On shape optimisation of optical waveguides using Inverse Problem techniques.

by
Thomas Felici & Heinz Engl.

1. Introduction

Optical waveguides are the basis of the optoelectonics and telecommunications industry. The best known example of an optical waveguide is the optical fibre, which has already replaced the telephone copper wire as the means of data transportation on all modern telecommunication networks. Less familiar, but just as important, are the optical waveguide components that now make up the optoelectronic systems which manipulate, filter, and dispatch incoming optical signals. These are usually fabricated via an etching process, and can have complicated structural features on the order of 0.25μm. Such "diffractive optics" devices have great advantages in terms of size and weight, and can often be designed to perform functions unattainable with traditional optical elements. For example, structures with spatially periodic features (diffraction gratings) are used as spectral filters, polarizers, waveguide couplers, etc. The development and application of this new technology increasingly relies on accurate mathematical models and numerical calculations both for the prediction of device behaviour and for the determination of "optimal" device designs. In contrast to the case of traditional optical structures, geometrical optics is generally not sufficiently accurate for these diffractive devices. The computational problem is much more challenging, requiring the solution of a full partial differential equation model.
A taper is a generic kind of optical waveguide with a cross-section that varies continuously along its length z. Tapers are used to couple light from a waveguide into a another waveguide with different cross sectional profile. It is well known that for large enough lengths, the light may be transmitted without power loss into the output waveguide. This is called the adiabatic regime of a taper. As the length decreases, the power loss increases. The aim of this study is to develop a formulation to minimise the taper length while keeping an acceptably low loss. This is achieved by varying the taper profile.
We start by setting out the direct problem for the propagation of the EM field in a generic waveguide, and define the optimisation problem. We then proceed to show how this can be in some sense ill posed. We then derive a modal formulation of the problem, and establish the evolution equations for the modal excitations in the taper. This is a convenient formulation if we want to determine the power excited in the fundamental mode of the exit waveguide, as this corresponds to the first coefficient of the modal expansion. The discrete optimisation problem is then derived by considering the taper as a sequence of uniform cross sections. The field in each section is expressed as an expansion of local modes, whose coefficients are determined by the field continuity condition across each sub section. We show how numerically the above mentioned ill posedness slows down the convergence of a classical optimisation algorithm with increasing discretisation refinement of the original structure. The ill posedness of the problem suggests the use of some sort of regularisation: we proceed to show how the power maximisation problem can be formulated as a non linear inverse problem, which can then be solved using established inverse problem regularisation techniques [5]. Numerical results presented here show that this new approach can lead to robust optimisation algorithms less sensitive to large discretisation refinements.

Inverse Problems **17** (2001) 1141-1162

For related work in shape optimisation of other types of optical devices, see [1] and [2].

2. Formulation of 2d direct problem.

For simplicity and clarity we restrict ourselves to the 2D case, although the entire analysis is directly applicable to the general 3D problems.
The source free Maxwell's equations in a continuous medium with varying linear permettivity ε are:

$$(1)\qquad \begin{cases} \nabla \wedge \mathbf{H} = \varepsilon \dot{\mathbf{E}} \\ \nabla \wedge \mathbf{E} = -\mu_0 \dot{\mathbf{H}} \\ \nabla \cdot (\varepsilon \mathbf{E}) = \nabla \cdot \mathbf{H} = \mathbf{0} \end{cases}$$

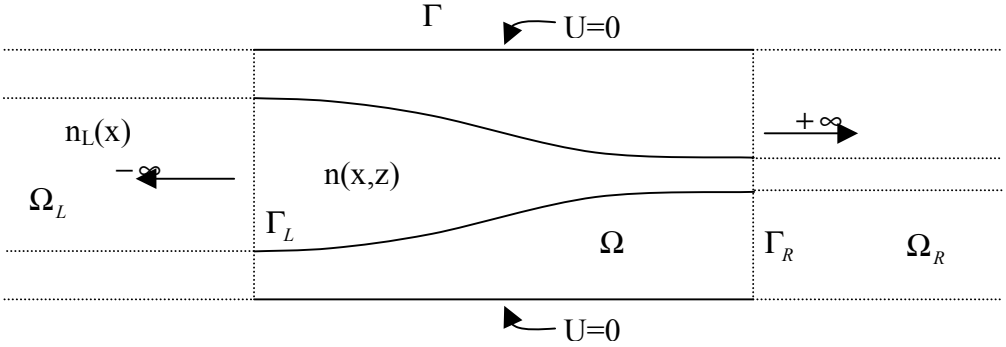
in 2-d, we assume that $\partial / \partial y = 0$. Then have a consistent solution with $H_y=E_x=E_z=0$ (the TE field) with:

$$(2)\qquad \begin{cases} \dfrac{\partial H_x}{\partial z} - \dfrac{\partial H_z}{\partial x} = -i\omega\varepsilon E_y \\ \dfrac{\partial E_y}{\partial z} = -i\omega\mu_0 H_x \\ \dfrac{\partial E_y}{\partial x} = i\omega\mu_0 H_z \end{cases}$$

where we have assumed a time harmonic dependency e^{-iwt}. This leads to the Helmholtz equation for E_y:

$\Delta E_y + k^2 n^2 E_y = 0$ with $\omega\sqrt{\mu_0\varepsilon_0} = k$ and $n^2 = \varepsilon/\varepsilon_0$.

Another set of solution exists with $E_y=H_x=H_z=0$ (the TM field), leading to a slightly modified equation, but we shall only deal with the TE modes here. The analysis is also valid in three dimensions if we assume a weakly guiding approximation, which essentially states that the variation of the refractive index throughout the region is small. For more details on this, as well as the treatment of the general 3D case, see [3].
We now formulate the propagation problem for the field in a generic waveguide.



From now on we write U instead of E_y. We assume that on the sides we have "hard walls" (U=0). This physically means that the field is reflected back into the region. In practice this is not a problem if we are seeking guided field solutions, which by definition are bound to the waveguide code and decrease exponentially on the outside. The field is then given by:

$$(3)\qquad \Delta U + n^2 U = 0 \qquad ; \quad \text{for} \quad (x,z) \in \Omega$$

(4) $U|_\Gamma = 0$

where we have set k=1 wlog.

Now, we need conditions at $\pm\infty$. First assume we have an input field from LHS coming from $-\infty$. This field U_I satisfies (3),(4) in Ω_L. We need to express the fact that the scattered field U-U_I is outward travelling on both sides. We can do this in terms of a modal expansion in Ω_L, Ω_R. So consider problem of finding eigen modes in these end regions.

