center is significant too. The further it is, the more the superior order terms of dispersion relations need to be taken into account.

3.1.2 Optimization taking into account the variance

In order for the program to not converge towards a pump that would not be physically viable (high amplitude for high values of $|\Omega|$), an initialization seed using the low-gain regime Gaussian pump, but with a different amplitude for different $\langle n \rangle$, can be taken. The Gaussian initial pump is taken to be real, as a global constant phase on the pump would not change the values of observables. Using such an initialization seed, the results of the optimization are shown in Fig. 3.1. Only the real part of the optimal pump is shown as the imaginary part has negligible amplitude in the chosen frame of reference. The Schmidt number for the optimized pumps at different values of $\langle n \rangle$ are shown in the table in Fig. 3.1.

A comparison between the Schmidt number of the optimal pumps with Gaussian seed and



$\langle n \rangle$	K
0.0001	1.218
0.1	1.206
0.2	1.197
0.5	1.172
1.0	1.144
2.0	1.110

Figure 3.1 Real part of the optimal pumps for different values of $\langle n \rangle$ with a Gaussian initialization seed. Note that the higher the gain is, the wider the optimal pump becomes. The real part of the pump is proportional to the imaginary part. It is then always possible to multiply the complex pump by another global phase such that the values of observables remain the same. In this new frame of reference, the imaginary part can be set to zero, which is why only the real part of the optimal pump is shown. The table under the visualization of the optimal pumps contains the Schmidt number obtained for each pump shown in the figure.

the Schmidt number with the best low-gain pump scaled so its mean number of photon pairs is $\langle n \rangle$ is shown in Fig. 3.2. With increasing $\langle n \rangle$, the difference in Schmidt number between both pumps increases.

While observations can be made with a Gaussian initial seed, it would be insightful to see what optimal results can be made with other initialization seeds, such as, again, with a

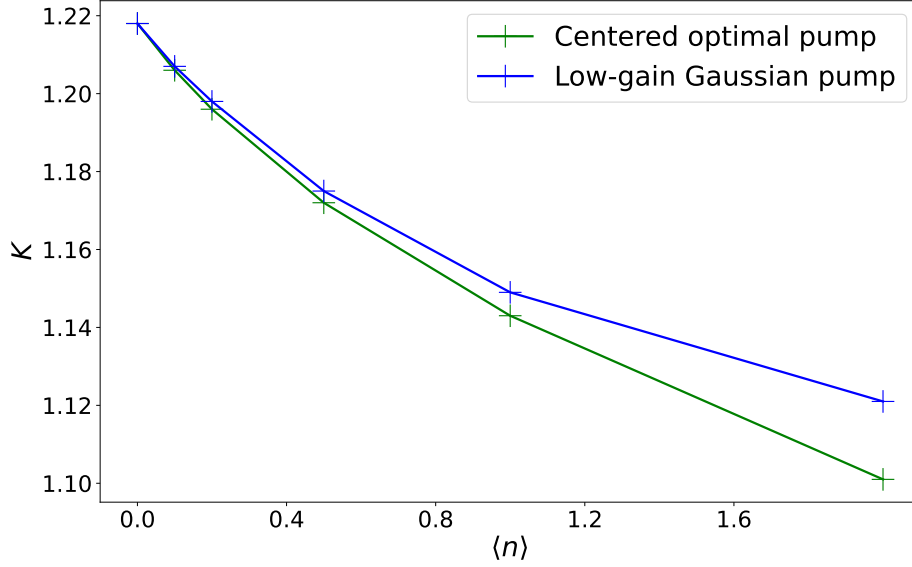


Figure 3.2 Comparison between the Schmidt number of the optimal pump using as seed the optimal Gaussian at low-gain regime and the Schmidt number of a Gaussian with width of said low-gain optimal pump and amplitude scaled to match $\langle n \rangle$. At low-gain, the Schmidt numbers are the same, but as $\langle n \rangle$ grows, the optimized pump has a lower value of K .

randomly distributed initialization. To keep the consideration of negligible group velocity dispersion valid, constraints on the variance of the pump can be added, as well as constraints forcing the center frequency of the pump to be set to zero. By normalizing the pump and treating it as a probability distribution, the variance can be expressed as

$$\sigma = \frac{\int d\Omega \Omega^2 |\beta(\Omega)|}{\int d\Omega |\beta(\Omega)|}, \quad (3.1)$$

$$= \frac{\sum_i \Omega_i^2 |\beta(\Omega_i)|}{\sum_j |\beta(\Omega_j)|}. \quad (3.2)$$

An inequality on this term can be set with a regularization. A constraint enforcing the expected value of frequency distribution of the pump to be zero, if treated as a probability, can be written as

$$\frac{\int d\Omega \Omega |\beta(\Omega)|}{\int d\Omega |\beta(\Omega)|} = \frac{\sum_i \Omega_i |\beta(\Omega_i)|}{\sum_j |\beta(\Omega_j)|} = 0. \quad (3.3)$$

Then, the regularized algorithm used is

```

pen = 5
j = 1
 $\beta = 0.01 \text{rand}(2N_\omega - 1)$ 
while  $\left(\text{Tr} [\mathbf{U}_{\text{SI}} \mathbf{U}_{\text{SI}}^\dagger] - \langle n \rangle\right)^2 \geq \epsilon$  do
     $\beta \rightarrow \text{argmin} \frac{\text{Tr} [\mathbf{U}_{\text{SI}} \mathbf{U}_{\text{SI}}^\dagger]^2}{\text{Tr} [(\mathbf{U}_{\text{SI}} \mathbf{U}_{\text{SI}}^\dagger)^2]} + \text{pen}^j \left( \left( \text{Tr} [\mathbf{U}_{\text{SI}} \mathbf{U}_{\text{SI}}^\dagger] - \langle n \rangle \right)^2 + \left( \frac{\sum_i \Omega_i |\beta(\Omega_i)|}{\sum_j |\beta(\Omega_j)|} \right)^2 \right) + \lambda \left( \frac{\sum_i \Omega_i^2 |\beta(\Omega_i)|}{\sum |\beta(\Omega_i)|} \right)$ 
     $j \rightarrow j + 1$ 
end while

```

where $\text{rand}(2N_\omega - 1)$ gives a randomly distributed vector of length $2N_\omega - 1$ varying between 0 and 1, λ is the Lagrange multiplier associated with the regularization on the variance of the pump, pen is the penalty coefficient that is being multiplied by its initial value every iteration and ϵ is the acceptable error on the mean photon number. The matrix $\mathbf{U}_{\text{SI}}$ is taken as a block of the matrix exponential as shown in Eq. (2.24). Taking $\lambda = 0.1$ and $\epsilon = \frac{\langle n \rangle}{10^3}$, the optimal pumps can be shown in Fig. 3.3 for different values of $\langle n \rangle$. Still only the real parts are shown, since the imaginary part is still proportional to the real part. This gives confidence that the optimization scheme can be applied to propagators of pseudo-quadratures where the pump is real. Subsequent plots will not be showing the optimal pump when $\langle n \rangle \ll 1$, since it is shown theoretically to be Gaussian [23]. The optimal pumps seem to be very similar to those of Fig. 3.1, meaning those same shapes can possibly correspond to global minima for the problem.

3.2 Optimization Over Pseudo-Quadratures

In the case where the equations of motion can be decoupled, the real part of the numerically optimal pumps for different values of $\langle n \rangle$ are shown in Fig. 3.4 using a randomly generated seed with constraints on the expected value of the frequency distribution and the variance of the pump. Once again, the results look very similar to those shown in Fig. 3.3. Note that the propagators of the annihilation and creation operators can be written as

$$\mathbf{U}_{SS} = \frac{1}{2} (\exp((\mathbf{K} + \mathbf{V}) \zeta) + \exp((\mathbf{K} - \mathbf{V}) \zeta)), \quad (3.4)$$

$$\mathbf{U}_{SI} = \frac{1}{2} (\exp((\mathbf{K} + \mathbf{V}) \zeta) - \exp((\mathbf{K} - \mathbf{V}) \zeta)), \quad (3.5)$$

as a consequence of eqs. (2.75), so the objective function and the constraint on $\langle n \rangle$ are made of functions of $\mathbf{U}_+(\zeta) = \exp[(\mathbf{K} + \mathbf{V}) \zeta]$ and $\mathbf{U}_-(\zeta) = \exp[(\mathbf{K} - \mathbf{V}) \zeta]$.

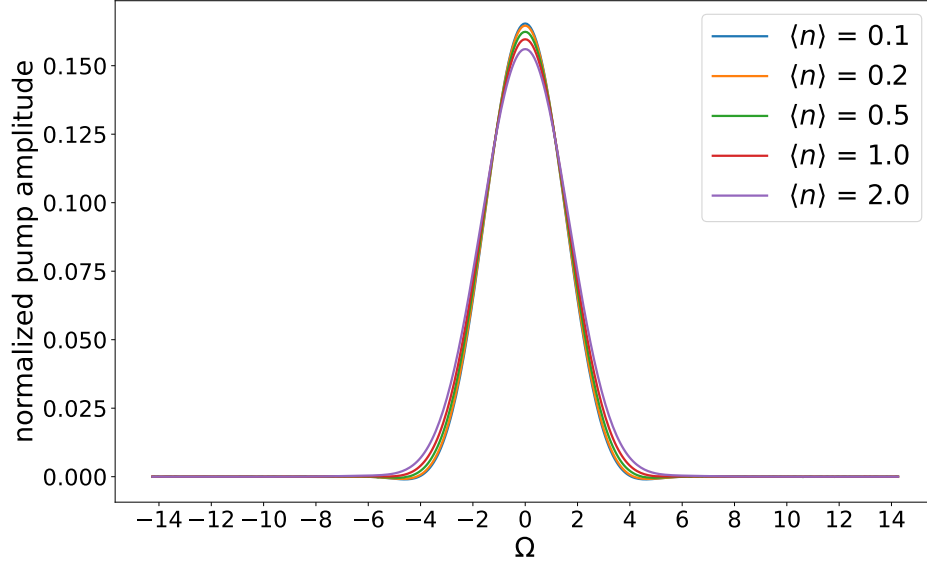


Figure 3.3 Real part of the optimal pumps obtained with random initialization seeds for different $\langle n \rangle$. Inequality on the variance implied by a regularization is added as well as a penalty on the expected value of the frequency distribution of the pump. Since the imaginary part is still proportional to the real part, we are again in the frame of reference where the imaginary part is negligible, hence why it was not shown. Note that it is very similar to Fig. 3.1.

3.3 Optimizing Over Gaussians

All previous optimal pumps look very similar to Gaussians, so to verify, instead of optimizing over every frequency pixel of the pump, the optimization scheme takes into account that the pump is always Gaussian and minimizes over its width and amplitude. Fig. 3.5 shows the optimal Gaussian when the equations of motion are decoupled for different values of $\langle n \rangle$. Again, the results look very similar to previous ones.

Defining the difference between vectors as the L2 norm $\sqrt{\frac{(\bar{v}-\bar{w})^\dagger(\bar{v}-\bar{w})}{\bar{v}^\dagger\bar{v}}}$, Table 3.1 shows the difference between the optimal Gaussian pump and the general optimal pump in Fig. 3.4 for different values of $\langle n \rangle$. Since the difference between the optimal pump obtained from a random initialization and the optimal Gaussian is very small, it seems the optimal pump from numerical optimization is still Gaussian even when not in the perturbative regime.

