SIAM J. OPTIM.
Vol. 8, No. 1, pp. 82–104, February 1998

004

RECONSTRUCTION WITH NOISY DATA: AN APPROACH VIA EIGENVALUE OPTIMIZATION*

DOMINIKUS NOLL†

Abstract. We present a nonlinear inverse filtering approach to problems such as power spectrum estimation of stationary time series or deconvolution of a blurred image. The technique is based on eigenvalue optimization and a numerical treatment may therefore be obtained using primal-dual interior-point methods for semidefinite programming.

Key words. inverse problems, power spectrum estimation, image restoration, Fisher information, eigenvalue optimization, semidefinite programming, large scale problems, finite-element discretization

AMS subject classifications. Primary, 49N50; Secondary, 90C06

PII. S105262349629856X

1. Introduction. In this paper we discuss a nonlinear inverse filtering approach to inverse problems such as the restoration of blurred images or the estimation of power spectra of stationary time series. In many situations, these inverse problems are difficult to treat due to a sizable imperfection of the available data.

Let us consider the problem of estimating an unknown nonnegative function $u(x) \geq 0$ on some domain $\Omega \subset \mathbb{R}^n$, given a finite number of measurements

$$\int_\Omega a_k(x)u(x)\,dx \;=\; b_k, \quad k = 0, \dots, m, \tag{1.1}$$

with $a_0(x), \dots, a_m(x)$ a given set of weight functions. This problem, with only a finite set of data available, is clearly underdetermined. More seriously, however, the data b_k may suffer from measurement errors and may make inversion of (1.1) a difficult task. As an example for such behavior consider the following.

Example 1.1 (power spectrum estimation). Here we wish to reconstruct the spectral density $u(x) \geq 0$, $x \in (-\pi, \pi)$ of a real stationary time series (X_t), based on knowledge of the first $m + 1$ sample autocovariances $\hat{b}_0, \dots, \hat{b}_m$, which have been obtained from a realization $x_0, \dots, x_N$ of (X_t) (with $m << N$), using the statistic

$$\hat{b}_k = \frac{1}{N} \sum_{t=0}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x}), \qquad k = 0, \dots, m.$$

Equivalently (cf. [8]), we have to exhibit $u(x) \geq 0$ satisfying

$$\int_{-\pi}^{\pi} \cos kx\, u(x)\,dx = \hat{b}_k, \quad k = 0, \dots, m, \tag{1.2}$$

which is precisely a problem of type (1.1). It is well known (cf. [8]) that the sample autocovariances $\hat{b}_k$ are increasingly unstable for large k, and that a direct inversion of (1.2) based on the Fourier series of $u(x)$ usually fails. □

*Received by the editors January 31, 1996; accepted for publication (in revised form) June 12, 1996.
http://www.siam.org/journals/siopt/8-1/29856.html
†Université Paul Sabatier, Laboratoire Approximation et Optimisation, 118 route de Narbonne, 31062 Toulouse Cedex, France (noll@dumbo.ups-tlse.fr).

As a second example, let us consider the following.

Example 1.2. (image restoration) Consider an image, represented by its distribution of gray levels $u(x) \geq 0$ over a region $\Omega \subset \mathbb{R}^2$. Suppose that during transmission through a news channel, the original image has suffered from various sources of degradation, some of them being signal dependent, some caused by random pixel noise. Assuming a linear model, the observed image $v(x) \geq 0$ may be represented by

$$v(x) = \int_\Omega q(x-y)\, u(y)\, dy + e(x), \qquad x \in \Omega, \tag{1.3}$$

where the convolutional term is responsible for the signal dependent blurring ($q(x)$ being referred to as the *point spread function* or simply the *mask*), and where $e(x)$ represents the signal independent random noise. Suppose now that the recorded image $v(x)$ has been sampled at the nodes $x_1, \ldots, x_m$ of a rectangular grid on Ω. Then the problem of deconvolving (1.3) will fit with the scheme (1.1), with the weight functions being $a_0 \equiv 1$, $a_k(x) = q(x_k - x)$, and the data $b_0 = 1$, $b_k = v(x_k)$. However, if the noise contribution $e(x)$ becomes sizable, it may be necessary to relax equation (1.1), $Au = b$, $\int_\Omega u\, dx = 1$, replacing it with a *tolerance model* of the form

$$|Au - b| \leq \epsilon, \qquad \int_\Omega u(x)\, dx = 1, \tag{1.4}$$

where $b = (b_1, \ldots, b_m)$, $Au = (\int_\Omega a_1 u dx, \ldots, \int_\Omega a_m u dx)$, and $|\cdot|$ denotes the Euclidean norm on $\mathbb{R}^m$. Various strategies for selecting an appropriate tolerance ϵ are known. For instance, if the noise variance σ^2 is known—a hypothesis which is often met in practice—the strong law of large numbers implies

$$\frac{1}{m} \sum_{i=1}^m e_i^2 \to \sigma^2 \quad (m \to \infty),$$

which in view of $|Ax - b|^2 = |e|^2 = \sum_{i=1}^m e_i^2 \approx m\sigma^2$ suggests the choice $\epsilon = \sigma\sqrt{m}$ (see [31, 32] for a detailed discussion). $\square$

A successful idea to stabilize an inversion problem of type (1.1) or (1.4) in the presence of noisy data b_k is to introduce a performance index

$$\mathcal{I}(u) = \int_\Omega h(x, u(x), \nabla u(x))\, dx$$

related to the problem under discussion, and to accept as a valid reconstruction the unique $u(x) \geq 0$ that minimizes $\mathcal{I}(u)$ under the constraints (1.1) (resp., (1.4)). Naturally, the strategy for designing $\mathcal{I}(u)$ should be justified by both theoretical models and performance in practice. While many different indices $\mathcal{I}(u)$ are being used in practice, a common feature seems to be their convexity.