In Ω_L: $\Delta U + n^2 U = 0$. Assume $U(x,z) = \tilde{U}(x)e^{i\beta z}$. Then

$$\frac{d^2\tilde{U}}{dx^2} + (n^2 - \beta^2)\tilde{U} = 0 \quad \text{for fixed z}$$

with $\tilde{U}(x_{\min}) = \tilde{U}(x_{\max}) = 0$

This is an eigen-problem. The "hard wall" boundary condition ensures that we have discrete eigen-modes $\{(\tilde{U}_k, \pm\beta_k); k = 1,2,\ldots\}$. Therefore:

$$(5) \qquad U(x,z) = \sum_{k=0}^{\infty} \left(C_k e^{i\beta_k z} + C_{-k} e^{-i\beta_k z}\right)\tilde{U}_k(x) \qquad \text{in } \Omega_L$$

An identical expression is true for the RHS.

Note, with appropriate reordering, following property holds:

$\max_{\Omega_{2D}}(n^2) \geq \beta_1^2 \geq \beta_2^2 \geq \ldots \to -\infty$.

In (5), set $\beta_k > 0$ if β_k is real , $\beta_k = i|\beta_k|$ if β_k is imaginary.

Under this notation, an outgoing wave U_L (to the LHS) is given by setting $C_{+k} = 0 \quad \forall\, k > 0$.

Hence (5) => $\alpha_k \equiv C_{-k} e^{i\beta_{-k} z} = \langle U_L, \tilde{U}_k \rangle \equiv \int_{x_{\min}}^{x_{\max}} U_L(x,z)\tilde{U}_k(x)dx$.

An incoming wave U_I (from the LHS) is given by setting $C_{-k} = 0 \quad \forall\, k > 0$.

Hence (5) => $\alpha_k \equiv C_k e^{i\beta_k z} = \langle U_I, \tilde{U}_k \rangle$.

Differentiating (5):

$$\frac{\partial U}{\partial z} = \sum_{k=1}^{\infty} i\beta_k(\alpha_k + \alpha_{-k})\tilde{U}_k$$

$$\Rightarrow \frac{\partial U_L}{\partial z} = -i\sum_{k=1}^{\infty} \beta_k^{(L)} \langle U_L, \tilde{U}_k^{(L)} \rangle \tilde{U}_k^{(L)} \quad \text{in } \Omega_L \text{ -this is the radiation condition for the LHS.}$$

$$\text{and } \frac{\partial U_I}{\partial z} = i\sum_{k=1}^{\infty} \beta_k^{(L)} \langle U_I, \tilde{U}_k^{(L)} \rangle \tilde{U}_k^{(L)} \quad \text{in } \Omega_L \text{ -this is the incoming condition from the LHS.}$$

Hence, since U= U_I+U_L:

$$\frac{\partial U}{\partial z} - \frac{\partial U_I}{\partial z} = -i\sum_{k=1}^{\infty} \beta_k^{(L)} \langle U - U_I, \tilde{U}_k^{(L)} \rangle \tilde{U}_k^{(L)}$$

$$\Rightarrow \frac{\partial U}{\partial z} - i\sum_{k=1}^{\infty} \beta_k^{(L)} \langle U_I, \tilde{U}_k^{(L)} \rangle \tilde{U}_k^{(L)} = -i\sum_{k=1}^{\infty} \beta_k^{(L)} \langle U, \tilde{U}_k^{(L)} \rangle \tilde{U}_k^{(L)} + i\sum_{k=1}^{\infty} \beta_k^{(L)} \langle U_I, \tilde{U}_k^{(L)} \rangle \tilde{U}_k^{(L)}$$

$$\Rightarrow \frac{\partial U}{\partial z} = -i\sum_{k=1}^{\infty} \beta_k^{(L)} \langle U, \tilde{U}_k^{(L)} \rangle \tilde{U}_k^{(L)} + 2i\sum_{k=1}^{\infty} \beta_k^{(L)} \langle U_I, \tilde{U}_k^{(L)} \rangle \tilde{U}_k^{(L)}$$

Analogously for the RHS we wave

$\frac{\partial U}{\partial z} = i\sum_{k=1}^{\infty} \beta_k^{(R)} \left\langle U, \widetilde{U}_k^{(R)} \right\rangle \widetilde{U}_k^{(R)}$ in Ω_R -this is the radiation condition for the RHS. (We assume no incoming wave from the right).

So complete problem is: Find $U(x,y,z)$ s.t.

$$(6) \qquad \begin{cases} \Delta U + n^2 U = 0 \quad ; \quad for \quad (x,y,z) \in \Omega \\ U|_{\Gamma} = 0 \qquad \text{(reflecting side walls)} \\ \dfrac{\partial U}{\partial z} + i\sum_{k=1}^{\infty} \beta_k^{(L)} \left\langle U, \widetilde{U}_k^{(L)} \right\rangle \widetilde{U}_k^{(L)} = 2i\sum_{k=1}^{\infty} \beta_k^{(L)} \left\langle U_I, \widetilde{U}_k^{(L)} \right\rangle \widetilde{U}_k^{(L)} \qquad on\ \Gamma_L \\ \dfrac{\partial U}{\partial z} - i\sum_{k=1}^{\infty} \beta_k^{(R)} \left\langle U, \widetilde{U}_k^{(R)} \right\rangle \widetilde{U}_k^{(R)} = 0 \qquad on\ \Gamma_R \end{cases}$$

where $\left\{\left(\widetilde{U}_k^{(L)}, \pm\beta_k^{(L)}\right), k>0\right\}, \left\{\left(\widetilde{U}_k^{(R)}, \pm\beta_k^{(R)}\right), k>0\right\}$ are the eigen-modes in Ω_L, Ω_R respectively.

Note:

1. Γ_L, Γ_R can be chosen at any position in Ω_L, Ω_R respectively.
2. Often the incoming wave front U_I is just a specific mode of the input waveguide, so that U_I is directly given in terms of its modal coefficients: $\left\langle U_I, \widetilde{U}_k^{(L)} \right\rangle = 1\ (k=1) \quad ; \quad = 0\ (k>1)$.

3. Formulation of the optimisation problem.

In general , we are interested in finding the optimal shape (or refractive index distribution) which in some sense maximises the power transfer of the waveguide. Alternatively we might want to optimise the shape according to some other physical requirement (e.g. minimising the length) while keeping the transmitted power loss below a specified threshold.

Often for practical purposes we are interested in the situation where the input field is an excitation of the fundamental mode of the input waveguide, and we are interested in the power remaining in the (guided) fundamental mode $\widetilde{U}_1^{(R)}$ of the output waveguide. In terms of the modal expansion (5), this corresponds to the coefficient $|C_1|^2$. In terms of the actual fields, this is given by:

$$(7) \qquad P(n^2) \equiv \left|\left\langle U, \widetilde{U}_1^{(R)} \right\rangle\right|^2 \equiv \left| \int_{x\in\Gamma_R} U(x, z_R) \widetilde{U}_1^{(R)}(x) dx \right|^2$$

Now, it is well known that maximum power transfer (P=1) can me achieved as the length of the taper tends to infinity. This can be seen using modal analysis [3]. It therefore makes sense to impose the additional constraint of keeping the taper length fixed.