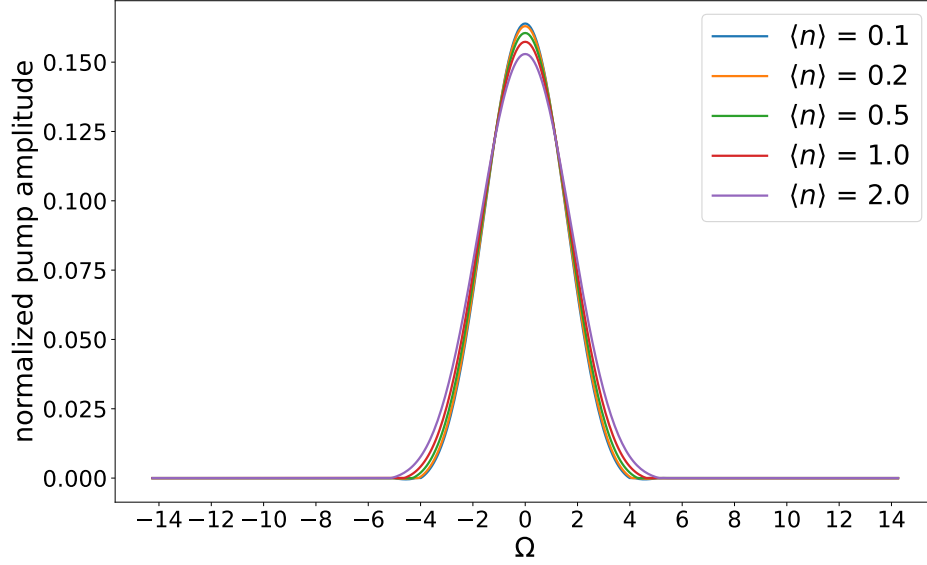


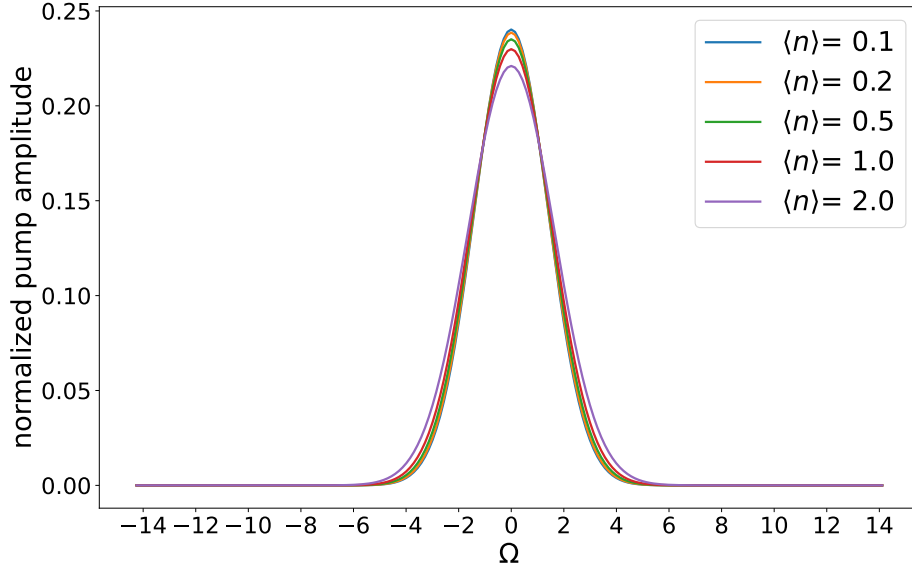
Figure 3.4 Normalized optimal pumps when optimizing over the pseudo-quadratures. Penalty on the frequency distribution is made such that the center frequency of the pump is zero, and regularization on the variance is also added. The results look very similar to the real part obtained in Fig. 3.3.

Table 3.1: Table showing the difference between the optimal pump and the optimal Gaussian pump. Since the value is less than 10^{-3} , it is reasonable to say the numerically optimal pump is always Gaussian.

Gain	Difference
0.1	8.74×10^{-4}
0.2	7.52×10^{-4}
0.5	6.47×10^{-4}
1.0	3.52×10^{-4}
2.0	9.37×10^{-5}

3.4 Convex Relaxation

Numerical optimization only gives local optima, not global ones. While results obtained with random seed are promising, there is no assurance of a global minimum. Then, one would wonder if there can exist a pump that can perform better than the Gaussians obtained numerically, say one that can make the matrix $\exp((\mathbf{K} + \mathbf{V})\zeta) - \exp((\mathbf{K} - \mathbf{V})\zeta)$ as low-rank as possible. Using the formalism in Sec. 2.7, while a non-trivial bound value does not seem reachable, feasibility results can still, however, be obtained. In the next section, the



$\langle n \rangle$	K
0.0001	1.218
0.1	1.207
0.2	1.197
0.5	1.172
1.0	1.143
2.0	1.109

Figure 3.5 Normalized optimal Gaussian pump for different values of $\langle n \rangle$. The width increases with the increase in gain. This may be a consequence of the spectral widening of signal and idler beams observed in experiments [1]. Note that the plot looks very similar to that of Fig. 3.4. The table below shows the Schmidt number evaluated at different values of $\langle n \rangle$ for optimal Gaussians. They are very close to the Schmidt number obtained when every discretization in frequency is allowed to vary freely. Then, we can infer that any discontinuity or negative values from prior optimization schemes may simply be a numerical artifact.

pseudo-quadrature formalism where the equations of motion are decoupled will be used.

3.4.1 Low-gain regime lower bounds with semidefinite relaxation

To solve the semidefinite relaxation of minimizing the spectral separability of signal and idler states under the constraints given in Sec. 2.7, MOSEK [76], an interior-point based semidefinite program solver, will be used through CVXPY [77], a Python-based convex optimization

language.

First, the constraints in Eq. (2.134), (2.135), (2.138), (2.148a), (2.148b), (2.142), (2.150) are imposed on the system with $\mathbf{R}_i$ being the elements of a matrix basis where only a single element is 1 and others 0. The power $|V|$ of the pump as defined in Eq. (2.148b) is taken to be the value needed for the low-gain optimal pump to obtain $\langle n \rangle = 10^{-4}$, which in this case is $|V| = 0.02531$. With Ω ranging from -10 to 10 , ζ ranging from 0 to 1 , a frequency window of 21 by 21 ($N_\omega = 21$) and a discretization of ζ of 6 ($N_z = 6$), the minimum of the objective function as defined in Sec. 2.7 is 0 . When mentioning $|V|$ that allows the low-gain pump to output $\langle n \rangle$, it signifies that $|V|$ is the value needed to multiply the matrix $\mathbf{V}$ obtained from the optimal low-gain pump, scaled such that $\text{Tr}[\mathbf{V}\mathbf{V}^\dagger] = 1$, to give $\langle n \rangle$ mean number of photon pairs per pulse. The reason behind the small frequency window and the low value of N_z is that semidefinite programs are quite costly to solve. For the program as defined by the given values of N_ω and N_z , approximately 80 gigabytes of RAM are occupied at maximum usage.

While disappointing, this result is not that surprising, considering that while the N_ω largest eigenvalues are big, as shown in Fig. 3.6, the rest of the eigenvalues are still not negligible. An explanation behind the zero value can be that the physics of the problem, or the relation between the linear part representing $\mathbf{U}_{+N_z}, \mathbf{U}_{-N_z}$ and $\mathbf{V}$, is not well respected by the semidefinite relaxation. Instead of establishing $\mathbf{U}_{\pm N_z} = \exp[(\mathbf{K} \pm \mathbf{V})\zeta_{N_z}]$, some other relations are being imposed. Upon further inspection, it can be seen that the block representing $\mathbf{U}_{+N_z}$ has the same values as the block representing $\mathbf{U}_{-N_z}$.

More constraints need to be added for the semidefinite program to make a distinction between $\mathbf{U}_{+N_z}$ and $\mathbf{U}_{-N_z}$ since the constraints regarding the dynamics and commutation relations are not strong enough. The need for more constraints can be illustrated with Fig. 3.7. It offers an intuitive picture of why more convex constraints may be needed for a better representation of the initial non-convex problem with 2 variables (x_1, x_2) . The convex envelope is the smallest convex set that contains the entire non-convex set. An SDR feasible point may give a lower objective function than any point in the non-convex feasible region, but it also may not be feasible in the context of the original problem. By cutting down the feasibility region, such a point may become outside of the newly restricted feasible set, increasing the minimum of the convex relaxation. Since there is no guarantee that the semidefinite relaxation is the convex envelope of the original problem, adding constraints would help cut down the feasible

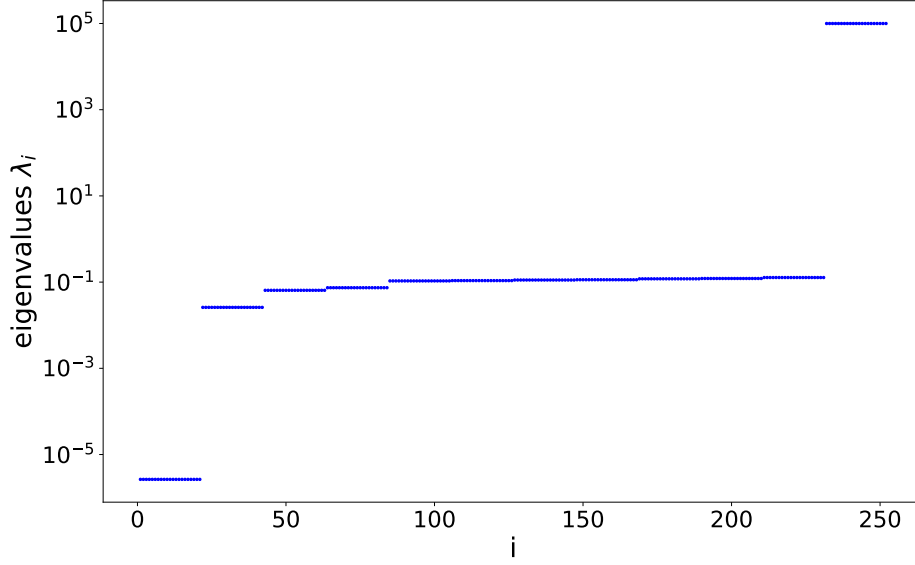


Figure 3.6 Eigenvalues of the optimal solution of the semidefinite relaxation in perturbative regime, where λ_i represents the i th eigenvalue. The eigenvalues are classified in increasing order, meaning that $\lambda_i \geq \lambda_j$ if $i \geq j$. The N_ω largest eigenvalues are on a scale of 10^5 , but the rest of the eigenvalues are not negligible, so the semidefinite matrix cannot be said to be of rank N_ω .

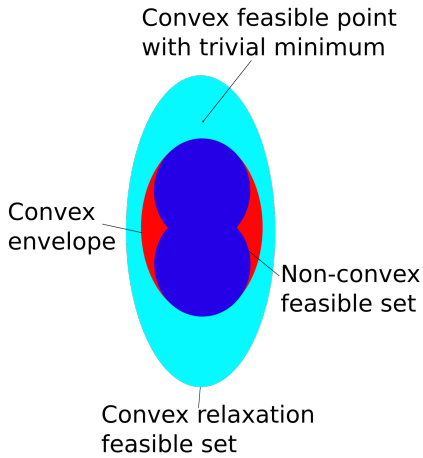


Figure 3.7 Representation of adding more convex constraints to convex relaxation with 2 variables. When more constraints are added, the convex feasible set will approximate that of the nonconvex set. Feasible points from the old relaxation with a suboptimal minimum will then be excluded from the new, tighter convex feasible set.

region so it becomes closer to the envelope.