In the power spectrum estimation problem (1.2), an objective often used in practice is the Burg entropy $\mathcal{I}(u) = -\int_\Omega \log u(x)\, dx$, which leads to a fast and, surprisingly, linear inverse method (see [7, 27]). For a theoretic justification see, e.g., [30]. By contrast, a nonlinear inverse filter is obtained when the Boltzmann–Shannon entropy/information measure $\mathcal{I}(u) = \int_\Omega u(x) \log u(x)\, dx$ is used, the discussion about the preference between the two choices being controversial; cf. [24].

The Boltzmann–Shannon entropy has also been tested for the image restoration problem (1.4), leading to a nonlinear inverse filter whose performance was found to be significantly superior to more standard linear inverse filters used for this type of problems; cf. [31, 32, 22, 37].

One such linear inverse filter for the image restoration problem (1.4) is based on the performance index $\mathcal{I}(u) = \int_\Omega |\Delta u(x)|^2 dx$. For this model, a fast algorithmic solution using the 2D fast Fourier transform was presented by Hunt [23]. Other inverse filters could be based on inverse Wiener or Kalman filtering, but these methods require knowledge of the cross spectrum of the true image and the noise, and knowledge of the dynamics of the input process, two pieces of information which typically would be lacking in practical situations; cf. [2, 26].

In [36, 5, 6], a nonlinear inverse method for (1.1) based on the *Fisher information*

$$\mathcal{I}(u) = \int_\Omega \frac{|\nabla u(x)|^2}{u(x)} dx \tag{1.5}$$

was proposed. In the one-dimensional setting, sufficiently fast algorithms for the power spectrum estimation problem (1.2) and for problems with polynomial weights have been presented in [6, 7]. These techniques, however, are not suited for higher dimensions, and so far a successful numerical approach in higher dimensions was lacking. The main issue of the present paper is to present one such approach based on techniques from eigenvalue optimization. Notice that it seems highly desirable to have codes based on the Fisher information index (1.5), in particular for higher dimensional problems like (1.4), since Fisher information was found to perform well in the one-dimensional case and in particular was seen to outperform the Boltzmann–Shannon objective in quite a number of situations, as has been reported in [6, 7].

The purpose of the present paper is to analyze the restoration programs (1.5), (1.1) (resp., (1.5), (1.4)) using convex programming duality. This leads to their reformulation as eigenvalue optimization (resp., semidefinite programming) problems, the main idea being to postpone discretization to the latest possible moment. While sections 2 and 3 present the details of the analysis, section 4 gives a convergence result for the finite element discretization. The final section 5 presents some numerical experiments.

Our notation in sections 2 and 3 follows standard references on variational problems and Sobolev spaces like [14, 1, 39]. Readers who find the technical arguments in sections 2 and 3 (leading to the central Theorems 3.4, 3.5) too cumbersome to follow in full detail are invited to obtain these dual programming formulations by formally applying convex duality to the Lagrangian (3.1). This gives the correct answers right away, and reading the following sections 4 and 5 is still possible.

Let us close this introductory section with a typical application for using the Fisher information as a performance index. For further motivation see [36] or [5, App. I].

Example 1.3 (robustness and higher order statistics (see [40])). Consider the problem of estimating the parameters a_i, b_j of a linear input-output ARMA system of the form

$$y_t + a_1 y_{t-1} + \cdots + a_m y_{t-m} = x_t + b_1 x_{t-1} + \cdots + b_n x_{t-n},$$

where $t \in \mathrm{Z}$, x_t is the input, and y_t the output sequence, and where the orders n, m are known. Suppose that the input is white noise with mean zero, variance σ^2, and probability density p.

Given a sample $y_1, \dots, y_N$ of outputs, the idea of maximum likelihood estimation of $\theta = (a_1, \dots, a_m, b_1, \dots, b_n)$ is to choose it to maximize a likelihood function of the form

$$L(\theta) = -\sum_{t=1}^{N} \ell\big(r_t(\theta)\big),$$

where $r_i(\theta) = y_i - x_i - b_1 x_{i-1} - \cdots - b_n x_{i-n} + a_1 y_{i-1} + \cdots + a_m y_{i-m}$ is the ith residual.

How to find the appropriate $\ell(r)$? This could be based on the following observation, essentially made already by Fisher [17]; see also Doob [15]. Under some mild assumptions on the input distribution p and the smoothness of ℓ, the estimates $\hat{\theta}_N$ based on ℓ are asymptotically normally distributed, i.e.,

$$\sqrt{N}(\hat{\theta}_N - \theta^*) \sim \mathcal{N}(0, \alpha A/\sigma^2) \qquad \text{as } N \to \infty.$$

Here the positive definite $(n+m) \times (n+m)$ matrix $A = A(\theta^*)$ depends only on the true parameter vector θ^*, while the scaling factors $\sigma^2 = \sigma^2(p)$ and $\alpha = \alpha(\ell, p)$ depend on the choice of ℓ and the underlying density p of the input law:

$$\alpha(\ell, p) = \int \ell'^2 p \, dx \bigg/ \bigg(\int \ell'' p \, dx \bigg)^2$$

and could therefore be influenced by our choice of ℓ.