4. Solution of the direct problem.

Due to the boundary conditions that we have imposed, system (6) is most conveniently solved using local modal expansions. This is a fairly standard technique, and we only give a basic outline of the method. For more details we refer to [3].

We start with the field equations (6):

$$(8) \qquad \begin{cases} \Delta U + n^2 U = 0 \quad ; \quad (x,z) \in \Omega \\ U|_{\Gamma} = 0 \quad \text{(reflecting side walls)} \end{cases}$$

Define $\mathrm{L}_t(U) \equiv \dfrac{\partial^2 U}{\partial x^2} + n^2(x,z)U$

Let Ω_z be the 1D cross sectional region at position z along the waveguide region Ω. Define the local basis set at each position z $\{(U_k, \beta_k)\,; k \geq 1\}$ given by the eigen problem:

$$(9) \qquad \begin{cases} \mathrm{L}_t(U_k) = \beta_k{}^2 U_k \quad in \quad \Omega_z \\ U_k\,|_{\partial\Omega_z} = 0 \end{cases}$$

Self adjointness of the operator $\mathbf{L}_t$ ensures that this set forms a complete basis set for any function $f \in \mathrm{C}(\Omega_z)$ with $f\,|_{\partial\Omega_z} = 0$, and the boundary conditions on U_k ensure that this set is discrete. Therefore any function U defined in the region Ω with $U\,|_{\partial\Omega_z} = 0$ can be expressed as a unique expansion of this local (orthogonal) basis set:

$$\exists!\ \mathbf{c}\ s.t.\ U = \sum_{k=1}^{\infty} c_k(z) U_k$$

Where the coefficients depend (only) on z.
Now, we are interested in a formulation which locally gives a representation of the electromagnetic field in terms of the forward and backwards local modal EM fields.
For a waveguide with constant cross section, we have [3]:

$$(10) \qquad \mathbf{E} = \sum_{k=1}^{\infty} (a_k + a_{-k})\mathbf{E}_k \quad ; \quad \mathbf{H} = \sum_{k=1}^{\infty} (a_k - a_{-k})\mathbf{H}_k$$

So that the a_k, a_{-k} 's are respectively the coefficients of forward, backward moving modal fields $\left(\mathbf{E}_k e^{i\beta_k z}, \mathbf{H}_k e^{i\beta_k z}\right)$, $\left(\mathbf{E}_k e^{i\beta_k z}, -\mathbf{H}_k e^{i\beta_k z}\right)$

Moreover, if these eigen-fields are normalised so that

$$\int \mathbf{E}_k \wedge \mathbf{H}_k . \hat{\mathbf{z}}\, ds = 1,$$

then the $|a_k|^2, |a_{-k}|^2$ will be the power in each excited mode.

In particular expression (7) for the output power of the fundamental mode is now just:

$$(11) \qquad P(n^2) = |a_1|^2$$

From (2):

$$\frac{\partial E_y}{\partial z} = -i\omega H_x = -i\omega \sum_{k=1}^{\infty} (a_k - a_{-k}) H_{xk} = \sum_{k=1}^{\infty} (a_k - a_{-k}) i\beta_k E_{xk}$$

We therefore chose our expansion for the function U solution to (6) in terms of the set of coefficients such that :

$$(12) \qquad \begin{cases} U = \sum_{k=1}^{\infty} (a_k + a_{-k}) U_k \\ \dfrac{\partial U}{\partial z} = \sum_{k=1}^{\infty} (a_k - a_{-k}) i\beta_k U_k \end{cases}$$

Hence these expressions are equivalent to the modal expansions (10) with $E_k = U_k$.
The U_k 's are the local eigen functions defined by (9) with normalisation:

$$(13) \qquad \int_{\Omega_z} U_k{}^2\, ds = \frac{1}{\beta_k}.$$

This ensures that the power normalisation is satisfied, and so $|a_k|^2$ is the power in each excited mode.

Using the fact that the derivative of (12)-a wrt z must be identical to (12)-b, and substituting (12) into (8)-a we deduce that the a_k's must obey the coupled ODE system:

(14) $$\dot{a}_k(z) - i\beta_k a_k(z) = \sum_{j\neq k,0} r_{kj}(z) a_j(z) \quad \forall k \neq 0$$

with $\beta_{-k} \equiv -\beta_k$ and $r_{kj}(z) = \dfrac{\int_{\Omega_z} \dfrac{\partial n^2}{\partial z} U_k U_j ds}{2(\beta_k - \beta_j)} \quad \forall\, j \neq k \;;\; j,k \neq 0$.

The boundary conditions at the beginning ($z = 0$) and end ($z = z_R$) translate as:

(15) $$a_k(0) = A_k \quad ; \quad a_{-k}(z_R) = 0 \quad \forall k > 0$$

where the A_k's are the given coefficients corresponding to the LHS input field U_I.

(14) can be rewritten as:

(16) $$\dot{\gamma}_k(z) = \sum_{j\neq k,0} r_{kj}(z) e^{i\int_0^z (\beta_j(z) - \beta_k(z))dz} \gamma_j(z) \quad \text{with} \quad a_k(z) = e^{i\int \beta_k(z)dz} \gamma_k(z)$$

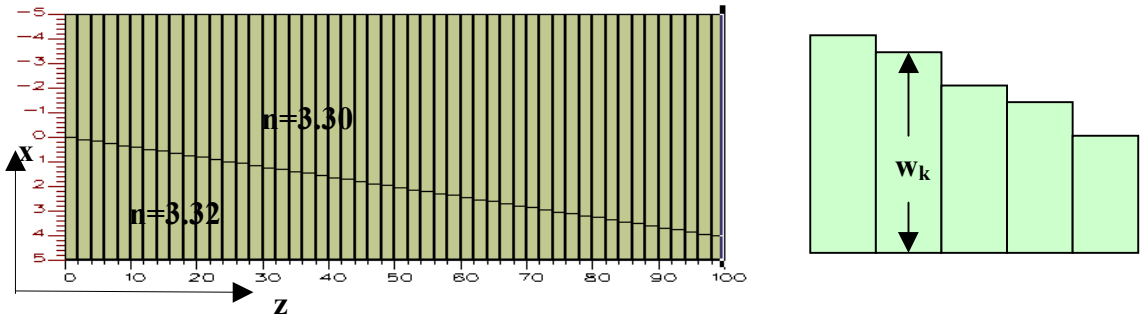
This is the evolution equation for the amplitudes $\gamma_k(z)$. Note that:

1. the RHS does not depend upon $\gamma_k(z)$, which means that this equation describes the cross-coupling between modes.
2. In a waveguide with constant cross section, $r_{kj} = 0$ so $\gamma_k(z)$ is constant, as expected.
3. An upper bound for the cross-coupling terms is given by: $r_{kj} \leq \dfrac{\int_{\Omega_z} \left|\dfrac{\partial n^2}{\partial z}\right| ds}{2|\beta_k - \beta_j|}$ so that the coupling influence tends to be strongest with its immediate neighbours, for which $|\beta_k - \beta_j|$ is smallest.
4. The coupling can have irregular behaviour at a point where neighbouring β's approach the same value. This can lead to some interesting behaviour to be investigated further. For example, one expects strong mode mixing to happen on a very short length scale.
5. This analysis assumes continuity in the refractive index *n(x,z)*. Jumps can be taken into account by relating the coefficients γ_k at either side of the jump using the field continuity conditions.