Constraints on the nuclear norm of $\mathbf{U}_{+N_z}$ and $\mathbf{U}_{-N_z}$ can be added to the problem to make a distinction between both matrices. Assume, for quadratically feasible propagators, meaning propagators from the feasible region of the quadratic matrix program, if d_{+i} are the singular values of $\frac{1}{\sqrt{\langle n \rangle}} \mathbf{U}_{+N_z}$ and d_{-i} are the singular values of $\frac{1}{\sqrt{\langle n \rangle}} \mathbf{U}_{-N_z}$, then the relation

$$|d_{+i} - d_{-i}| \leq 2 \quad (3.6)$$

must be true. The definition of $d_{\pm i}$ are the normalized singular values for $\mathbf{U}_{\pm N_z}$. Because

$$\sqrt{\langle n \rangle} = \frac{1}{2} \sqrt{\langle n \rangle \operatorname{Tr} \left[\frac{1}{\langle n \rangle} (\mathbf{U}_{+N_z} - \mathbf{U}_{-N_z}) (\mathbf{U}_{+N_z} - \mathbf{U}_{-N_z})^\dagger \right]} = \frac{1}{2} \sqrt{\langle n \rangle \sum_i (d_{+i} - d_{-i})^2}, \quad (3.7)$$

then $\sum_i (d_{+i} - d_{-i})^2 = 4$, leading to the inequality $(d_{+i} - d_{-i})^2 \leq 4$. In fact, if $\frac{1}{\sqrt{\langle n \rangle}} (\mathbf{U}_{+N_z} - \mathbf{U}_{-N_z})$ has j significant singular values, then

$$|d_{+i} - d_{-i}| \leq 2 \quad \forall i \leq j, \quad (3.8)$$

$$|d_{+i} - d_{-i}| = 0 \quad \forall i > j. \quad (3.9)$$

Taking into account $d_{+i} = \frac{1}{\langle n \rangle d_{-i}}$ from the commutation relations, assuming only the first two singular values are significant,

$$2 - \epsilon \leq d_{+1} - d_{-1} \leq 2, \quad (3.10a)$$

$$-\sigma \leq d_{+2} - d_{-2} \leq \sigma, \quad (3.10b)$$

$$d_{+i} - d_{-i} = 0 \quad \forall i > 2, \quad (3.10c)$$

where it is taken into account that the largest singular value of $\frac{1}{\sqrt{\langle n \rangle}} \mathbf{U}_{+N_z}$ is always greater or equal to the singular value of $\frac{1}{\sqrt{\langle n \rangle}} \mathbf{U}_{-N_z}$ with the same index when the difference between

these matrices is low-rank. The inequalities on d_{+1} and d_{-1} can lead to

$$2 - \epsilon \leq d_{+1} - \frac{1}{d_{+1} \langle n \rangle} \leq 2, \quad (3.11a)$$

$$d_{+1}^2 - \frac{1}{\langle n \rangle} - 2d_{+1} \leq 0, \quad (3.11b)$$

$$d_{+1} \leq 1 + \sqrt{1 + \frac{1}{\langle n \rangle}}, \quad (3.11c)$$

$$2 - \epsilon \leq \frac{1}{\langle n \rangle d_{-1}} - d_{-1}, \quad (3.11d)$$

$$d_{-1}^2 + (2 - \epsilon) d_{-1} - \frac{1}{\langle n \rangle} \leq 0, \quad (3.11e)$$

$$d_{-1} \leq -\frac{2 - \epsilon}{2} + \sqrt{\left(\frac{2 - \epsilon}{2}\right)^2 + \frac{1}{\langle n \rangle}}, \quad (3.11f)$$

with $0 \leq \epsilon \leq 2$. The inequalities for the second singular value can be seen as

$$d_{-2} \leq \frac{\sigma}{2} + \sqrt{\frac{\sigma}{2} + \frac{1}{\langle n \rangle}}, \quad (3.12a)$$

$$d_{+2} \leq \frac{\sigma}{2} + \sqrt{\frac{\sigma}{2} + \frac{1}{\langle n \rangle}}, \quad (3.12b)$$

with $0 \leq \sigma \leq 2$. Assuming that the matrix $\mathbf{U}_{+N_z} - \mathbf{U}_{-N_z}$ is low-rank enough, then it is not too restrictive to say that $|d_{+i} - d_{-i}|$ will be insignificant for $i \geq 2$. Then, constraints

$$\sum_i d_{+i} \leq 1 + \sqrt{1 + \frac{1}{\langle n \rangle}} + \frac{\sigma}{2} + \sqrt{\left(\frac{\sigma}{2}\right)^2 + \frac{\langle n \rangle}{2}} + \frac{(N_\omega - 2)}{\sqrt{\langle n \rangle}}, \quad (3.13)$$

$$\sum_i d_{-i} \leq -\frac{2 - \epsilon}{2} + \sqrt{\left(\frac{2 - \epsilon}{2}\right)^2 + \frac{1}{\langle n \rangle}} + \frac{\sigma}{2} + \sqrt{\left(\frac{\sigma}{2}\right)^2 + \frac{\langle n \rangle}{2}} + \frac{(N_\omega - 2)}{\sqrt{\langle n \rangle}} \quad (3.14)$$

can be added to the program in an attempt to relate the commutation constraints as well as the mean photon constraints, which are quadratic, to the linear parts of the propagators. While propagators respecting Ineqs. (3.13) and (3.14) do not necessarily respect Ineqs. (3.11c), (3.11f) and (3.12), the reverse is true, meaning for small ϵ and σ , propagators with rank close to 1 always respect Ineqs. (3.13) and (3.14). For example, in the case where the pump is allowed to be decentered, .

The constraints from Ineq. (3.13) and (3.14) bring forth a few ideas:

1. Any rank one or close to rank one $\frac{1}{\sqrt{\langle n \rangle}} (\mathbf{U}_{+N_z} - \mathbf{U}_{-N_z})$ respects Ineqs. (3.13) and (3.14).
2. The relations $\mathbf{U}_{+N_z} \mathbf{U}_{-N_z}^\dagger = \mathbf{I}$ and $\frac{1}{4} \text{Tr} [\mathbf{U}_{+N_z} \mathbf{U}_{+N}^\dagger + \mathbf{U}_{-N_z} \mathbf{U}_{-N}^\dagger - 2\mathbf{I}] - \langle n \rangle = 0$ are attempted to be enforced on the linear parts. Attempted because the constraints are not given on individual singular values, instead on the sum of them.
3. They enforce an infeasibility constraint. Any infeasible point for the semidefinite relaxation will be infeasible for the original quadratically constrained program. If for some values of ϵ and σ , the semidefinite problem is infeasible, then the original problem is not feasible either.

These constraints on $\mathbf{U}_{+N_z}$ and $\mathbf{U}_{-N_z}$ immediately make the bound better. Using the same parameters that gave a bound of 0, fixing $\epsilon = \sigma = 0.1$, the lower bound becomes 9.77×10^{-1} instead. The value is lower than but very close to 1, which is a bound well aligned with the numerical optimization results when there are no constraints on centering the pump, as there exists a pump, while maybe physically non-viable, that gives perfectly separable idler and signal states.

While for numerical optimization, regularization on the variance of the pump divided by the sum of its absolute values is used, here the pump is decomposed into a Hermite basis instead, which is to say

$$\beta(\Omega) = \sum_n^M \frac{c_n}{\sqrt{2^n n!} \sqrt{\pi}} H_n(\Omega \tau) \exp \left[-\frac{\Omega^2 \tau^2}{2} \right], \quad (3.15)$$

where $\beta(\Omega)$ is the function in frequency of the pump (the first and last row of $\mathbf{V}$), $H_n(\Omega)$ are Hermite polynomials, and τ is related to the width of elements of the basis. For a small enough M , Eq. (3.15) limits the feasible pumps to be centered enough such that any pumps that have large amplitudes too far from the center frequency are excluded. With the same parameters and with Hermite basis constraints with τ^2 chosen as the value of the optimal low-gain Gaussian pump ($\frac{\tau^2}{2} \approx \frac{1}{4}$) and $M = 10$, the lower bound is 1.22. This means there is no pump whose amplitude is located around the central frequency that can give perfect (or even almost perfect) spectral separability for $\langle n \rangle \ll 1$. While a clear relation between the Schmidt number and the normalized nuclear norm is difficult to give, an idea of the scale of the value of the Schmidt number given by a weighted nuclear norm of 1.43 can be given by keeping in mind that with a Schmidt number of 1.43 for the perturbative optimal pump, the weighted nuclear norm is approximately 1.55.

3.4.2 High-gain regime lower bounds with semidefinite relaxation

The observations for the perturbative regime do not hold true for the high-gain regime. Table 3.2 shows how, with increasing $|V|$, the objective function tends to zero for $\langle n \rangle = 1$, where Inf means the bound is ∞ , or the problem is infeasible. The nuclear norm and Hermite basis constraints are imposed with the same values of ϵ , σ , M and τ as in Sec. 3.4.1. The value of $|V|$ needed for the low-gain optimal pump to obtain $\langle n \rangle = 1$ is $|V| \approx 2.145$, giving the scale of $|V|$ needed for $\langle n \rangle = 1$. However, for even lower $|V|$, the optimal value of the objective function drops towards zero. It is possible that the higher is $|V|$, the less the semidefinite program respects the physics. An interesting observation is that adding the constraint that

Table 3.2: Lower bounds on the objective function for different values of $|V|$ when $\langle n \rangle = 1$ with Hermite basis constraint and without Hermite basis constraint. When $|V| \geq 1.0$, the bound is driven to 0. However, there exists a quadratically feasible pump that can give a mean number of photon pairs per pulse of 1 for $|V| \geq 1.0$, meaning the semidefinite heuristics used in low-gain are not effective in high-gain. For some values of $|V|$, only setting the pump to be a superposition of Hermite basis elements can make the problem infeasible with constraints on the nuclear norm of propagators.

$ V $	Lower bound on obj. f.	Lower bound on obj. f. with Hermite basis constraints
0.0001	Inf	Inf
0.1	Inf	Inf
0.2	3.67×10^{-1}	Inf
0.25	2.98×10^{-1}	Inf
0.3	2.50×10^{-1}	2.50×10^{-1}
0.4	1.60×10^{-1}	1.60×10^{-1}
0.5	6.95×10^{-2}	6.95×10^{-2}
1.0	4.46×10^{-8}	1.31×10^{-7}
1.5	9.47×10^{-9}	1.05×10^{-8}

the pump is made of Hermite basis elements gives an infeasible solution for some values of $|V|$. Since these constraints do not affect the physics, only the pump, it may be possible that a pump that is the superposition of $M = 10$ hermite modes of width $\tau \approx \frac{1}{\sqrt{2}}$ may never give an objective function very close to 1. However, there is no certainty to this statement since the problem is not infeasible for all values of $|V|$.

A reason why the semidefinite relaxation is not giving good bounds may be that the quadratic term from the description of dynamics in Eq. (2.148b) is not well-defined. In fact, for $\mathbf{U}_{+N_z}$,

the quadratic part is

$$\begin{bmatrix} \mathbf{0} & \mathbf{0} & \dots & \frac{|V|\Delta\zeta}{2} \exp[\mathbf{K}\zeta_{N_z-1}] \mathbf{R}_i \\ \mathbf{0} & \mathbf{0} & \dots & \frac{|V|\Delta\zeta}{2} \exp[\mathbf{K}\zeta_{N_z-2}] \mathbf{R}_i \\ \vdots & \vdots & \ddots & \vdots \\ \frac{|V|\Delta\zeta}{2} \mathbf{R}_i^\dagger \exp[-\mathbf{K}\zeta_{N_z-1}] & \frac{|V|\Delta\zeta}{2} \mathbf{R}_i^\dagger \exp[-\mathbf{K}\zeta_{N_z-2}] & \dots & \mathbf{0} \end{bmatrix}. \quad (3.16)$$

This matrix only has off-diagonal elements that are significant. This is very different from the quadratic matrix governing the dynamics of the electromagnetic field [78], where hidden convexity is observed for some problems. For electromagnetic fields, the amplitude of the quadratic matrix is highly diagonal with decaying off-diagonal elements. Attempts at adding terms on the diagonal to the dynamics can be made with finite differences.