Let us assume in the following that the elements p of the class $\mathcal{P}$ have identical variance $\sigma_0^2 = \sigma^2(p)$. Designing a robust estimate now leads us to consider the worst case p_0 among the probability densities $p \in \mathcal{P}$ and then improve the asymptotic efficiency of the estimate $\hat{\theta}_N$ by choosing ℓ to minimize $\sup_p \alpha(\ell, p) = \alpha(\ell, p_0)$. That is, solve the minimax problem

$$\inf_{\ell} \sup_{p \in \mathcal{P}} \alpha(\ell, p).$$

Switching the infimum and the supremum, the inner minimization (now over the ℓ for a fixed p) admits a solution of the form $\ell = -\rho \log p$ (ρ constant). As $\alpha(\ell, p) = \alpha(\frac{1}{\rho}\ell, p)$, the problem is now equivalent to maximizing $\alpha(-\log p, p) = \mathcal{I}(p)^{-1}$, or equivalently, minimizing $\mathcal{I}(p)$, the Fisher information, over the admissible family $\mathcal{P}$ of probability densities.

The classical choice is of course $\ell_0(r) = r^2$. This is explained by choosing as the family $\mathcal{P}$ of admissible input distributions all laws p with given moments up to order 2. Indeed, the solution p_0 to the problem of minimizing the Fisher information $\mathcal{I}(p)$ subject to the constraints $\int p(x)\,dx = 1$, $\int p(x) x \, dx = 0$, and $\int p(x) x^2 \, dx = \sigma^2$ is just the Gaussian with mean zero and variance σ^2, which leads to $\ell_0(r) = -\log p_0(r) = \mathcal{O}(r^2)$.

Suppose now that the white noise processes x_t has higher-order moments, say, up to order $m > 2$, m even. In this case, the classical choice $\ell_0(r) = r^2$ should be replaced by $\ell_0 = -\log p_0$, where p_0 now minimizes Fisher information over the set $\mathcal{P}$ of densities p with given moments up to order m. The results presented by Zivojnovic [40] suggest that this choice of ℓ_0 will indeed improve the asymptotic efficiency of the maximum likelihood approach. The maximization of $L(\theta)$ would then, with no explicit ℓ_0 available, have to be performed numerically. $\square$

2. Variational formulation. In this section we will analyze the Fisher variational problem. First observe that the precise definition of the Fisher integrand $h(u, p)$

is

$$h(u,p)=\begin{cases}|p|^2/u & \text{if } u>0,\\ 0 & \text{if } u=0 \text{ and } p=0,\\ +\infty & \text{otherwise,}\end{cases} \tag{2.1}$$

which is a normal convex integrand in the sense of Rockafellar [35]. The variational problem we consider is now

$$\text{(P)}\qquad \begin{array}{ll}\text{minimize} & \mathcal{I}(u)=\int_\Omega h(u(x),\nabla u(x))\,dx=\int_\Omega \frac{|\nabla u(x)|^2}{u(x)}dx\\ \text{subject to} & \int_\Omega a_k(x)u(x)\,dx=b_k,\ k=0,\dots,m,\\ & u(x)\ge 0,\ u\in W^{1,1}(\Omega).\end{array}$$

In tandem with (P) and motivated by the image restoration problem (1.4) we also consider the relaxed Fisher program

$$(\mathrm{P}_r)\qquad \begin{array}{ll}\text{minimize} & \mathcal{I}(u)=\int_\Omega \frac{|\nabla u(x)|^2}{u(x)}dx\\ \text{subject to} & u(x)\ge 0,\ u\in W^{1,1}(\Omega),\\ & \int_\Omega a_0(x)u(x)\,dx=b_0,\\ & |Au-b|\le\epsilon,\end{array}$$

where $Au=(\int_\Omega a_1u\,dx,\dots,\int_\Omega a_mu\,dx)$ and $b=(b_1,\dots,b_m)$. Notice that (P) and (P_r) are convex programs.

Throughout the following we shall assume that programs $(\mathrm{P}),(\mathrm{P}_r)$ are feasible. Moreover, to avoid pathological situations, we shall assume throughout that $u\equiv 0$ is not an optimal solution of $(\mathrm{P}),(\mathrm{P}_r)$. Let us introduce the substitution

$$u(x)=v(x)^2,\quad 2v(x)\nabla v(x)=\nabla u(x), \tag{2.2}$$

which turns problems (P), (P_r) into the equivalent but nonconvex programs $(\tilde{\mathrm{P}}),(\tilde{\mathrm{P}}_r)$:

$$(\tilde{\mathrm{P}})\qquad \begin{array}{ll}\text{minimize} & \mathcal{J}(v)=4\int_\Omega |\nabla v(x)|^2\\ \text{subject to} & v\in W^{1,2}(\Omega),\\ & \int_\Omega a_k(x)v(x)^2dx=b_k,\ k=0,\dots,m,\end{array}$$

resp.,

$$(\tilde{\mathrm{P}}_r)\qquad \begin{array}{ll}\text{minimize} & \mathcal{J}(v)=4\int_\Omega |\nabla v(x)|^2dx\\ \text{subject to} & v\in W^{1,2}(\Omega),\ |Av^2-b|\le\epsilon,\\ & \int_\Omega a_0(x)v(x)^2dx=b_0.\end{array}$$

We are now in the position to prove existence and unicity of solutions of (P) and (P_r).