5. Solving numerically the optimisation problem.

Reduction to a discrete optimisation problem is obtained by parametrising the refractive index profile. The parametrisation chosen here is a discretisation of the continuous shape into a sequence of N sub sections, each with fixed length and uniform in the z direction. The refractive index profile in each subsection is then controlled by a finite set of parameters. In this example, the unknown parameters are the position of the waveguide boundary in each sub section $\{w_1, w_2, \ldots, w_N\}$.

This piecewise continuous discretisation implies that we impose very weak regularity constraints on the taper profile. In particular, we are allowing for solutions with profile discontinuities. The field is calculated via local modal analysis in each sub section. The field expansion coefficients in each subsection are uniquely determined by the continuity condition at each interface as well as the initial field excitation on the LHS.

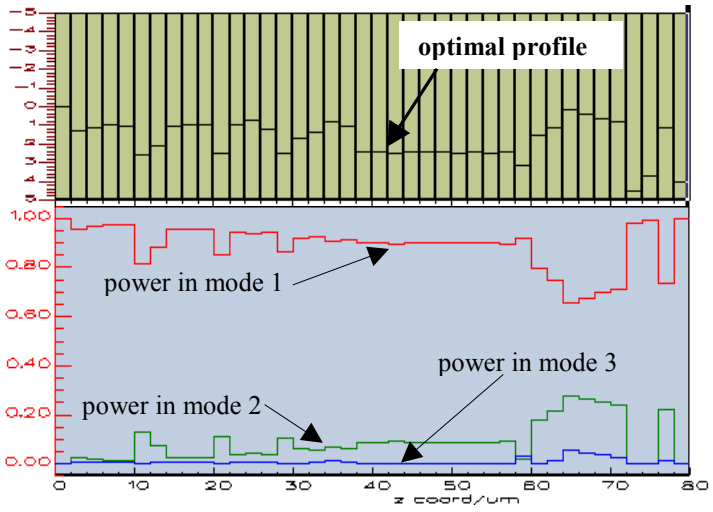


Symmetric upper half of the taper modelled as a sequence of constant sections.

To maximise the resulting objective function $P(w_1,w_2,\ldots,w_N)$ a variable metric optimisation routine was used [4]. The derivatives required by the method were calculated using finite differences. The initial shape was chosen to be the linear taper (see above figure) with two different discretisations N.

6. *Physical interpretation of optimal solutions.*

The results, shown in appendices 1, 2, are interesting in their own right, since they contradict currently held ideas in the optoelectonics community as to what an optimal shape should be for a device which maximises power preservation in the fundamental mode. Indeed it is generally believed that the shape should be taper-like, i.e. a smooth transition between the input and output waveguide cross-sections. It is well known that as the taper length increases, the power excited in the input fundamental mode will tend to remain in the local fundamental mode, which of course is changing along the taper. The "optimal" shape would then be the one that somehow maximises this adiabatic behaviour *throughout* the waveguide. On the contrary, these results show that real optimal shapes seem to be based upon quite a different underlying mechanism - the resonance between adjacent modes: as can be seen from the graph, the beginning of the optimal profile comprises of regularly spaced "teeth" repeating every 10µm. This periodic variation produces a discrete coupling such that that the power, although initially entirely in mode 1, nevertheless remains throughout this region almost entirely in the first two local modes. After a middle straight section, the final "bump" just before the end re-injects all the power now in mode 2 back into mode 1. The total power lost from the fundamental mode is 0.5%. This is achieved in only 80µm, while an equivalent straight taper would have to be 5 times longer before achieving a comparable efficiency!



7. *Ill posedness of the optimisation problem.*

It is evident on observation of the convergence graphs in appendices 1,2 for the optimiser that the straight forward optimisation strategy outlined above runs into difficulties for finer discretisations (appendix 2). The problem gets worse if we further increase the number of subsections N. This excessive slowdown in convergence rate is due to the ill posed nature of the optimisation problem itself, in the sense that we now explain.

In fact the above optimisation problem is ill posed in the sense that we can always find an arbitrarily large variation in the optimal solution, which still gives an arbitrarily close field U. To see this we do the following: suppose that we have any solution n Suppose we apply a perturbation δn^2 to n^2. Then the resulting field U will undergo a change δU is given to first order by perturbing (6):

$$(17)\quad \begin{cases} \Delta\ \delta U + n^2 \delta U = -U\delta n^2 \quad ; \quad for \quad (x,z)\in\Omega \\ \delta U\,|_{\Gamma} = 0 \\ \dfrac{\partial \delta U}{\partial z} + i\sum_{k=1}^{\infty} \beta_k^{(L)} \left\langle \delta U, \widetilde{U}_k^{(L)} \right\rangle \widetilde{U}_k^{(L)} = 0 \qquad on\ \Gamma_L \\ \dfrac{\partial \delta U}{\partial z} - i\sum_{k=1}^{\infty} \beta_k^{(R)} \left\langle \delta U, \widetilde{U}_k^{(R)} \right\rangle \widetilde{U}_k^{(R)} = 0 \qquad on\ \Gamma_R \end{cases}$$

it is convenient to re-express (17) in integral form, so we introduce the Green function $G(\mathbf{r},\mathbf{r}')$, with $\mathbf{r} = (x,z)$, $\mathbf{r}' = (x',z')$, defined by:

$$(18)\quad \begin{cases} \Delta G(\mathbf{r},\mathbf{r}') + n(\mathbf{r})^2 G(\mathbf{r},\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}') \quad ; \quad for \quad \mathbf{r},\mathbf{r}'\in\Omega \\ G(\mathbf{r},\mathbf{r}')|_{\mathbf{r}\in\Gamma} = 0 \quad ; \quad for \quad \mathbf{r}'\in\Omega \\ \dfrac{\partial G(\mathbf{r},\mathbf{r}')}{\partial z} + i\sum_{k=1}^{\infty} \beta_k^{(L)} \left\langle G, \widetilde{U}_k^{(L)} \right\rangle \widetilde{U}_k^{(L)} = 0 \qquad on\ \Gamma_L \\ \dfrac{\partial G(\mathbf{r},\mathbf{r}')}{\partial z} - i\sum_{k=1}^{\infty} \beta_k^{(R)} \left\langle G, \widetilde{U}_k^{(R)} \right\rangle \widetilde{U}_k^{(R)} = 0 \qquad on\ \Gamma_R \end{cases}$$

The field δU is then given in the usual way by evaluating the expression:

$G\Delta\delta U - \delta U \Delta G \equiv \nabla\cdot(G\nabla\delta U - \delta U \nabla G)$