For higher $\langle n \rangle$ values at higher pump power, because of the observations made previously, more constraints relating the linear parts of the propagators and the pump together as well as constraints relating the $\mathbf{V}\mathbf{U}_{\pm N}^\dagger$ parts to $\mathbf{U}_{\pm i}\mathbf{U}_{\pm j}^\dagger$ parts need to be imposed, as well as constraints relating linear parts of $\mathbf{V}$, $\mathbf{U}_{+i}$ and $\mathbf{U}_{-i}$. It is known that a first order derivative can be approximated by finite differences of higher accuracy orders [79]. Let a function $f(\zeta_i) = \exp[(\mathbf{K} + \mathbf{V})\zeta_i]$ for $\zeta_i \in [-N_z\Delta\zeta, N_z\Delta\zeta]$ where $\Delta\zeta$ is the size of discretization, with the propagators defined as

$$\mathbf{U}_{\pm i} = \exp[(\mathbf{K} \pm \mathbf{V})\zeta_i] \quad \forall \quad i \in [0, N_z], \quad (3.17)$$

$$\mathbf{U}_{\pm i}^\dagger = \exp[(-\mathbf{K} \pm \mathbf{V})\zeta_i] \quad \forall \quad i \in [0, N_z], \quad (3.18)$$

for $f(\zeta)$ evaluated at $\zeta \leq 0$, it can be said that

$$f(-\zeta_i) = \mathbf{U}_{-i}^\dagger, \quad (3.19a)$$

$$f(\zeta_i) = \mathbf{U}_{+i}. \quad (3.19b)$$

Then, knowing the first derivative of $f(\zeta)$ is $\partial_z f(\zeta) = (\mathbf{K} + \mathbf{V})f(\zeta)$, evaluating the derivative when $\zeta_i = 0$, the derivative can be written as

$$\partial_\zeta f(\zeta)|_{\zeta=0} = \mathbf{K} + \mathbf{V}. \quad (3.20)$$

It can be said that, for N_z big enough and $\Delta\zeta$ small enough,

$$\mathbf{K} + \mathbf{V} \approx \frac{1}{\Delta\zeta} \sum_{i=1}^{N_z} c_i (f(\zeta_i) - f(-\zeta_i)), \quad (3.21)$$

where c_i are central finite difference coefficients. With Eqs. (3.19), the relation

$$\mathbf{K} + \mathbf{V} \approx \frac{1}{\Delta\zeta} \sum_{i=1}^{N_z} c_i (\mathbf{U}_{+i} - \mathbf{U}_{-i}^\dagger) \quad (3.22)$$

can be established. For a backwards finite difference, it can be seen that

$$\mathbf{K} \pm \mathbf{V} \approx \frac{1}{\Delta\zeta} \sum_{i=0}^{N_z} d_i \mathbf{U}_{\pm i}, \quad (3.23)$$

or

$$-\mathbf{K} \pm \mathbf{V} \approx \frac{1}{\Delta\zeta} \sum_{i=0}^{N_z} d_i \mathbf{U}_{\pm i}^\dagger. \quad (3.24)$$

Assume the approximation is good enough, then

$$(-\mathbf{K} \pm \mathbf{V}) \mathbf{U}_{\pm j}^\dagger \approx \frac{1}{\Delta\zeta} \sum_{i=0}^{N_z} d_i \mathbf{U}_{\pm i} \mathbf{U}_{\pm j}^\dagger \quad \forall j \leq N_z. \quad (3.25)$$

This way, since the term $\mathbf{V} \mathbf{U}_{\pm j}^\dagger$ is in the dynamics, a relation between $\mathbf{U}_{\pm j} \mathbf{U}_{\pm j}^\dagger$ and the dynamics is established. In other words, the SPDC gain should now influence how the dynamics evolve. The central finite difference gives a direct relationship between the linear parts of the propagators and that of the pump.

An issue with the finite differences is that with the parameters used (Ω and ζ range, discretization in frequency and in ζ , etc.), the numerical error on the approximation is too big. An attempt is still made with the same parameters, except setting ζ to range from 0 to 0.1. In this regime, the low-gain optimal pump with $\zeta \in [0, 1]$ can give $\langle n \rangle = 1$ with $|V| \approx 9.97$. Adding the constraints on finite differences and the nuclear norm inequalities, the bound on the objective function is still very close to 0. However, adding finite difference constraints does decrease the feasible region. Table 3.3 shows the bounds when varying the value of $|V|$ when the finite differences are implemented and when they are not for $\langle n \rangle = 1$ and ζ ranging from 0 to 0.1. While the minimum on the objective function seems to be higher, at values of $|V|$ that would give a quadratically feasible point, it is still too low, implying that while they

Table 3.3: Optimal value of objective function as a function of $|V|$ for when finite difference constraints are applied and when they are not. Here, the parameter ζ is between 0 and 0.1. When the finite difference constraints are respected, the optimal value does improve, but it still approaches 0 when $|V|$ is close to the amplitude for which the low-gain optimal pump would give $\langle n \rangle = 1$.

$ V $	Optimal obj. f. without finite difference	Optimal obj. f. with finite difference
2.0	3.69×10^{-1}	Inf
3.0	2.50×10^{-1}	3.42×10^{-1}
4.0	2.20×10^{-1}	3.27×10^{-1}
5.0	1.29×10^{-1}	2.54×10^{-1}
6.0	3.93×10^{-2}	1.32×10^{-1}
7.0	2.50×10^{-7}	1.16×10^{-1}
8.0	2.43×10^{-10}	6.02×10^{-7}

may help make the bounds better, they are not the needed constraints for acceptable bounds.

When setting $\langle n \rangle = 1$, Table 3.4 shows the value of $\text{Tr} [\mathbf{V}_{\text{opt}} \mathbf{V}_{\text{opt}}^\dagger]$ with increasing $|V|$ for $\mathbf{V}_{\text{opt}}$ being the linear part of the pump from the optimal semidefinite matrix. There may be a correlation between when the objective function decreases and when $\text{Tr} [\mathbf{V}_{\text{opt}} \mathbf{V}_{\text{opt}}^\dagger]$ decreases.

Table 3.4: Value of $\text{Tr} [\mathbf{V}_{\text{opt}} \mathbf{V}_{\text{opt}}^\dagger]$ as a function of $|V|$. While compared to Table 3.2, lower values of $\text{Tr} [\mathbf{V}_{\text{opt}} \mathbf{V}_{\text{opt}}^\dagger]$ are often associated with a lower minimum of objective function. But when $\text{Tr} [\mathbf{V}_{\text{opt}} \mathbf{V}_{\text{opt}}^\dagger]$ is closer to 1, the optimal value of objective function is indisputably higher than when $\text{Tr} [\mathbf{V}_{\text{opt}} \mathbf{V}_{\text{opt}}^\dagger]$ is much lesser than 1.

$ V $	$\text{Tr} [\mathbf{V}_{\text{opt}} \mathbf{V}_{\text{opt}}^\dagger]$
0.2	2.43×10^{-1}
0.25	1.74×10^{-2}
0.3	1.37×10^{-4}
0.4	1.01×10^{-4}
0.5	9.04×10^{-6}
1.0	1.03×10^{-5}

From Table 3.4, the constraint $\text{Tr} [\mathbf{V}_{\text{opt}} \mathbf{V}_{\text{opt}}^\dagger] = 1$ is never respected. This comes from how the rank of the semidefinite matrix is not low enough. The semidefinite program only enforces the constraint $\text{Tr} [\mathbf{V}_{\text{opt}} \mathbf{V}_{\text{opt}}^\dagger] \leq 1$. When adding affine constraints on $\mathbf{V}$ such that it becomes a fixed pump $\mathbf{V}_{\text{fix}}$ where $\text{Tr} [\mathbf{V}_{\text{fix}} \mathbf{V}_{\text{fix}}^\dagger] = 1$, the lower bound for different values of $\langle n \rangle$ is shown in Table 3.5. $\mathbf{V}_{\text{fix}}$ is the low-gain optimal Gaussian pump, and the $|V|$ used is the value

which would make $\mathbf{V}_{\text{fix}}$ give $\langle n \rangle$. The bounds seem to have a fixed error compared to the

Table 3.5: Bounds given by the optimization scheme when fixing the pump to be $\mathbf{V}_{\text{fix}}$, with $\mathbf{V}_{\text{fix}}$ the matrix given by the low-gain optimal pump normalized such that $\text{Tr} [\mathbf{V}_{\text{fix}} \mathbf{V}_{\text{fix}}^\dagger] = 1$. The value of objective function with matrix exponential is given by when $\mathbf{U}_{\pm N_z} = \exp[(\mathbf{K} + |V| \mathbf{V}_{\text{fix}}) \zeta_f]$ with $\zeta_f = 1$ and the objective function after optimization being the objective function obtained with SDR. It can be seen they are not very far off considering the numerical errors in approximating integration by summation.

Obj. f. with matrix exponential	Obj. f. after optimization	$\langle n \rangle$
1.57	1.31	0.0001
1.55	1.30	0.2
1.54	1.29	0.3
1.52	1.27	0.5
1.50	1.25	1.0

value obtained with matrix exponential, possibly stemming from the numerical error on the approximation of the integral on the dynamics by a discretized sum with details shown in Appendix D. Then, if one can constrain $\text{Tr} [\mathbf{V}_{\text{fix}} \mathbf{V}_{\text{fix}}^\dagger]$ to be as close to 1 as possible, a good bound may be obtained. In fact, the eigenvalues of the semidefinite matrix where the pump is constrained to be fixed are shown in Fig. 3.8 with $\langle n \rangle = 1$. The N_ω largest eigenvalues are the most prominent and, while the $N_\omega + 1$ th eigenvalue is large, the other eigenvalues decrease in amplitude quite quickly. The solution can be approximated as

$$\mathbf{Y} \approx \begin{bmatrix} \mathbf{X} \\ \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{X}^\dagger & \mathbf{I} \end{bmatrix}, \quad (3.26)$$

hence the reason why the bound is close to the theoretical value.

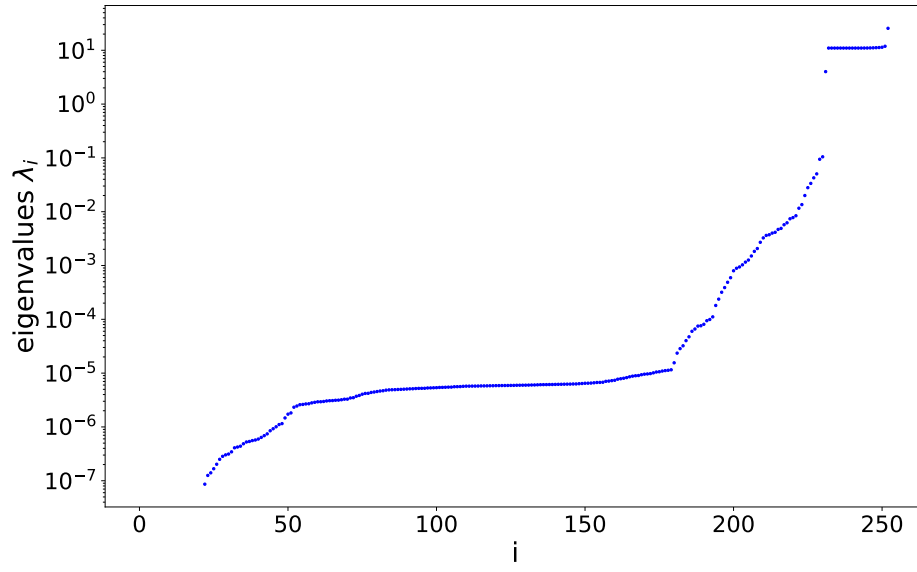


Figure 3.8 Eigenvalues of the semidefinite matrix when affine constraints on the linear part of the pump are applied to the pump in order to fix it to be $\mathbf{V}_{\text{fix}}$. λ_i is the i th eigenvalue. It is noteworthy that the smallest eigenvalues are smaller than those in Fig. 3.6, showing the solution is lower rank.