PROPOSITION 2.1. *Let $a_0,\dots,a_m\in\mathcal{L}^\infty(\Omega)$, and assume $\int_\Omega a_0(x)dx\ne 0$. Let Ω be a bounded domain of class $\mathcal{C}^{0,1}$. Then program* (P) *has a unique solution in $W^{1,1}(\Omega)$, and similarly,* (P_r) *has a unique solution in $W^{1,1}(\Omega)$.*

Proof. (1) We prove existence of a solution in the case of program (P_r). Using the transform (2.2), we may consider $(\tilde{\mathrm{P}}_r)$. Let $v_n \in W^{1,2}(\Omega)$ be a minimizing sequence for $(\tilde{\mathrm{P}}_r)$,

$$\mathcal{J}(v_n) \to \min, \quad |Av_n^2 - b| \le \epsilon, \quad \int_\Omega a_0 v_n^2 dx = b_0. \tag{2.3}$$

Observe that $|v_n|$ is as well a minimizing sequence, since $|\nabla v_n| = |\nabla |v_n||$ a.e. Hence we may assume $v_n \ge 0$.

The first condition in (2.3) implies the boundedness of ∇v_n in $\mathcal{L}^2$. Suppose v_n was not bounded in $W^{1,2}(\Omega)$. Then, since ∇v_n is bounded we may decompose $v_n = w_n + c_n$, where w_n is a bounded sequence in $W^{1,2}(\Omega)$, and the c_n are constants tending to ∞. Indeed, let $c_n = \int_\Omega v_n dx$ and $w_n = v_n - c_n$. Then $\int_\Omega w_n dx = 0$ in tandem with the boundedness of $\|\nabla w_n\|_2 = \|\nabla v_n\|_2$ implies boundedness of w_n in $\mathcal{L}^2$ (using a Poincaré inequality like, for instance, Ziemer [39, Thm. 4.2.1]). But then

$$b_0 = \int_\Omega v_n^2 a_0 \, dx = c_n^2 \int_\Omega a_0 \, dx + \mathcal{O}(c_n),$$

which implies $\int_\Omega a_0 \, dx = 0$, a contradiction. This implies the boundedness of v_n in $W^{1,2}$. We select a weakly convergent subsequence, also noted v_n, i.e., $v_n \rightharpoonup v$. The convexity of $\mathcal{J}$ implies its weak lower semicontinuity, hence

$$\liminf \mathcal{J}(v_n) \ge \mathcal{J}(v),$$

and so if v turns out to be feasible, it must be a minimum for $(\tilde{\mathrm{P}}_r)$.

The fact that Ω is of class $\mathcal{C}^{0,1}$ implies the compactness of the inclusion $W^{1,2}(\Omega) \to \mathcal{L}^2(\Omega)$; cf. [21]. Hence v_n has a norm convergent subsequence in $\mathcal{L}^2$, also noted v_n, i.e., $v_n \to v$ in $\mathcal{L}^2$. This implies $v_n^2 \to v^2$ in $\mathcal{L}^1$, since $\|v_n^2 - v^2\|_1 \le \|v_n - v\|_2 \|v_n + v\|_2 \to 0$. It also implies $v \ge 0$. Now $\epsilon \ge |Av_n^2 - b| \to |Av^2 - b|$ and $b_0 = \int a_0 v_n^2 dx \to \int a_0 v^2 dx$ together imply the feasibility of v. With v a minimum for $(\tilde{\mathrm{P}}_r)$, $u = v^2$ is a minimum for (P_r).

(2) To prove unicity, we may return to the original program (P_r). The argument is now verbally the same as for the one-dimensional case given in [5, Thm. 2.1]. Notice that by contrast the solution of $(\tilde{\mathrm{P}}_r)$ is not unique, but a unique nonnegative solution exists. ∎

In order to obtain more information on the optimal solution of (P) (resp., (P_r)), we shall need a Lagrange multiplier result. For this we recall the concept of a pseudo-Haar family which was introduced in [4]: Call $a_0, \ldots, a_m$ pseudo-Haar if $a_0|A, \ldots, a_m|A$ are linearly independent on any measurable set $A \subset \Omega$ with positive measure.

PROPOSITION 2.2. *Suppose that the weight functions $a_0, \ldots, a_m$ are pseudo-Haar and of class $\mathcal{C}^{t-1}$ for some $t \ge 1$. Suppose that Ω is bounded and of class $\mathcal{C}^{t+1}$. Let $\bar{v} \in W^{1,2}(\Omega)$ be the unique nonnegative optimal solution of program $(\tilde{P})$ (resp., $(\tilde{\mathrm{P}}_r)$). Then there exist real numbers $\bar{\lambda}_0, \ldots, \bar{\lambda}_m$ such that $\bar{v}$ satisfies the linear elliptic PDE with Neumann boundary conditions:*

$$-\Delta \bar{v}(x) + \sum_{k=0}^m \bar{\lambda}_k a_k(x) \bar{v}(x) = 0 \ \ on\ \Omega, \tag{2.4}$$

$$\frac{\partial \bar{v}(x)}{\partial n_\Omega} = 0 \ \ on\ \partial\Omega. \tag{2.5}$$

Moreover, $\bar{v} \in W^{1+t,2}(\Omega)$*, and in particular, for* $t \geq k + \frac{n}{2} - 1$*, we have* $\bar{v} \in \mathcal{C}^k(\bar{\Omega})$.