Using the divergence theorem over the domain Ω we obtain:

$$\delta U(\mathbf{r}) = \int_{\mathbf{r}'\in\Omega} -G(\mathbf{r},\mathbf{r}')U(\mathbf{r}')\delta n^2(\mathbf{r}')dv - \int_{\mathbf{r}'\in\partial\Omega} \left(G(\mathbf{r},\mathbf{r}')\frac{\partial U(\mathbf{r}')}{\partial \mathbf{n}} - U\frac{\partial G(\mathbf{r}')}{\partial \mathbf{n}} \right)\cdot \mathbf{n} ds'$$

The contributions of the side walls Γ vanish since $G|_{\Gamma} = U|_{\Gamma} = 0$. The contributions from Γ_L, Γ_R also vanish, e.g.:

$$\int_{\Gamma_L} \left(G\frac{\partial U}{\partial \mathbf{n}} - U\frac{\partial G}{\partial \mathbf{n}} \right)\cdot \mathbf{n} ds = -i\sum_{k=1}^{\infty} \beta_k^{(L)} \left\langle G, \widetilde{U}_k^{(L)} \right\rangle \int_{\Gamma_L} U\widetilde{U}_k^{(L)} ds + i\sum_{k=1}^{\infty} \beta_k^{(L)} \left\langle U, \widetilde{U}_k^{(L)} \right\rangle \int_{\Gamma_L} G\widetilde{U}_k^{(L)} ds = 0$$

same goes for contribution on Γ_R.

Hence have:

(19) $$\delta U(\mathbf{r}) = -\int_{\mathbf{r}'\in\Omega} G(\mathbf{r},\mathbf{r}')U(\mathbf{r}')\delta n^2(\mathbf{r}')dv\,.$$

This can be functionally expressed as a mapping:

$$K : C^0(\Omega) \to K\left(C^0(\Omega)\right) \text{ defined by}$$

$$\delta n^2 :\to \delta U(\mathbf{r}) \equiv -\int_{\mathbf{r}'\in\Omega} G(\mathbf{r},\mathbf{r}')U(\mathbf{r}')\delta n^2(\mathbf{r}')dv$$

where $C^0(\Omega)$ is the space of all continuous functions over Ω and $K\left(C^0(\Omega)\right)$ its image under K. This can be viewed as an integral equation for δn^2 for a given δU.

It is well known [5] that these equations (Fredholm integral equations of first kind) are ill posed in the sense that for any admissible δU and its solution δn^2, for a given arbitrarily small ε there exists another function $\delta n^2{}_1$ for any δn^2 such that

$$\left\| K(\delta n^2{}_1) - \delta U \right\| < \varepsilon\,.$$

This observation implies the following:

1. Since the objective functions and constraints in both the above optimisation problems only depend explicitly (and continuously) on the field U, and not on the refractive index distribution n^2, small variations in the field imply small variations in the constraint and objective functions. Hence if n^2 is an **optimal** solution, then we can choose an arbitrarily different distribution $n^2{}_1$ which will also minimise the objective, and satisfy the constraints, to arbitrary precision.
2. Any optimisation scheme which in some way uses the above linearisation to find a local direction of descent will therefore manifest instabilities if it makes no additional regularity assumptions on the function search space, i.e. it attempts to search for an optimal solution n^2 in the entire space $C^0(\Omega)$.
3. This physically manifests the fact that the light propagation U is insensitive to variations in the refractive index profile which are smaller in scale than the propagating "wavelength" of U.
4. We can recover numerical stability to our optimisation problem by limiting our search space to a smaller function space, for example we could search for the optimal n^2 in the space $C^1(\Omega)$. Equation (19) would then become a mapping $K : C^1(\Omega) \to K\left(C^1(\Omega)\right)$ which has a continuous inverse in $C^0(\Omega)$, due to compactness of $C^0(\Omega)$ in $C^1(\Omega)$, if such an inverse exists. If this inverse does not exist, we can always consider the generalised inverse. This will be the basis of the alternative algorithm described in the next chapter.

8. The Inverse Problem approach.

This approach is based on a power conservation property for (6), namely that:

$$\mathrm{Im}\int_{z\in\Omega_z} U\frac{\partial\overline{U}}{\partial z}dx$$

is conserved along the waveguide. Hence the total input power must be the same as the total output power on the right plus the reflected power at the input:

$$1 = \sum_{k>0}\left|a_{-k}(0)\right|^2 + \sum_{k>0}\left|a_k(z_R)\right|^2$$

The second sum contains the coefficients of the modal expansion for $U(x, z_R; n)$ at the exit of the taper:

$$U(x, z_R; n) = \sum_{k=1}^{\infty} a_k(z_R) U_k^R$$

which are uniquely determined by (14) and (15) for a given refractive index distribution *n*. (from (15), the backward coefficients have already been set to zero). It follows that the power transmitted in the RHS fundamental mode, $|a_1(z_R)|^2$, is always smaller than 1, and equality is obtained only when:

$$\begin{bmatrix} |a_1(z_R)| \\ |a_2(z_R)| \\ |a_3(z_R)| \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix}. \tag{20}$$

We have therefore re expressed the optimisation problem of minimising $P(n)$ with solving (20), for the unknown refractive index distribution *n*. Viewed in this way, the problem is known to be ill posed, and is very similar in concept to inverse scattering problems [??]. Moreover the dependency on n $U(x, z_R; n)$ is non linear, and a solution may not even exist in reality, as there might not be a shape with finite length which can totally convey all the input power into the RHS fundamental mode (*P=1*). Nevertheless this formulation has the advantage that we are taking into account the values of $a_2(z_R), a_3(z_R), \ldots$ as well as $a_1(z_R)$. Although this is not strictly necessary, as power conservation implies that if $a_1(z_R) = 1$ then $a_2(z_R), a_3(z_R), \ldots = 0$, numerically it should help to drive the coefficients towards these desired values.
The other advantage is that we can leverage techniques used for solving non linear inverse problems. We now outline a typical approach using a Newton-Raphson like technique.
We rewrite (20) as

$$\mathbf{a}(n) = \mathbf{a_1}$$

with

$$\mathbf{a_1} \equiv \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix}; \mathbf{a}(n) \equiv \begin{bmatrix} |a_1(z_R)| \\ |a_2(z_R)| \\ |a_3(z_R)| \\ \vdots \end{bmatrix}$$

Next Newton step $n + \delta n$ is given by the linearised problem

$$\left.\frac{\partial \mathbf{a}}{\partial n}\right|_n \delta n = \mathbf{a_1} - \mathbf{a}(n) \tag{21}$$

The expression on the LHS is to be understood as a functional derivative. Note that such a δn is a descent direction for $\|\mathbf{a}(n) - \mathbf{a_1}\|_2$ (easily seen by taking its derivative and using (21)).
After parametrising $n(x, z) \to \mathbf{n} = \{n_1, n_2, \cdots, n_N\}$ the Newton step equation becomes

$$\left.\frac{\partial \mathbf{a}}{\partial \mathbf{n}}\right|_n \delta \mathbf{n} = \mathbf{a_1} - \mathbf{a}(\mathbf{n}) \tag{22}$$

where the Jacobian $\left.\frac{\partial \mathbf{a}}{\partial \mathbf{n}}\right|_n$ is a (M x N) matrix, M being the last retained mode coefficient.