CHAPTER 4 CONCLUSION

4.1 Summary of Works

A literature review leading to the importance of pump design in high-gain SPDC has been established. How the perturbative results are obtained from the high-gain formalism has also been shown. A description of a scheme to decouple the equations of motion has been given. Then, details about numerical optimization heuristics and a convex relaxation of the spectral separability problem for photon pairs from the SPDC process have been shown. Results from numerical optimization showed even in high-gain regime, the spectral distribution minimizing the frequency entanglement between idler and signal states is Gaussian when in symmetric group velocity matching regime, only that the width increases with mean photon number pairs per pulse. A scheme to turn the dynamics of SPDC into quadratic equations is presented, bypassing the need for matrix exponentiation. The results from convex relaxation show that, in some cases, if the pump is made of superposition of Hermite basis elements with width equal to that of the low-gain pump, perfect separability cannot be achieved. However, these results cannot be generalized. The convex relaxation method of the spectral separability as presented in this project is then inconclusive.

4.2 Limitations

While numerical optimization gives pump shapes that minimize the spectral entanglement of the idler and signal states for high-gain SPDC through the use of BFGS at different mean photon pairs per pump pulse, there is still no guarantee there does not exist a spectral envelope that can do better. The convex relaxation is inconclusive for general pump power, so the method presented in this manuscript is not enough to find an acceptable lower bound.

4.3 Future Research

A few improvements can be made to the convex optimization model shown. An intuitive modification is to solve the dual of the quadratic matrix program with the nuclear norm objective function and the inequalities on the nuclear norm of the propagators instead of solving the semidefinite relaxation. Another possibility involves finding the dual of a problem with the same constraints, but instead of minimizing the nuclear norm, the spectral norm is maximized. Additional possible improvements include, with the possibility of loss of convexity,

adding constraints that are imposed on the linear parts of the pump to force its Frobenius norm to be 1 ($\text{Tr} [\mathbf{V}\mathbf{V}^\dagger] = 1$).

REFERENCES

- [1] K. Y. Spasibko, T. S. Iskhakov, and M. V. Chekhova, “Spectral properties of high-gain parametric down-conversion,” *Optics express*, vol. 20, no. 7, pp. 7507–7515, 2012.
- [2] M. A. Taylor and W. P. Bowen, “Quantum metrology and its application in biology,” *Physics Reports*, vol. 615, pp. 1–59, 2016.
- [3] H. J. Kimble, “The quantum internet,” *Nature*, vol. 453, no. 7198, pp. 1023–1030, 2008.
- [4] J. Shi and T. Byrnes, “Gaussian boson sampling with partial distinguishability,” *arXiv preprint arXiv:2105.09583*, 2021.
- [5] D. Grier, D. J. Brod, J. M. Arrazola, M. B. de Andrade Alonso, and N. Quesada, “The complexity of bipartite gaussian boson sampling,” *Quantum*, vol. 6, p. 863, 2022.
- [6] C. C. Gerry and P. L. Knight, *Introductory quantum optics*. Cambridge university press, 2023.
- [7] C.-K. Hong, Z.-Y. Ou, and L. Mandel, “Measurement of subpicosecond time intervals between two photons by interference,” *Physical review letters*, vol. 59, no. 18, p. 2044, 1987.
- [8] G. Buller and R. Collins, “Single-photon generation and detection,” *Measurement Science and Technology*, vol. 21, no. 1, p. 012002, 2009.
- [9] A. Sipahigil, M. L. Goldman, E. Togan, Y. Chu, M. Markham, D. J. Twitchen, A. S. Zibrov, A. Kubanek, and M. D. Lukin, “Quantum interference of single photons from remote nitrogen-vacancy centers in diamond,” *Physical review letters*, vol. 108, no. 14, p. 143601, 2012.
- [10] N. Somaschi, V. Giesz, L. De Santis, J. Loredó, M. P. Almeida, G. Hornecker, S. L. Portalupi, T. Grange, C. Anton, J. Demory *et al.*, “Near-optimal single-photon sources in the solid state,” *Nature Photonics*, vol. 10, no. 5, pp. 340–345, 2016.
- [11] R. W. Boyd, A. L. Gaeta, and E. Giese, “Nonlinear optics,” in *Springer Handbook of Atomic, Molecular, and Optical Physics*. Springer, 2008, pp. 1097–1110.
- [12] S. Castelletto and R. Scholten, “Heralded single photon sources: a route towards quantum communication technology and photon standards,” *The European Physical Journal-Applied Physics*, vol. 41, no. 3, pp. 181–194, 2008.

- [13] R. W. Boyd, “Order-of-magnitude estimates of the nonlinear optical susceptibility,” *journal of modern optics*, vol. 46, no. 3, pp. 367–378, 1999.
- [14] X. Guo, C.-l. Zou, C. Schuck, H. Jung, R. Cheng, and H. X. Tang, “Parametric down-conversion photon-pair source on a nanophotonic chip,” *Light: Science & Applications*, vol. 6, no. 5, pp. e16 249–e16 249, 2017.
- [15] Y. P. Varshni, “Temperature dependence of the energy gap in semiconductors,” *physica*, vol. 34, no. 1, pp. 149–154, 1967.
- [16] L. Helt and N. Quesada, “Degenerate squeezing in waveguides: a unified theoretical approach,” *Journal of Physics: Photonics*, vol. 2, no. 3, p. 035001, 2020.
- [17] C. Hong and L. Mandel, “Theory of parametric frequency down conversion of light,” *Physical Review A*, vol. 31, no. 4, p. 2409, 1985.
- [18] N. Quesada, L. Helt, M. Menotti, M. Liscidini, and J. Sipe, “Beyond photon pairs—nonlinear quantum photonics in the high-gain regime: a tutorial,” *Advances in Optics and Photonics*, vol. 14, no. 3, pp. 291–403, 2022.
- [19] F. Riexinger, M. Kutas, B. Haase, P. Bickert, D. Molter, M. Bortz, and G. von Freymann, “General simulation method for spontaneous parametric down-and parametric up-conversion experiments,” *SciPost Physics Core*, vol. 6, no. 1, p. 022, 2023.
- [20] Z. Yang, M. Liscidini, and J. E. Sipe, “Spontaneous parametric down-conversion in waveguides: a backward heisenberg picture approach,” *Physical Review A—Atomic, Molecular, and Optical Physics*, vol. 77, no. 3, p. 033808, 2008.
- [21] R.-B. Jin, W.-H. Cai, C. Ding, F. Mei, G.-W. Deng, R. Shimizu, and Q. Zhou, “Spectrally uncorrelated biphotons generated from “the family of bbo crystal”,” *Quantum Engineering*, vol. 2, no. 2, p. e38, 2020.
- [22] W. P. Grice, A. B. U’Ren, and I. A. Walmsley, “Eliminating frequency and space-time correlations in multiphoton states,” *Physical Review A*, vol. 64, no. 6, p. 063815, 2001.
- [23] N. Quesada and A. M. Brańczyk, “Gaussian functions are optimal for waveguided nonlinear-quantum-optical processes,” *Physical Review A*, vol. 98, no. 4, p. 043813, 2018.
- [24] F. Graffitti, J. Kelly-Massicotte, A. Fedrizzi, and A. M. Brańczyk, “Design considerations for high-purity heralded single-photon sources,” *Physical Review A*, vol. 98, no. 5, p. 053811, 2018.

- [25] N. Bruno, A. Martin, T. Guerreiro, B. Sanguinetti, and R. T. Thew, “Pulsed source of spectrally uncorrelated and indistinguishable photons at telecom wavelengths,” *Optics express*, vol. 22, no. 14, pp. 17 246–17 253, 2014.
- [26] M. M. Fejer, G. Magel, D. H. Jundt, and R. L. Byer, “Quasi-phase-matched second harmonic generation: tuning and tolerances,” *IEEE Journal of quantum electronics*, vol. 28, no. 11, pp. 2631–2654, 1992.
- [27] P. B. Dixon, J. H. Shapiro, and F. N. Wong, “Spectral engineering by gaussian phase-matching for quantum photonics,” *Optics express*, vol. 21, no. 5, pp. 5879–5890, 2013.
- [28] A. M. Brańczyk, A. Fedrizzi, T. M. Stace, T. C. Ralph, and A. G. White, “Engineered optical nonlinearity for quantum light sources,” *Optics express*, vol. 19, no. 1, pp. 55–65, 2011.
- [29] F. Graffitti, D. Kundys, D. T. Reid, A. M. Brańczyk, and A. Fedrizzi, “Pure down-conversion photons through sub-coherence-length domain engineering,” *Quantum Science and Technology*, vol. 2, no. 3, p. 035001, 2017.
- [30] S. Poveda-Hospital, Y.-A. Peter, and N. Quesada, “Custom nonlinearity profile for integrated quantum light sources,” *Physical Review Applied*, vol. 19, no. 5, p. 054033, 2023.
- [31] E. Meyer-Scott, N. Montaut, J. Tiedau, L. Sansoni, H. Herrmann, T. J. Bartley, and C. Silberhorn, “Limits on the heralding efficiencies and spectral purities of spectrally filtered single photons from photon-pair sources,” *Physical Review A*, vol. 95, no. 6, p. 061803, 2017.
- [32] C. Drago and A. M. Brańczyk, “Hong–ou–mandel interference: a spectral–temporal analysis,” *Canadian Journal of Physics*, 2024.
- [33] P. J. Mosley, J. S. Lundeen, B. J. Smith, P. Wasylczyk, A. B. U’Ren, . f. C. Silberhorn, and I. A. Walmsley, “Heralded generation of ultrafast single photons in pure quantum states,” *Physical Review Letters*, vol. 100, no. 13, p. 133601, 2008.
- [34] J. Eberly, “Schmidt analysis of pure-state entanglement,” *Laser physics*, vol. 16, pp. 921–926, 2006.
- [35] C. Law, I. A. Walmsley, and J. Eberly, “Continuous frequency entanglement: effective finite hilbert space and entropy control,” *Physical Review Letters*, vol. 84, no. 23, p. 5304, 2000.