Proof. We give the proof in the case of program $(\tilde{\mathrm{P}}_r)$. The argument for $(\tilde{\mathrm{P}})$ is similar. Introducing an auxiliary variable $e \in \mathbb{R}^m$, program $(\tilde{\mathrm{P}}_r)$ takes the equivalent form

$$\begin{aligned} \text{minimize} \quad & f(v,e) = \int_\Omega |\nabla v(x)|^2 dx \\ \text{subject to} \quad & g_0(v,e) = \int_\Omega a_0(x) v(x)^2 dx - b_0 = 0, \\ & g_i(v,e) = \int_\Omega a_i(x) v(x)^2 dx - b_i - e_i = 0, \quad i = 1, \dots, m, \\ & g_{m+1}(v,e) = \epsilon^2 - |e|^2 \geq 0. \end{aligned}$$

We claim that there exist Lagrange multipliers $\bar{\lambda}_0, \dots, \bar{\lambda}_m$ and $\bar{\lambda}_{m+1} \geq 0$ such that the optimal pair $(\bar{v}, \bar{e})$, with $\bar{e}_i = \int a_i \bar{v}^2 dx - b_i$, satisfies

1. $\nabla_v f(\bar{v}, \bar{e}) + \sum_{k=0}^m \bar{\lambda}_k \nabla_v g_k(\bar{v}, \bar{e}) = 0$;
2. $-\bar{\lambda}_i - 2\bar{\lambda}_{m+1} \bar{e}_i = 0, \quad i = 1, \dots, m$;
3. $\bar{\lambda}_{m+1}(\epsilon^2 - |e|^2) = 0$.

To prove this, we need a Lagrange multiplier result in a Banach space setting. For instance, the multiplier rule obtained by Ginsburg and Ioffe [20, Thm. 3.1] applies to the above program. (Notice that this result is obtained through techniques of nonsmooth analysis and needs no constraint qualification. In particular, the linear independence of the a_i is not needed for this argument.) Now let the associated Lagrangian be defined as

$$L(v, e; \mu, \lambda_0, \dots, \lambda_m, \lambda_{m+1}) = \mu f(v,e) + \sum_{k=0}^{m+1} \lambda_k g_k(v,e);$$

then the quoted result provides multipliers $\bar{\mu} \geq 0, \bar{\lambda}_0, \dots, \bar{\lambda}_m, \bar{\lambda}_{m+1} \geq 0$ not all zero such that $\bar{\lambda}_{m+1} g_{m+1}(\bar{v}, \bar{e}) = 0$ and such that

$$(0,0) \in \partial_a L(\bar{v}, \bar{e}; \bar{\mu}, \dots, \bar{\lambda}_{m+1}). \tag{2.6}$$

Here ∂_a refers to the *approximate subdifferential* (in the sense of [20]) with respect to (v,e). Notice, however, that the functions f, g_k are all Fréchet differentiable with respect to the variables v, e, so ∂_a coincides with the Fréchet derivative. Evaluating (2.6) therefore, along with the constraints, gives the following conditions:

1.′ $\bar{\mu} \nabla_v f(\bar{v}, \bar{e}) + \sum_{k=0}^m \bar{\lambda}_k \nabla_v g_k(\bar{v}, \bar{e}) = 0$;
2.′ $-\bar{\lambda}_i - 2\bar{\lambda}_{m+1} \bar{e}_i = 0, \quad i = 1, \dots, m$;
3.′ $\bar{\lambda}_{m+1} g_{m+1}(\bar{v}, \bar{e}) = 0$.

This is almost what we wish to prove, and it remains to check that the case $\bar{\mu} = 0$ may be excluded, i.e., that the problem is normal. It is this part of the proof which requires our qualification hypothesis.

Suppose we had $\bar{\mu} = 0$. Then we must have $\bar{\lambda}_{m+1} > 0$. Indeed, by item $2'$, $\bar{\lambda}_{m+1} = 0$ implied $\bar{\lambda}_1 = \cdots = \bar{\lambda}_m = 0$. Hence $\bar{\lambda}_0 \neq 0$. But item $1'$ now reads as $\bar{\lambda}_0 \int_\Omega a_0(x) \bar{v}(x) h(x)\, dx = 0$ for all test functions h, hence $a_0 \bar{v} = 0$, and since $\bar{v} \neq 0$ this contradicts the pseudo-Haar assumption. So we must have $\bar{\lambda}_{m+1} > 0$.

Consequently, item $1'$ with $\bar{\mu} = 0$ gives us $\sum_1^m \bar{\lambda}_k \nabla_v g_k(\bar{v}, \bar{e}) = 0$, and by the definition of the Fréchet derivative of the g_k with respect to v, this implies

$$\sum_{k=1}^m \bar{\lambda}_k \int_\Omega a_k(x) \bar{v}(x) h(x)\, dx = 0$$

for all test functions $h \in \mathcal{C}^1(\Omega)$. Hence $\sum_1^m \bar\lambda_k a_k \bar v \equiv 0$. Since the a_k are assumed pseudo-Haar, and since $\bar v$ does not vanish identically, we deduce $\bar\lambda_1 = \cdots = \bar\lambda_m = 0$. Inserting this in item $2'$ gives $\bar e_1 = \cdots = \bar e_m = 0$, as $\bar\lambda_{m+1}$ had been agreed to be strictly positive. But then the inequality constraint could not have been active, and by item $3'$, we had to have $\bar\lambda_{m+1} = 0$, a contradiction. This reasoning proves $\bar\mu > 0$, and we may assume $\bar\mu = 1$.

We now exploit item 1 in the appropriate sense. Observe that f, g_i are all Fréchet differentiable on $W^{1,2}(\Omega) \times \mathbb{R}^m$ with $\langle \nabla_v f(v,e), h\rangle_{W^{1,2}} = \langle \nabla v, \nabla h\rangle_{\mathcal{L}^2}$ and $\langle \nabla_v g_k(v,e), h\rangle_{W^{1,2}} = \langle a_k \bar v, h\rangle_{\mathcal{L}^2}$. This implies that $\bar v$ is a weak solution of the Neumann problem (2.4), (2.5). Applying a boundary regularity result like [21, Thms. 9.1.15, 9.1.17] implies that the solution $\bar v$ is in fact in $W^{1+t,2}(\Omega)$, and using Green's formula, the equation (2.4) with Neumann boundary condition (2.5) are satisfied.