Note that depending on the choice of N and M we may have an under determined or over determined system. Clearly the choice of N depends on the degree of refinement we require. The choice of M is related to the accuracy of the solution to the direct problem: of course, the more coefficients we keep in (14), the more accurate the solution to the direct problem. The choice of the number M of these coefficients to include in (20) may seem arbitrary (we may or may not include all coefficients included in solving the direct problem (14)), but the optimal choice may well be related to the resolution N [??].

Either way, we are now faced with the problem of choosing the $\delta\mathbf{n}$ which "best" satisfies (22), which may have no solution, or a whole subspace of solutions. However, it is not strictly necessary to SOLVE the above equation, but simply to find a $\delta\mathbf{n}$ that is still a direction of descent to the original problem. We now show that this may obtained by considering the Singular Value Decomposition of the jacobian:

$$\left.\frac{\partial \mathbf{a}}{\partial \mathbf{n}}\right|_{\mathbf{n}} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathbf{T}}$$

U = (M x N) , **V**=(N x N) are **unitary matrices** - M=dim(**a**), **D**=(N x N) matrix with (nonzero) **singular values** $\{d_1,\cdots,d_{\min(M,N)}\}$ on the diagonal.

This is an SVD of a matrix which arises from discretising an integral kernel, so since (21) is closely related to (19), we expect the singular values to decrease to zero: $d_k \to 0$ as $k \to \infty$.

This is the manifestation of the ill posed nature of the linearised inverse problem (19). We therefore choose the step $\delta\mathbf{n}$ given by:

(23) $$\delta\mathbf{n} = \mathbf{V}\mathbf{D}_{red}^{-1}\mathbf{U}^{\mathbf{T}}\left[\mathbf{a_1} - \mathbf{a}(\mathbf{n})\right]$$

This is just the generalised inverse, with **D** replaced by $\mathbf{D}_{red}$ - the matrix **D** with $d_k < d_{\max}\alpha$ set to zero for a given regularisation parameter $0 < \alpha < 1$.

The effect of the regularisation parameter is to eliminate the singular values that are considered to be too small, and that therefore lead to large numerical instabilities in the determination of $\delta\mathbf{n}$. This is directly related to regularisation by truncation of singular values of integral operators [5]. Note that this is one of many regularisation techniques that may be used. The other most notable one being Tikhonov regularisation [5] which would be equivalent to adding a regularisation factor $\alpha\delta\mathbf{n}$ on the lhs of (22). The latter would have been the algorithm of choice if for practical reasons it would be too lengthy to calculate the SVD, which would be the case if M,N were too large. Conventional wisdom indicates, however, that regularisation by truncation of singular values is preferable if the singular values are available.

In particular, the solution to (23) still gives a descent direction:

$$\begin{aligned}
\delta\left\|\mathbf{a_1} - \mathbf{a}(\mathbf{n})\right\|_2 &= -2(\mathbf{a_1} - \mathbf{a}(\mathbf{n})) \bullet \left.\frac{\partial \mathbf{a}}{\partial \mathbf{n}}\right|_n \delta\mathbf{n} \\
&= -2\mathbf{b} \bullet \left.\frac{\partial \mathbf{a}}{\partial \mathbf{n}}\right|_n \delta\mathbf{n} \quad ; \quad \mathbf{b} = \mathbf{a_1} - \mathbf{a}(\mathbf{n}) \\
&= -2\mathbf{b}^{\mathbf{T}}\left(\mathbf{U}\mathbf{D}\mathbf{V}^{\mathbf{T}}\right)\left(\mathbf{V}\mathbf{D}_{red}^{-1}\mathbf{U}^{\mathbf{T}}\mathbf{b}\right) \\
&= -2\left(\mathbf{U}^{\mathbf{T}}\mathbf{b}\right)^{\mathbf{T}}\mathbf{D}\mathbf{D}_{red}^{-1}\left(\mathbf{U}^{\mathbf{T}}\mathbf{b}\right) \\
&< 0
\end{aligned}$$

This is the basic requirement for a convergent root finding algorithm.

9. Numerical experiments using the inverse problem approach.

We use the above regularised descent estimate to implement the following classical search algorithm:

For given regularisation parameter $0 < \alpha < 1$ *and current estimate* $\mathbf{n}_k$*, next step given by:*

$$\delta\mathbf{n} = \mathbf{V}\mathbf{D}_{red}^{-1}\mathbf{U}^{\mathrm{T}}\left[\mathbf{a_1} - \mathbf{a}(\mathbf{n}_k)\right]$$

Where $\mathbf{U}$,$\mathbf{V}$ $\mathbf{D}_{red}$ *are obtained as described above.*
Do line search by minimising

$$F(\lambda) = \left\|\mathbf{a_1} - \mathbf{a}(n_k + \lambda\,\delta n)\right\|_2$$

and set

$$n_{k+1} = n_k + \lambda_{\min}\delta n$$

Repeat till $\left\|\mathbf{a_1} - \mathbf{a}(\mathbf{n}_k)\right\|_2$ *stops decreasing.*

At this stage we're only interested in studying the effectiveness of the search direction, its behaviour for various values of regularisation parameter, and comparison with performance of the direct optimisation approach. To this end we have run both the new and the previous optimisation algorithm for a fixed number of outer iterations (line searches), and have compared the power coupling $P = |a_1|^2$ resulting in each case.

The combined results for low and high resolutions are shown in the appendices.
The first significant fact to note is the variation of the solution found with respect to the regularisation parameter. In both cases (low and high resolution) there clearly seems to an optimal region for this parameter. This is a familiar result widely documented in inverse problem literature, and is due to the following reasons:

- If the regularisation parameter is close to 1, only large singular values are retained. These are the ones that give the smoother contribution to (23), thus resulting in a smoother correction $\delta\mathbf{n}$ at each outer iteration. However, as can be seen from the numerical results, the real solution might well be a highly irregular (non smooth) function. Convergence is therefore hampered by the poor correction given by the descent step.
- As the regularisation parameter is decreased, the smaller singular values are retained in (23), thus resulting in a $\delta\mathbf{n}$ which models better the final solution. However as regularisation parameter decreases further convergence slows down again due to increased numerical instabilities resulting in inclusion of excessively small singular values.

The second important point is that given the correct choice for the regularisation parameter, the algorithm's convergence rate, although being merely comparable to the classical optimisation approach for small resolutions (appendix 1), it becomes much better at high resolutions (appendix 2). This shows that indeed regularisation techniques can be used to accelerate convergence by filtering out the numerical instabilities associated with high resolutions.