- [36] R. Grobe, K. Rzazewski, and J. Eberly, “Measure of electron-electron correlation in atomic physics,” *Journal of Physics B: Atomic, Molecular and Optical Physics*, vol. 27, no. 16, p. L503, 1994.
- [37] A. Christ, K. Laiho, A. Eckstein, K. N. Cassemiro, and C. Silberhorn, “Probing multi-mode squeezing with correlation functions,” *New Journal of Physics*, vol. 13, no. 3, p. 033027, 2011.
- [38] A. Christ and C. Silberhorn, “Limits on the deterministic creation of pure single-photon states using parametric down-conversion,” *Physical Review A*, vol. 85, no. 2, p. 023829, 2012.
- [39] M. Barbieri, T. Weinhold, B. Lanyon, A. Gilchrist, K. Resch, M. Almeida, and A. White, “Parametric downconversion and optical quantum gates: two’s company, four’s a crowd,” *Journal of Modern Optics*, vol. 56, no. 2-3, pp. 209–214, 2009.
- [40] J. F. Bulmer, S. Paesani, R. S. Chadwick, and N. Quesada, “Threshold detection statistics of bosonic states,” *Physical Review A*, vol. 106, no. 4, p. 043712, 2022.
- [41] S. Sempere-Llagostera, G. Thekkadath, R. Patel, W. Kolthammer, and I. Walmsley, “Reducing $g(2)(0)$ of a parametric down-conversion source via photon-number resolution with superconducting nanowire detectors,” *Optics Express*, vol. 30, no. 2, pp. 3138–3147, 2022.
- [42] N. Quesada and J. Sipe, “Effects of time ordering in quantum nonlinear optics,” *Physical Review A*, vol. 90, no. 6, p. 063840, 2014.
- [43] —, “Time-ordering effects in the generation of entangled photons using nonlinear optical processes,” *Physical review letters*, vol. 114, no. 9, p. 093903, 2015.
- [44] P. Sharapova, G. Frascella, M. Riabinin, A. Pérez, O. Tikhonova, S. Lemieux, R. Boyd, G. Leuchs, and M. Chekhova, “Properties of bright squeezed vacuum at increasing brightness,” *Physical Review Research*, vol. 2, no. 1, p. 013371, 2020.
- [45] M. Houde and N. Quesada, “Waveguided sources of consistent, single-temporal-mode squeezed light: The good, the bad, and the ugly,” *AVS Quantum Science*, vol. 5, no. 1, 2023.
- [46] D. Liu, Y. Tan, E. Khoram, and Z. Yu, “Training deep neural networks for the inverse design of nanophotonic structures,” *Acs Photonics*, vol. 5, no. 4, pp. 1365–1369, 2018.

- [47] P. Chao, B. Strekha, R. Kuate Defo, S. Molesky, and A. W. Rodriguez, “Physical limits in electromagnetism,” *Nature Reviews Physics*, vol. 4, no. 8, pp. 543–559, 2022.
- [48] Z. Liu, D. Zhu, L. Raju, and W. Cai, “Tackling photonic inverse design with machine learning,” *Advanced Science*, vol. 8, no. 5, p. 2002923, 2021.
- [49] A. Hakansson, J. Sánchez-Dehesa, and L. Sanchis, “Inverse design of photonic crystal devices,” *IEEE Journal on selected areas in communications*, vol. 23, no. 7, pp. 1365–1371, 2005.
- [50] S. Goel, S. Leedumrongwatthanakun, N. H. Valencia, W. McCutcheon, A. Tavakoli, C. Conti, P. W. Pinkse, and M. Malik, “Inverse design of high-dimensional quantum optical circuits in a complex medium,” *Nature Physics*, pp. 1–8, 2024.
- [51] D. Jang, S. Kim, and J. Kim, “Deep-learning-based inverse design of colloidal quantum dots,” *Optics Communications*, p. 130384, 2024.
- [52] E. Rozenberg, A. Karnieli, O. Yesharim, J. Foley-Comer, S. Trajtenberg-Mills, D. Freedman, A. M. Bronstein, and A. Arie, “Inverse design of spontaneous parametric down-conversion for generation of high-dimensional qudits,” *Optica*, vol. 9, no. 6, pp. 602–615, 2022.
- [53] M. Houde and N. Quesada, “Perfect pulsed inline twin-beam squeezers,” *AVS Quantum Science*, vol. 6, no. 2, 2024.
- [54] G. Triginer, M. D. Vidrighin, N. Quesada, A. Eckstein, M. Moore, W. S. Kolthammer, J. Sipe, and I. A. Walmsley, “Understanding high-gain twin-beam sources using cascaded stimulated emission,” *Physical Review X*, vol. 10, no. 3, p. 031063, 2020.
- [55] N. Quesada, G. Triginer, M. D. Vidrighin, and J. Sipe, “Theory of high-gain twin-beam generation in waveguides: From maxwell’s equations to efficient simulation,” *Physical Review A*, vol. 102, no. 3, p. 033519, 2020.
- [56] K. Zielnicki, K. Garay-Palmett, D. Cruz-Delgado, H. Cruz-Ramirez, M. F. O’Boyle, B. Fang, V. O. Lorenz, A. B. U’Ren, and P. G. Kwiat, “Joint spectral characterization of photon-pair sources,” *Journal of Modern Optics*, vol. 65, no. 10, pp. 1141–1160, 2018.
- [57] M. Manngård, J. M. Böling, and H. T. Toivonen, “Subspace identification for mimo systems in the presence of trends and outliers,” in *Computer Aided Chemical Engineering*. Elsevier, 2017, vol. 40, pp. 307–312.

- [58] P. Wolfe, “Convergence conditions for ascent methods,” *SIAM review*, vol. 11, no. 2, pp. 226–235, 1969.
- [59] J. C. Meza, “Steepest descent,” *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 6, pp. 719–722, 2010.
- [60] J. Dennis and R. Schnabel, “Globally convergent modifications of newton’s method,” *Philadelphia, PA, USA: SIAM*, pp. 111–154, 1996.
- [61] Y.-H. Dai, “Convergence properties of the bfgs algorithm,” *SIAM Journal on Optimization*, vol. 13, no. 3, pp. 693–701, 2002.
- [62] R. M. Freund, “Penalty and barrier methods for constrained optimization,” *Lecture Notes, Massachusetts Institute of Technology*, 2004.
- [63] A. B.-T. . A. Nemirovski, “Optimization iii convex analysis,” Spring 2013.
- [64] W. J. Fu, “Penalized regressions: the bridge versus the lasso,” *Journal of computational and graphical statistics*, vol. 7, no. 3, pp. 397–416, 1998.
- [65] A. P. Ruszczyński, *Nonlinear optimization*. Princeton university press, 2006, vol. 13.
- [66] S. Boyd and L. Vandenberghe, *Convex Optimization*, 7th ed. Cambridge: Cambridge University Press, 2009.
- [67] J. J. Sakurai and J. Napolitano, *Modern quantum mechanics*. Cambridge University Press, 2020.
- [68] J. Park and S. Boyd, “General heuristics for nonconvex quadratically constrained quadratic programming,” *arXiv preprint arXiv:1703.07870*, 2017.
- [69] F. Zhang, *The Schur complement and its applications*. Springer Science & Business Media, 2006, vol. 4.
- [70] L. Vandenberghe and S. Boyd, “Semidefinite programming,” *SIAM review*, vol. 38, no. 1, pp. 49–95, 1996.
- [71] A. Beck, “Quadratic matrix programming,” *SIAM Journal on Optimization*, vol. 17, no. 4, pp. 1224–1238, 2007.
- [72] J. Peng, T. Zhu, H. Luo, and K.-C. Toh, “Semi-definite programming relaxation of quadratic assignment problems based on nonredundant matrix splitting,” *Computational Optimization and Applications*, vol. 60, pp. 171–198, 2015.

- [73] R. Fletcher, *Practical methods of optimization*. John Wiley & Sons, 2000.
- [74] M. Blondel, Q. Berthet, M. Cuturi, R. Frostig, S. Hoyer, F. Llinares-López, F. Pedregosa, and J.-P. Vert, “Efficient and modular implicit differentiation,” *arXiv preprint arXiv:2105.15183*, 2021.
- [75] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, “JAX: composable transformations of Python+NumPy programs,” 2018. [Online]. Available: <http://github.com/google/jax>
- [76] M. ApS, *MOSEK Optimizer API for Python 10.2.1*, 2019. [Online]. Available: <https://docs.mosek.com/latest/pythonapi/index.html>
- [77] S. Diamond and S. Boyd, “CVXPY: A Python-embedded modeling language for convex optimization,” *Journal of Machine Learning Research*, vol. 17, no. 83, pp. 1–5, 2016.
- [78] W. Xue, H. Zhang, A. Gopal, V. Rokhlin, and O. D. Miller, “Fullwave design of cm-scale cylindrical metasurfaces via fast direct solvers,” *arXiv preprint arXiv:2308.08569*, 2023.
- [79] B. Fornberg, “Generation of finite difference formulas on arbitrarily spaced grids,” *Mathematics of computation*, vol. 51, no. 184, pp. 699–706, 1988.

APPENDIX A PASSAGE FROM TIME DEPENDENT HEISENBERG EQUATIONS OF MOTION TO FREQUENCY MODES EQUATIONS OF MOTION

For the dynamics of the signal mode, the derivation starts with the Fourier transform of the operator $\hat{\psi}_S(z, t)$

$$\hat{\psi}_J(z, t) = \frac{1}{\sqrt{2\pi v_J}} \int d\omega \exp[-i(\omega + \bar{\omega}_J)t] \exp\left[i\omega \frac{z}{v_P}\right] \hat{a}_J(z, \omega). \quad (\text{A.1})$$

Then,

$$\partial_t \hat{\psi}_S(z, t) = \frac{-i}{\sqrt{2\pi v_S}} \int d\omega (\omega + \bar{\omega}_S) \exp[-i(\omega + \bar{\omega}_S)t] \exp\left[i\omega \frac{z}{v_P}\right] \hat{a}_S(z, \omega), \quad (\text{A.2})$$

$$\begin{aligned} i\xi(z) \langle \hat{\psi}(z, t) \rangle \hat{\psi}(z, t) &= \frac{i\xi(z)}{2\pi\sqrt{v_P v_I}} \int d\omega d\omega' \exp[-i(\omega + \omega' + \bar{\omega}_I + \bar{\omega}_S)t] \\ &\quad \exp\left[i(\omega + \omega') \frac{z}{v_P}\right] \exp[i(\omega' + \bar{\omega}_I)t] \exp\left[-i\omega' \frac{z}{v_P}\right] \langle \hat{a}_P(z, \omega + \omega') \rangle \hat{a}_I^\dagger(z, \omega') \\ &= \frac{i\xi(z)}{2\pi\sqrt{v_P v_I}} \int d\omega d\omega' \exp[-i(\omega + \bar{\omega}_S)t] \exp\left[i\omega \frac{z}{v_P}\right] \langle \hat{a}_P(z, \omega + \omega') \rangle \hat{a}_I^\dagger(z, \omega'), \end{aligned} \quad (\text{A.3})$$

$$v_S \partial_z \hat{\psi}_S(z, t) = \frac{v_S}{\sqrt{2\pi v_S}} \int d\omega \exp[-i(\omega + \bar{\omega}_S)t] \exp\left[i\omega \frac{z}{v_P}\right] \left[i \frac{\omega}{v_P} \hat{a}_S(z, \omega) + \partial_z \hat{a}_S(z, \omega) \right]. \quad (\text{A.4})$$

Suppose the terms being integrated over $\int d\omega$ are equal, then

$$\partial_t \hat{\psi}_S(z, t) - i\xi(z) \langle \hat{\psi}_P(z, t) \rangle \hat{\psi}_I^\dagger(z, t) + v_S \partial_z \hat{\psi}_S(z, t) + i\bar{\omega}_S \hat{\psi}_S(z, t) = 0 \quad (\text{A.5})$$

$$(\text{A.6})$$

becomes

$$\begin{aligned}
& \frac{-i}{\sqrt{v_S}} (\omega + \bar{\omega}_S) \hat{a}_S(z, \omega) - \frac{i\xi(z)}{\sqrt{2\pi v_P v_I}} \int d\omega' \langle \hat{a}_P(z, \omega + \omega') \rangle \hat{a}_I^\dagger(z, \omega') + \sqrt{v_S} \left[i \frac{\omega}{v_P} \hat{a}_S(z, \omega) + \partial_z \hat{a}(z, \omega) \right] \\
& + \frac{i}{\sqrt{v_S}} \bar{\omega}_S \hat{a}(z, \omega) \\
& = \frac{-i}{v_S} \omega \hat{a}_S(z, \omega) - \frac{i\xi(z)}{\sqrt{2\pi v_P v_I v_S}} \int d\omega' \langle \hat{a}_P(z, \omega + \omega') \rangle \hat{a}_I^\dagger(z, \omega') + \frac{i\omega}{v_P} \hat{a}_S(z, \omega) + \partial_z \hat{a}_S(z, \omega) \\
& = -\frac{i\xi(z)}{\sqrt{2\pi v_P v_I v_S}} \int d\omega' \langle \hat{a}_P(z, \omega + \omega') \rangle \hat{a}_I^\dagger(z, \omega') + i\omega \left(\frac{1}{v_P} - \frac{1}{v_S} \right) \hat{a}_S(z, \omega) + \partial_z \hat{a}_S(z, \omega) = 0.
\end{aligned} \tag{A.7}$$

The procedure is the same with $\hat{\psi}_I(z, t)$.