Finally, observe that the Sobolev embedding theorem gives $W^{1+t,2}(\Omega) \subset \mathcal{C}^k(\bar\Omega)$ whenever $t+1 > k + n/2$ (cf. [21, Thm. 6.2.30]), and this implies the last statement. $\square$

Remarks. 1) For $\mathcal{C}^\infty$ weight functions a_i, the classical inner regularity theory implies that the solution $\bar v$ of $(\tilde{\mathrm{P}})$ (resp., $(\tilde{\mathrm{P}}_r)$) is of class $\mathcal{C}^\infty(\Omega)$; see [21].

2) For applications it seems more natural to consider $\mathcal{C}^{0,1}$ domains, which includes, in particular, polyhedral domains used for finite elements. Here at least in dimensions $n = 2, 3$, and for smooth a_i, the generalized solution is in $\mathcal{C}^\infty(\Omega) \cap \mathcal{C}^{0,\alpha}(\bar\Omega)$ for some Hölder exponent $0 < \alpha < 1$, so (2.4) is satisfied classically, but (2.5) has to be interpreted in the variational sense. The Hölder boundary regularity here may be derived from Theorem 3.2 of Murthy and Stampacchia [29].

3) In the imaging case $n = 2$, choosing $t \geq 2$ implies that the solution $\bar v$ is of class $\mathcal{C}^1$. The equation (2.4) and the boundary condition (2.5) are then satisfied in the classical sense.

A case of particular interest is Example 1.3. Here $\Omega = \mathbb{R}$, and the weight functions are smooth but unbounded, so Proposition 2.2 does not apply directly. We have the following result, which is not the most general one but covers the type of applications motivated by Example 1.3.

Corollary 2.3. *Consider the problem* (P) *of minimizing Fisher information* $\mathcal{I}(u)$ *subject to the moment constraints*

$$\int_{-\infty}^{\infty} x^k u(x)\,dx = b_k, \quad k = 0, \ldots, m.$$

Then if $m = 2r$ is even, problem (P) *has a unique solution $\bar u$ which is positive on* $\mathbb{R}$. *Moreover, there exist multipliers $\bar\lambda_0, \ldots, \bar\lambda_{2r}$ such that $\bar v = \bar u^{1/2}$ is analytic and satisfies the linear differential equation with boundary conditions*

$$v''(x) + \sum_{k=0}^{2r} \bar\lambda_k x^k v(x) = 0, \qquad v(\pm\infty) = 0.$$

Proof. 1) We first prove existence. As before, we solve the transformed problem $(\tilde{\mathrm{P}})$. Consider the weighted Sobolev space $\mathcal{W}$ of functions $v \in W^{1,2}(\mathbb{R})$ satisfying $\int v^2(x) x^{2r}\,dx < +\infty$. A corresponding Hilbert space norm is

$$\|v\| = \left(\int_{-\infty}^{\infty} v'(x)^2 dx + \int_{-\infty}^{\infty} v(x)^2 x^{2r}\,dx \right)^{1/2}.$$

Reasoning as in Proposition 2.1, consider a minimizing sequence v_n for $(\tilde{\mathrm{P}})$. Then $\|v_n'\|_2 = \mathcal{O}(1)$ is clear. Now $\int x^{2r} v_n(x)^2 dx = b_{2r}$ implies the boundedness of v_n in $\mathcal{W}$. It is this part of the proof which requires m to be even. Selecting a weakly convergent subsequence, the rest of the proof remains unchanged.

2) The second part of the argument leading to the Euler equation is now exactly the same as in Proposition 2.2. Notice, however, that the elements of $\mathcal{W}$ vanish at infinity, so the boundary value problem is now actually a Dirichlet problem, which explains the boundary condition $\bar{v}(\pm\infty) = 0$. As the weight functions are analytic, the same must be true for the solution $\bar{v}$.

3) It remains to argue that $\bar{v}$ is positive on $\mathbb{R}$. This may be shown by reasoning similar to [5, Thm. 3.2(1)]. ◻

Remark. Notice that for m odd, the Fisher problem with moment constraints up to order m in general fails to have a solution. In the case $m = 1$, this may be easily seen, e.g., for $b_0 = 1, b_1 = 0$ by constructing a sequence of symmetric $u_n > 0$ satisfying $\int u_n dx = 1$ and $\mathcal{I}(u_n) \to 0$. By symmetry, one always has $\int x u_n(x)\, dx = 0$.

3. Exploiting convexity. On transforming the programs $(\mathrm{P}), (\mathrm{P}_r)$ into the equivalent form $(\tilde{\mathrm{P}}), (\tilde{\mathrm{P}}_r)$, we were able to prove the existence of solutions and we obtained necessary optimality conditions using a Lagrange multiplier result. However, the convexity of the problems was not preserved by (2.2), so we lost some information, which turns out to be crucial for an appropriate numerical formulation. In the present section, we will therefore apply techniques from convex duality theory to obtain the missing information.

Introducing dummy variables p, e, program (P_r) takes the equivalent form

$$\begin{array}{ll} \text{minimize} & \mathcal{I}(u,p) := \displaystyle\int_\Omega h(u(x), p(x))\, dx = \int_\Omega \frac{|p(x)|^2}{u(x)} dx \\ \text{subject to} & p = \nabla u,\ Au - b - e = 0,\ u \in W^{1,2}(\Omega), \\ & u(x) \geq 0,\ \displaystyle\int_\Omega a_0(x) u(x)\, dx = b_0, \\ & \epsilon^2 - |e|^2 \geq 0. \end{array}$$

Recall that in section 2, the setting chosen for the programs $(\mathrm{P}), (\mathrm{P}_r)$ was $W^{1,1}(\Omega)$. However, as Proposition 2.2 shows the optimal solution to lie in a better space, restricting (P_r) to $W^{1,2}(\Omega)$ is justified and will provide a more convenient setting for the reasoning in this section.