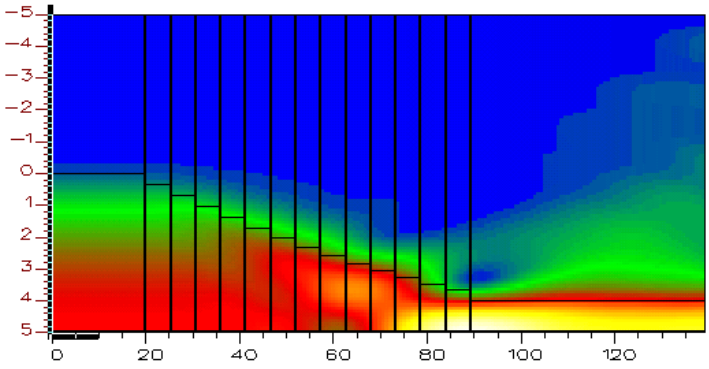
10. conclusions and outlook

As expected, convergence of classical descent algorithms slows down considerably with increasing refinement of shape discretisation. This initial study indicates that the Inverse Problem approach is less sensitive to increase in refinement. There is an optimal choice of

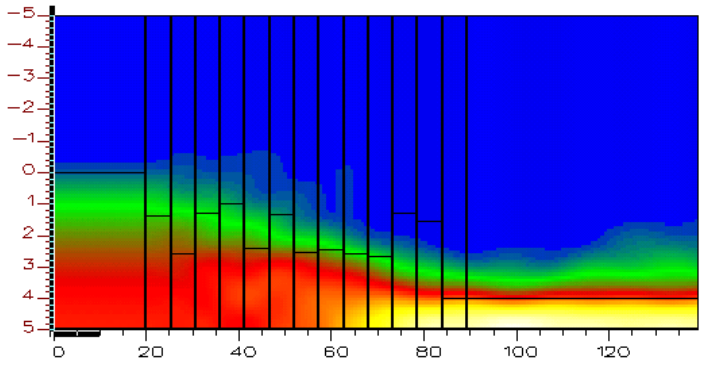
regularisation parameter: trade off between "smoothness" of Newton step (large regularisation parameter), and numerical instability (small regularisation parameter).
On the whole when large resolutions are used in the discretisation of the original problem, the algorithm based on the Inverse Problem approach exhibits much faster rates of convergence, given an appropriate choice of regularisation parameter.
There are of course still open questions.

- In practice a criterion is needed for the "optimal" choice of regularisation parameter at each Newton step. This is closely related to choosing the optimal regularisation parameter in solving inverse problems, and several criteria have been established for such a choice, in particular for linear problems [5]. The main difference is that in an inverse problem the optimal parameter is a function of the error estimate of the given data, while here it will depend on other criteria to be established, e.g. the error of the current estimate.
- The technique presented here is very similar in strategy to the one employed for non linear inverse problems. For the latter, more sophisticated regularisation techniques have been established which also involve the line search, and not just the search for a local descent direction [??]. This will provide further gains in computational efficiency.
- The Jacobian $\left.\frac{\partial \mathbf{a}}{\partial \mathbf{n}}\right|_{\mathbf{n}}$ was calculated using finite differences. Although this could not be avoided for the optimisation algorithm, in the inverse problem approach this can be approximated using Multidimensional Secant Methods (e.g. Broyden's Method). This would result in a dramatic reduction in the computational cost. Again, this is related to methods used for non linear inverse problems. The approximation needs to be chosen with care, and is also dependent on the regularization strategy employed [??].

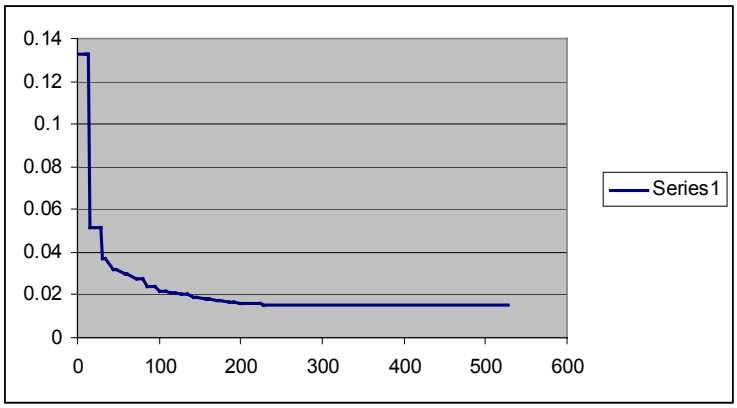
11. appendix 1: Optimisation using 13 subsections, 10 modes



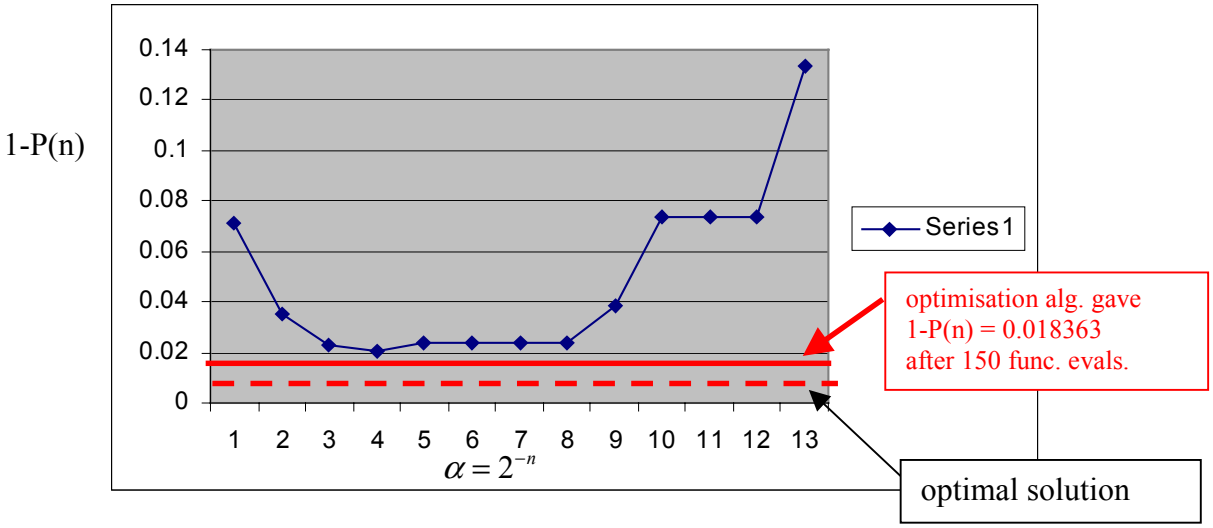
Initial shape. 1-P(n) = 0.1331



Optimal shape. 1-P(n) = 0.0150

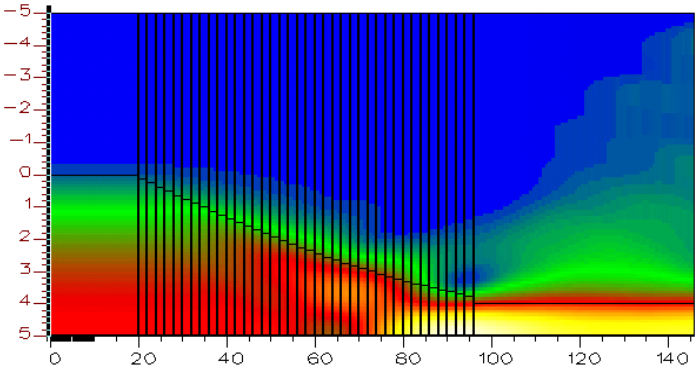


Convergence occurs at 1-P(n)= 0.0150 ,which is reached after 26 outer iterations.(370 func. evals.)