APPENDIX B FORMULATION OF SCHMIDT NUMBER FROM THE PROPAGATOR

Taking the definition of $g^{(2)}$ being

$$g^{(2)} = \frac{\Delta\omega^2 \left\langle \sum_{ij} \left(\hat{a}_S^\dagger(z, \omega_i) \hat{a}_S(z, \omega_i) \right) \left(\hat{a}_S^\dagger(z, \omega_j) \hat{a}_S(z, \omega_j) \right) \right\rangle - \Delta\omega \left\langle \sum_i \hat{a}_S^\dagger(z, \omega_i) \hat{a}_S(z, \omega_i) \right\rangle^2}{\Delta\omega^2 \left(\left\langle \sum_i \hat{a}_S^\dagger(z, \omega_i) \hat{a}_S(z, \omega_i) \right\rangle \right)^2} = 1 + \frac{1}{K}, \quad (\text{B.1})$$

it can be send that

$$\left\langle \sum_i \Delta\omega N(z, \omega_i) \right\rangle = \left\langle \sum_n \Delta\omega (\hat{\mathbf{a}}_S(z))_n^\dagger (\hat{\mathbf{a}}_S(z))_n \right\rangle = \sum_n \langle vac | (\Delta\omega \hat{\mathbf{a}}_S(z))_n^\dagger (\hat{\mathbf{a}}_S(z))_n | vac \rangle, \quad (\text{B.2})$$

$$= \sum_n \langle vac | \Delta\omega \left(\sum_m (\mathbf{U}_{n,m}^{SS})^* (\hat{\mathbf{a}}_S(z_0))_m^\dagger + \sum_m (\mathbf{U}_{n,m}^{SI})^* (\hat{\mathbf{a}}_I(z_0))_m \right) \quad (\text{B.3})$$

$$\begin{aligned} & \left(\sum_{m'} \mathbf{U}_{n,m'}^{SS} (\hat{\mathbf{a}}_S(z_0))_{m'} + \sum_{m'} \mathbf{U}_{n,m'}^{SI} (\hat{\mathbf{a}}_I^\dagger(z))_{m'} \right) | vac \rangle, \\ &= \sum_n \langle vac | \Delta\omega \left(\sum_m (\mathbf{U}_{n,m}^{SI})^* (\hat{\mathbf{a}}_I(z))_m \sum_{m'} (\mathbf{U}_{n,m'}^{SI} (\hat{\mathbf{a}}_I(z))_{m'}^\dagger) \right) | vac \rangle \\ &= \Delta\omega \sum_{nmm'} \mathbf{U}_{n,m}^{SI*} \mathbf{U}_{n,m'}^{SI} \frac{\delta_{mm'}}{\Delta\omega} \\ &= \text{Tr} \left[\mathbf{U}^{SI\dagger} \mathbf{U}^{SI} \right], \quad (\text{B.4}) \end{aligned}$$

from the commutation relation $\left[(\hat{\mathbf{a}}_I(z))_i, (\hat{\mathbf{a}}_I^\dagger(z))_j \right] = \delta(\omega_i - \omega_j) = \frac{\delta_{ij}}{\Delta\omega}$, with $\delta(\omega_i - \omega_j)$ Dirac's Delta and δ_{ij} Kronecker's Delta. $\mathbf{U}^{SI} \equiv \mathbf{U}_{SI}$ is used for notation simplicity. Also,

$$\Delta\omega^2 \left\langle \sum_{in} \hat{N}(z, \omega_i) \hat{N}(z, \omega_n) \right\rangle = \Delta\omega^2 \left\langle \sum_i (\hat{\mathbf{a}}_S(z))_i^\dagger (\hat{\mathbf{a}}_S(z))_i \sum_n (\hat{\mathbf{a}}_S(z))_n^\dagger (\hat{\mathbf{a}}_S(z))_n \right\rangle, \quad (\text{B.5})$$

$$= \Delta\omega^2 \sum_{in} \left\langle (\hat{\mathbf{a}}_S(z))_i^\dagger (\hat{\mathbf{a}}_S(z))_i (\hat{\mathbf{a}}_S(z))_n^\dagger (\hat{\mathbf{a}}_S(z))_n \right\rangle, \quad (\text{B.6})$$

$$= \Delta\omega^2 \sum_{in} \left\langle (\hat{\mathbf{a}}_S(z))_i^\dagger (\hat{\mathbf{a}}_S(z))_n^\dagger (\hat{\mathbf{a}}_S(z))_i (\hat{\mathbf{a}}_S(z))_n + (\hat{\mathbf{a}}_S(z))_i^\dagger (\hat{\mathbf{a}}_S(z))_n \frac{\delta_{jm}}{\Delta\omega} \right\rangle, \quad (\text{B.7})$$

$$\begin{aligned} &= \sum_{in} \langle vac | \Delta\omega^2 \left[\sum_k \mathbf{U}^{SS*}{}_{ik} (\hat{\mathbf{a}}_S(z_0))_k^\dagger + \sum_k \mathbf{U}^{s,i}{}_{ik} (\hat{\mathbf{a}}_I(z))_k \right] \\ &\quad \left[\sum_{k'} \mathbf{U}^{SS*}{}_{nk'} (\hat{\mathbf{a}}_S(z_0))_{k'}^\dagger + \sum_{k'} \mathbf{U}^{SI}{}_{nk'} (\hat{\mathbf{a}}_I(z_0))_{k'} \right] \\ &\quad \left[\sum_{k''} \mathbf{U}^{SS}{}_{ik''} (\hat{\mathbf{a}}_S(z_0))_{k''} + \sum_{k''} \mathbf{U}^{s,i*}{}_{ik''} (\hat{\mathbf{a}}_I(z_0))_{k''}^\dagger \right] \\ &\quad \left[\sum_{k'''} \mathbf{U}^{SS}{}_{nk'''} (\hat{\mathbf{a}}_S(z_0))_{k'''} + \sum_{k'''} \mathbf{U}^{SI*}{}_{nk'''} (\hat{\mathbf{a}}_I(z_0))_{k'''}^\dagger \right] | vac \rangle \\ &\quad + \sum_i \langle vac | \Delta\omega (\hat{\mathbf{a}}_S(z))_i^\dagger (\hat{\mathbf{a}}_S(z))_i | vac \rangle \\ &= \left\langle \sum_i \hat{N}(z, \omega_i) \right\rangle + \Delta\omega^2 \sum_{in} \langle vac | \sum_{kk'k''k'''} \mathbf{U}^{SI}{}_{ik} \mathbf{U}^{SI}{}_{nk'} \mathbf{U}^{SI*}{}_{ik''} \mathbf{U}^{SI*}{}_{nk'''} \\ &\quad (\hat{\mathbf{a}}_I(z_0))_k (\hat{\mathbf{a}}_I(z_0))_{k'} (\hat{\mathbf{a}}_I(z_0))_{k''}^\dagger (\hat{\mathbf{a}}_I(z_0))_{k'''}^\dagger | vac \rangle \end{aligned} \quad (\text{B.8})$$

$$= \left\langle \sum_i \hat{N}(z, \omega_i) \right\rangle + \Delta\omega^2 \sum_{in} \sum_{kk'k''k'''} \mathbf{U}^{SI}{}_{ik} \mathbf{U}^{SI}{}_{nk'} \mathbf{U}^{SI*}{}_{ik''} \mathbf{U}^{SI*}{}_{nk'''} \quad (\text{B.9})$$

$$\left(\frac{\delta_{kk'''} \delta_{k'k''}}{\Delta\omega^2} + \frac{\delta_{kk''} \delta_{k'k'''}}{\Delta\omega^2} \right) \quad (\text{B.10})$$

$$= \left\langle \sum_i \hat{N}(z, \omega_i) \right\rangle + \left\langle \sum_i \hat{N}(z, \omega_i) \right\rangle^2 + \text{Tr} \left[\mathbf{U}^{(SI)\dagger} \mathbf{U}^{(SI)} \mathbf{U}^{(SI\dagger)} \mathbf{U}^{(SI)} \right]. \quad (\text{B.11})$$

Then, the Schmidt number can be said to be

$$\begin{aligned}
K &= \frac{\Delta\omega^2 \left\langle \sum_i \hat{N}(z, \omega_i) \right\rangle^2}{\Delta\omega^2 \left\langle \sum_{ij} \hat{N}(z, \omega_i) \hat{N}(z, \omega_j) \right\rangle - \Delta\omega \left\langle \sum_i \hat{N}(z, \omega_i) \right\rangle - \Delta\omega^2 \left\langle \sum_i \hat{N}(z, \omega_i) \right\rangle^2} \\
&= \frac{\text{Tr} \left[\mathbf{U}^{(SI)} \mathbf{U}^{(SI)\dagger} \right]^2}{\text{Tr} \left[\mathbf{U}^{(SI)} \mathbf{U}^{(SI)\dagger} \mathbf{U}^{(SI)} \mathbf{U}^{(SI)\dagger} \right]}.
\end{aligned} \tag{B.12}$$

APPENDIX C COMMUTATION RELATION BETWEEN THE PSEUDO-QUADRATURE OPERATORS

The commutation relations for the pseudo-quadrature operators are shown using the commutation relations between the annihilation and creation operators of the signal and idler mode.

$$\begin{aligned}
& [\hat{X}_+(z, \omega_i), \hat{X}_+^\dagger(z, \omega_j)] \\
&= \frac{1}{2} \left[\left(\hat{a}_S(z, \omega) + \hat{a}_I^\dagger(z, \omega_i) \right) \left(\hat{a}_S^\dagger(z, \omega_j) + \hat{a}_I(z, \omega_j) \right) - \left(\hat{a}_S^\dagger(z, \omega_j) + \hat{a}_I(z, \omega_j) \right) \left(\hat{a}_S(z, \omega_i) + \hat{a}_I^\dagger(z, \omega_i) \right) \right] \\
&= \frac{1}{2} \left[\hat{a}_S(z, \omega_i) \hat{a}_S^\dagger(z, \omega_j) - \hat{a}_S^\dagger(z, \omega_j) \hat{a}_S(z, \omega_i) + \hat{a}_I^\dagger(z, \omega_i) \hat{a}_I(z, \omega_j) - \hat{a}_I(z, \omega_j) \hat{a}_I^\dagger(z, \omega_i) \right] \\
&+ \frac{1}{2} \left[\hat{a}_S(z, \omega_i) \hat{a}_I(z, \omega_j) + \hat{a}_I^\dagger(z, \omega_i) \hat{a}_S^\dagger(z, \omega_j) - \hat{a}_S^\dagger(z, \omega_j) \hat{a}_I^\dagger(z, \omega_i) - \hat{a}_I(z, \omega_j) \hat{a}_S(z, \omega_i) \right] \\
&= \frac{1}{2} (\delta_{ij} - \delta_{ji}) = 0.
\end{aligned} \tag{C.1}$$