The Lagrangian associated with this formulation is

$$\begin{aligned} L(u,p,e;w,\lambda,\mu,\nu) = {} & \mathcal{I}(u,p) + \langle w, \nabla u - p\rangle + \lambda \cdot (Au - b - e) \\ & + \mu\left(\int_\Omega a_0(x) u(x)\, dx - b_0\right) + \nu(\epsilon^2 - |e|^2), \end{aligned} \tag{3.1}$$

where the state constraint $u \geq 0$ need not be modeled explicitly into the Lagrangian, since it is implicit in the definition of the Fisher functional $\mathcal{I}(u,p)$. With (3.1) program (P_r) takes the equivalent form

$$\begin{aligned} V(\mathrm{P}_r) = \inf\Big\{ & \sup\left\{L(u,p,e;w,\lambda,\mu,\nu) : w \in \mathcal{L}^2,\ \lambda \in \mathbb{R}^m,\ \mu \in \mathbb{R}, \nu \geq 0\right\} : \\ & u \in W^{1,2},\ p \in \mathcal{L}^2,\ |e| \leq \epsilon \Big\}. \end{aligned}$$

The dual program (D_r) associated with (P_r) is then defined as

$$V(\mathrm{D}_r) = \sup\left\{ \inf_{u,p,e} L(u,p,e;w,\lambda,\mu,\nu) : w,\lambda,\mu,\nu \right\}$$

with the obvious relation $V(\mathrm{D}_r) \le V(\mathrm{P}_r)$. Under the following Slater-type constraint qualification the two programs will be seen to be equivalent:

$$\text{(CQ)} \qquad \begin{array}{l} \text{there exists } \hat{u} \in \mathcal{C}^1(\bar{\Omega}),\ \hat{u} > 0, \text{ such} \\ \text{that } \displaystyle\int_\Omega a_0 \hat{u}\, dx = b_0 \text{ and } |A\hat{u} - b| < \epsilon. \end{array}$$

Proposition 3.1. *Suppose that* (CQ) *holds for* (P_r), *then there exist multipliers* $\bar{w} \in \mathcal{L}^2(\Omega)$, $\bar{\lambda} \in \mathbb{R}^m$, *and* $\bar{\mu} \in \mathbb{R}$, $\bar{\nu} \ge 0$ *such that*

$$V(\mathrm{D}_r) = V(\mathrm{P}_r) = \inf_{u,p,e} L(u,p,e;\bar{w},\bar{\lambda},\bar{\mu},\bar{\nu}) = L(\bar{u},\nabla\bar{u},\bar{e};\bar{w},\bar{\lambda},\bar{\mu},\bar{\nu})$$

are satisfied. That is, $(\bar{u},\nabla\bar{u},\bar{e};\bar{w},\bar{\lambda},\bar{\mu},\bar{\nu})$ *is a saddle point of* L *in* $W^{1,2}(\Omega)\times\mathcal{L}^2(\Omega)\times\mathbb{R}^m\times\mathcal{L}^2(\Omega)\times\mathbb{R}^m\times\mathbb{R}\times\mathbb{R}$.

Proof. The result is standard, so we only sketch the argument. Define a lower semicontinuous proper convex function $f : \mathcal{L}^2(\Omega)\times\mathbb{R}^m\times\mathbb{R}\times\mathbb{R} \to \mathbb{R}\cup\{+\infty\}$ by setting

$$\begin{aligned} f(z,\theta,\alpha,\sigma) = \inf\Bigg\{ &\mathcal{I}(u,p) : u \in W^{1,2},\ p \in \mathcal{L}^2,\ \nabla u - p = z, Au - b - e = \theta, \\ &\int_\Omega a_0 u dx - b_0 = \alpha, \epsilon^2 - |e|^2 \ge \sigma \Bigg\}. \end{aligned}$$

Then the value of (P_r) is $V(\mathrm{P}_r) = f(0,0,0,0) > -\infty$. It remains to show that $(0,0,0,0)$ is in the interior of $\mathrm{dom} f$ as a subset of $\mathcal{L}^2\times\mathbb{R}^m\times\mathbb{R}\times\mathbb{R}$, for then $\partial f(0,0,0,0)$ is nonempty and any $(-\bar{w},-\bar{\lambda},-\bar{\mu},-\bar{\nu}) \in \partial f(0,0,0,0)$ will provide the desired Lagrange multipliers. Now the constraint qualification (CQ) is seen to imply $f(z,\theta,\mu,\nu) < +\infty$ on a full neighborhood of the origin in $\mathcal{L}^2(\Omega)\times\mathbb{R}^m\times\mathbb{R}\times\mathbb{R}$, whence the result. ☐

Corollary 3.2. *Suppose that* (CQ) *is satisfied for* (P_r) *and let* $\bar{u}$ *be its solution. Then* $E = \{x \in \Omega : \bar{u}(x) = 0\}$ *is a Lebesgue null set and the function* $\log\bar{u}$ *is in* $W^{1,2}(\Omega)$. *Moreover,* $\bar{u} > 0$ *on* Ω *if either* (i) *the dimension of the problem is* $n = 1$ *or* (ii) *the weight functions* a_k *are analytic on* Ω.