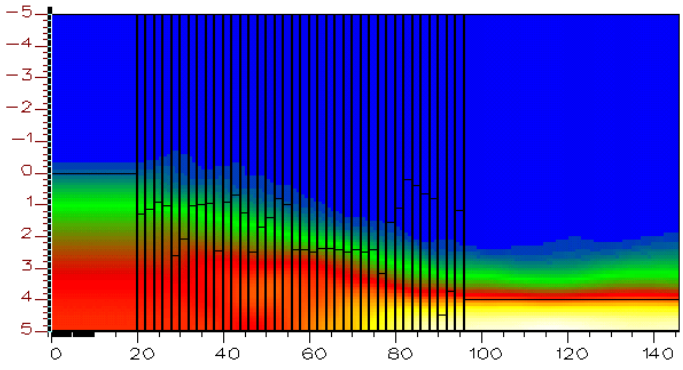


Graph of behaviour of IP algorithm
power loss / regularisation parameter after 5 line searches (~150 func. evals.)

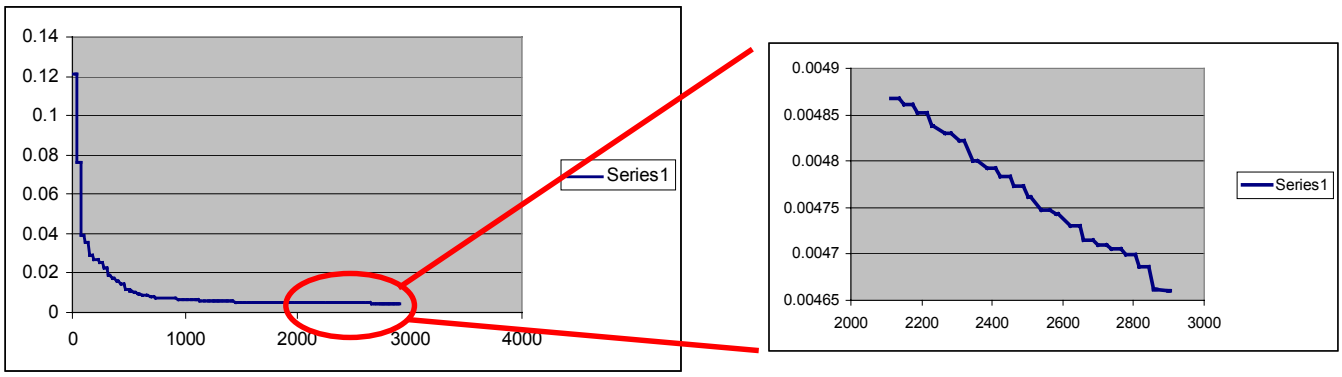
12. appendix 2: Optimisation using 48 subsections, 15 modes



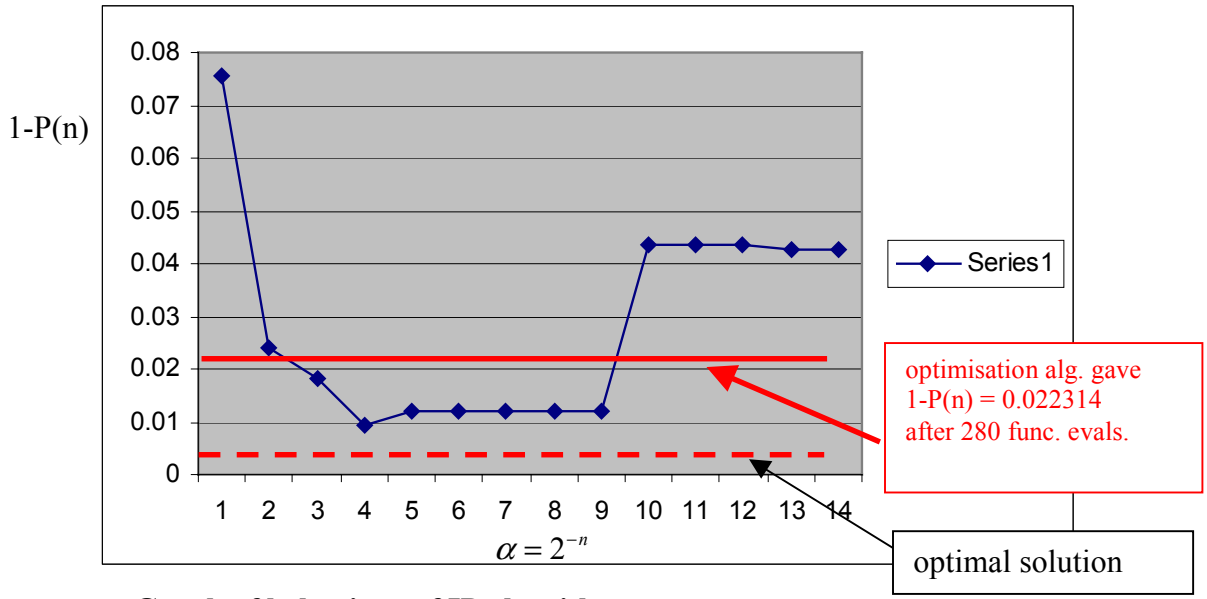
Initial shape. 1-P(n) = 0.12150



Optimal shape. 1-P(n) = 0.00466



Convergence is never reached.
1-P(n)= 0.00466, after 75 line searches.(2900 func evals)



Graph of behaviour of IP algorithm
power loss / regularisation parameter after 5 line searches (~280 func. evals.)

13. References

1. David C. Dobson, *"Optimal shape design of blazed diffraction gratings"* , Appl. Math. Opt., **40** (1999), 61-78.
2. Gang Bao and David C. Dobson, *"Modeling and optimal design of diffractive optical structures",* Surv. Math. Ind., **8** (1998), 37-62.
3. A.W. Snyder, J.D. Love *,"Optical Waveguide Theory"* , Chapman and Hall.
4. Direction Set (Powell's) Methods in Multidimensions: *"Numerical recipes"*
5. H.W.Engl, M.Hanke,A.Neubauer: *"Regularization of Inverse Problems"*, Kluwer Academic Publishers.