The rest of the commutation relations from Eq. (2.80) are obtained in similar fashion and give 0, except $[\hat{X}_+(z, \omega_i), \hat{X}_-^\dagger(z, \omega_j)]$, which is

$$\begin{aligned}
& [\hat{X}_+(z, \omega_i), \hat{X}_-^\dagger(z, \omega_j)] \\
&= \frac{1}{2} \left[\left(\hat{a}_S(z, \omega) + \hat{a}_I^\dagger(z, \omega_i) \right) \left(\hat{a}_S^\dagger(z, \omega_j) - \hat{a}_I(z, \omega_j) \right) - \left(\hat{a}_S^\dagger(z, \omega_j) - \hat{a}_I(z, \omega_j) \right) \left(\hat{a}_S(z, \omega_i) + \hat{a}_I^\dagger(z, \omega_i) \right) \right] \\
&= \frac{1}{2} \left[\hat{a}_S(z, \omega_i) \hat{a}_S^\dagger(z, \omega_j) - \hat{a}_S^\dagger(z, \omega_j) \hat{a}_S(z, \omega_i) - \hat{a}_I^\dagger(z, \omega_i) \hat{a}_I(z, \omega_j) + \hat{a}_I(z, \omega_j) \hat{a}_I^\dagger(z, \omega_i) \right] \\
&+ \frac{1}{2} \left[-\hat{a}_S(z, \omega_i) \hat{a}_I(z, \omega_j) + \hat{a}_I(z, \omega_j) \hat{a}_S(z, \omega_i) + \hat{a}_I^\dagger(z, \omega_i) \hat{a}_S^\dagger(z, \omega_j) - \hat{a}_S^\dagger(z, \omega_j) \hat{a}_I^\dagger(z, \omega_i) \right] \\
&= \frac{1}{2} (\delta_{ij} + \delta_{ji}) = \delta_{ij}.
\end{aligned} \tag{C.2}$$

APPENDIX D ERROR ON THE NUMERICAL APPROXIMATION OF THE EQUATIONS OF MOTION

From Schmidt number convergence, for a frequency range from -5 to 5 , a frequency window of 21×21 seems to be enough. With these parameters and setting $\frac{1}{\Delta v} = 1$, number of discretization in ζ of 6 ($N_z = 6$) and $\zeta_f = 1$, the error on the numerical approximation of the integrals can be seen in Table 3.5, where the error is the max norm, or $\text{error}(\mathbf{X}, \mathbf{Y}) = \max(\text{abs}(\mathbf{X}_{ij} - \mathbf{Y}_{ij})) \forall i, j \in \dim(\mathbf{X})$. The pump used is the low-gain optimal one.

Table D.1: Error on numerical approximation of the integral on the dynamics. When $\langle n \rangle$ increases, the error does not. Its value is however quite large, so rescaling the frequency range or the waveguide parameters may be needed to obtain better accuracy. But to do so requires an larger positive semidefinite matrix, which requires more computational resources.

Error of real part	Error on imaginary part	$\langle n \rangle$
3.53×10^{-1}	2.31×10^{-1}	0.0001
3.39×10^{-1}	2.23×10^{-1}	0.2
3.32×10^{-1}	2.20×10^{-1}	0.3
3.21×10^{-1}	2.14×10^{-1}	0.5
2.80×10^{-1}	1.89×10^{-1}	1.0

Setting $\langle n \rangle = 1$, for different discretization of ζ , the numerical error is shown in Table D.2.

Table D.2: Error on numerical approximation of the integral on the dynamics at $\langle n \rangle = 1$ with low-gain optimal pump. When N_z increases, the error decreases. Indeed, rescaling the problem can lead to better accuracy in numerical approximation.

Error of real part	Error on imaginary part	N_z
2.18×10^{-1}	1.46×10^{-1}	4
5.12×10^{-2}	3.45×10^{-2}	5
2.80×10^{-2}	1.89×10^{-2}	6
1.81×10^{-2}	1.22×10^{-2}	7
1.28×10^{-2}	8.65×10^{-3}	8
7.46×10^{-3}	5.03×10^{-3}	10
2.98×10^{-3}	2.01×10^{-3}	15

APPENDIX E EQUIVALENCE IN OBJECTIVE FUNCTION OF NUCLEAR NORM AND SCHMIDT NUMBER

Since the term in Eq. (2.86) is positive, minimizing it is equivalent to minimizing its square root. The objective function then becomes

$$\min_{\beta} f(\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta)) = \frac{\text{Tr}[(\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta))(\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta))^\dagger]}{\sqrt{\text{Tr}[(\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta))(\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta))^\dagger]^2}}. \quad (\text{E.1})$$

The numerator is the sum of squared singular values and the denominator is the square root of the sum of singular values to the power of 4. Since the singular values are always nonnegative, the objective function can also be stated as

$$\min_{\beta} \frac{|\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta)|_{\text{Nuc}}}{\sqrt{\text{Tr}[(\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta))(\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta))^\dagger]}}, \quad (\text{E.2})$$

where $|\cdot|_{\text{Nuc}}$ is the nuclear norm defined as $|\mathbf{A}|_{\text{Nuc}} = \text{Tr}[\sqrt{\mathbf{A}\mathbf{A}^\dagger}]$. Then, supposing that $\text{Tr}[\mathbf{U}_{SI}(\zeta)\mathbf{U}_{SI}^\dagger(\zeta)] = \langle n \rangle = \frac{1}{4} \text{Tr}[(\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta))(\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta))^\dagger]$, $f(\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta))$ is equivalent to

$$\min_{\beta} \frac{|\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta)|_{\text{Nuc}}}{2\sqrt{\langle n \rangle}}. \quad (\text{E.3})$$

In fact, the argmin of those functions are the same. Consider Eq. (E.1), let d_i be the i th singular value of $\mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta)$ and $\sigma_i = d_i^2$, then objective in Eq. (E.1) has a minimum value of 1 as in

$$\frac{\sum \sigma_i}{\sqrt{\sum \sigma_i^2}} \geq 1. \quad (\text{E.4})$$

It can be proved by contradiction; suppose $\frac{\sum \sigma_i}{\sqrt{\sum \sigma_i^2}} < 1$, then

$$\sum \sigma_i - \sqrt{\sum \sigma_i^2} < 0, \quad (\text{E.5})$$

$$\sqrt{\sum \sigma_i^2 + \sum_{i \neq j} \sigma_i \sigma_j} - \sqrt{\sum \sigma_i^2} < 0, \quad (\text{E.6})$$

$$\sum_{i \neq j} \sigma_i \sigma_j < 0, \quad (\text{E.7})$$

but it is taken into account that $\sigma_i \geq 0$, so $\frac{\sum \sigma_i}{\sqrt{\sum \sigma_i^2}}$ must not be less than 1.

If $\mathbf{A} = \mathbf{U}_+(\zeta) - \mathbf{U}_-(\zeta)$ is of rank 1, then $\mathbf{A} = a\mathbf{v}\mathbf{v}^\dagger$ where $\mathbf{v}^\dagger\mathbf{v} = \mathbf{w}^\dagger\mathbf{w} = 1$ and

$$\frac{\text{Tr} [\mathbf{A}\mathbf{A}^\dagger]}{\sqrt{\text{Tr} [(\mathbf{A}\mathbf{A}^\dagger)^2]}} = \frac{\text{Tr} [a^2\mathbf{v}\mathbf{v}^\dagger]}{\sqrt{\text{Tr} [a^4\mathbf{v}\mathbf{v}^\dagger]}} = 1. \quad (\text{E.8})$$

The case where $\mathbf{A}$ is of rank 1 then gives the minimum of Eq. (E.1). For the objective function in Eq. (E.2), the minimum is also 1 by taking the same argument as done previously, while remembering the singular values are always defined to be non-negative. For a rank 1 matrix, Eq. (E.2) becomes

$$\frac{\text{Tr} [\sqrt{\mathbf{A}\mathbf{A}^\dagger}]}{\sqrt{\text{Tr} [\mathbf{A}\mathbf{A}^\dagger]}} = \frac{\text{Tr} [a\sqrt{\mathbf{v}\mathbf{v}^\dagger}]}{\sqrt{\text{Tr} [a^2\mathbf{v}\mathbf{v}^\dagger]}} = 1. \quad (\text{E.9})$$

To find the maximum of Schmidt number over the constraint $\frac{1}{4} \text{Tr} [\mathbf{A}\mathbf{A}^\dagger] = \langle n \rangle$, first it can be seen that

$$\frac{\sum_{i=1}^N \sigma_i}{\sqrt{\sum_{i=1}^N \sigma_i^2}} = \sqrt{\frac{\sum_{i=1}^N \sigma_i^2 + \sum_{i \neq j}^N \sigma_i \sigma_j}{\sum_{i=1}^N \sigma_i^2}} = \sqrt{1 + \frac{\sum_{i \neq j}^N \sigma_i \sigma_j}{\sum_{i=1}^N \sigma_i^2}} \leq \sqrt{N}, \quad (\text{E.10})$$

where N is the size of $\mathbf{A}$. To prove this, first let

$$(\sigma_i - \sigma_j)^2 = \sigma_i^2 + \sigma_j^2 - 2\sigma_i\sigma_j \geq 0, \quad (\text{E.11})$$

then one can prove

$$\sum_{i \neq j}^N \sigma_i \sigma_j \leq \frac{N}{2} \sum_i \sigma_i^2 + \frac{N}{2} \sum_j \sigma_j^2 - \sum_i \sigma_i^2 = (N-1) \sum_i \sigma_i^2. \quad (\text{E.12})$$

Then,

$$\sqrt{1 + \frac{\sum_{i \neq j}^N \sigma_i \sigma_j}{\sum_{i=1}^N \sigma_i^2}} \leq \sqrt{1 + (N-1)} = \sqrt{N}. \quad (\text{E.13})$$

which is the maximum of Eq. (E.1). For the objective function in Eq. (E.2), the maximum is the exact same value through the same argument.

For Eq. (E.1), the maximum is attained when the matrix $\mathbf{A}$ is full rank with every singular value being $2\sqrt{\frac{\langle n \rangle}{N}}$. Evaluating the equation at this point, Eq. (2.86) takes on the value

$$\frac{\sum_{i=1}^N \frac{4\langle n \rangle}{N}}{\sqrt{\sum_{i=1}^N \frac{16\langle n \rangle^2}{N^2}}} = \frac{4\langle n \rangle}{\frac{4\langle n \rangle}{\sqrt{N}}} = \sqrt{N}. \quad (\text{E.14})$$

For Eq. (E.2), at the same conditions, the value of the function becomes

$$\frac{\sum_{i=1}^N 2\sqrt{\frac{\langle n \rangle}{N}}}{\sum_{i=1}^N \frac{4\langle n \rangle}{N}} = \frac{2\sqrt{N\langle n \rangle}}{\sqrt{4\langle n \rangle}} = \sqrt{N}. \quad (\text{E.15})$$

Then, from an optimization point of view, Eq. (E.1) and (E.2) are equivalent objective functions and the nuclear norm formulation can replace the Schmidt number.