Proof. a) We first prove that $|\nabla\bar{u}|/\bar{u}$ is square integrable on $\Omega\setminus E = \{x \in \Omega : \bar{u}(x) > 0\}$. Since $(\bar{u},\nabla\bar{u},\bar{e};\bar{w},\bar{\lambda},\bar{\mu},\bar{\nu})$ is a saddle point of L, we find that

$$\begin{aligned} 0 &\le \frac{1}{t}\left(L(\bar{u}+t,\nabla\bar{u},\bar{e};\bar{w},\bar{\lambda},\bar{\mu},\bar{\nu}) - L(\bar{u},\nabla\bar{u},\bar{e};\bar{w},\bar{\lambda},\bar{\mu},\bar{\nu})\right) \\ &= \int_{\{\bar{u}>0\}} \frac{-|\nabla\bar{u}|^2}{\bar{u}(\bar{u}+t)} dx + \int_\Omega \left(A^*\bar{\lambda} + \bar{\mu}a_0\right) dx < +\infty, \end{aligned}$$

which uses the fact that $\nabla\bar{u} = 0$ a.e. on $\{\bar{u} = 0\}$. By convexity, the integrand is nondecreasing in $t > 0$, and hence monotone convergence allows us to pass to the limit $t \to 0$ under the integral sign. This implies $|\nabla\bar{u}|/\bar{u} \in \mathcal{L}^2(\Omega)$ on $\{\bar{u} > 0\}$.

b) We now argue that E is Lebesgue null. We first show how to derive this in dimension $n = 2$. Since $\bar{u} \in W^{1,2}(\Omega)$, [39, Thm. 2.1.4] implies that for almost all

$\xi \in \mathbb{R}$, $\bar{u}(\xi, \cdot)$ is absolutely continuous on $\Omega_\xi = \{\eta \in \mathbb{R} : (\xi, \eta) \in \Omega\}$. Moreover, by part a), $\partial_2 \bar{u}/\bar{u}$ is integrable on $\Omega \setminus E$, and hence for almost all ξ, $\partial_2 \bar{u}(\xi, \cdot)/\bar{u}(\xi, \cdot)$ is integrable on $\Omega_\xi \cap \{\eta : \bar{u}(\xi, \eta) > 0\}$. Fix some ξ satisfying both conditions. We claim that the following alternative holds: Either $\bar{u}(\xi, \cdot) > 0$ on Ω_ξ, or $\bar{u}(\xi, \cdot) = 0$ identically there. Indeed, suppose we had $\bar{u}(\xi, \eta) = 0$ for some $\eta \in \Omega_\xi$, but $\bar{u}(\xi, \cdot) > 0$ near the zero η, say on $I = (\eta, \eta')$. Then $\partial_2 \bar{u}(\xi, \cdot)/\bar{u}(\xi, \cdot)$ fails to be integrable on I, which is a contradiction. Indeed, for $\delta > 0$ we have

$$\int_{\eta+\delta}^{\eta'} \frac{\partial_2 \bar{u}(\xi, \eta)}{\bar{u}(\xi, \eta)} d\eta = \log \bar{u}(\xi, \eta') - \log \bar{u}(\xi, \eta + \delta) \to +\infty \qquad (\delta \to 0).$$

This proves the claimed alternative.

Suppose now that $\bar{u}(\xi, \cdot) \equiv 0$ on some Ω_ξ. This shows that every $\bar{u}(\cdot, \eta)$ vanishes somewhere on $\Omega^\eta = \{\xi : (\xi, \eta) \in \Omega\}$. Repeating the above argument with respect to the second coordinate gives the same alternative as above, but with the first possibility now excluded. Hence for almost every η, we have $\bar{u}(\cdot, \eta) = 0$ identically, i.e., $\bar{u} \equiv 0$ a.e. This is impossible, and hence E must be a null set in dimension $n = 2$. For larger n we have to iterate the above argument using Fubini's theorem.

With $\bar{u} > 0$ a.e., it follows that $\log \bar{u}$ is defined a.e., and such that $\nabla(\log \bar{u}) \in \mathcal{L}^2(\Omega)$ by part a). It follows via a Poincaré inequality that $\log \bar{u}$ is in $W^{1,2}(\Omega)$.

c) Statement (i) now follows, for in dimension $n = 1$ the embedding theorem gives $W^{1,2}(\Omega) \subset \mathcal{C}(\bar{\Omega})$, and so $\log \bar{u}$ is continuous, which means $\bar{u} > 0$. As for statement (ii), observe that the differential operator in (2.4) has now analytic coefficients, so by Lewy's classical argument (cf. [12]) the solution $\bar{v}$, and hence $\bar{u} = \bar{v}^2$, must be analytic on Ω. For analytic $\bar{u}$, $\log \bar{u}$ could not be integrable in a neighborhood of a zero of $\bar{u}$, so $\bar{u} > 0$. □

Remark. The above reasoning shows that the set E could not contain any line segments or even pieces of curves.

In the following, assume that (CQ) is satisfied with $\hat{u} \geq \eta > 0$ of class $\mathcal{C}^1(\bar{\Omega})$, so that Proposition 3.1 applies. This implies

$$L(\hat{u} + u, \nabla \hat{u}, \hat{e}; \bar{w}, \bar{\lambda}, \bar{\mu}, \bar{\nu}) \geq V(P_r) = V(D_r) > -\infty$$

for every $u \in \mathcal{C}^1(\bar{\Omega})$ having $\|u\|_\infty \leq \eta/2$. Since $\mathcal{I}(\hat{u} + u, \nabla \hat{u})$ is bounded on $\|u\|_\infty \leq \eta/2$, the linear form $u \to \langle \nabla u, \bar{w} \rangle$ is bounded below (and hence bounded) on $\|u\|_\infty \leq \eta/2$. It extends to a bounded linear form on $\mathcal{C}(\bar{\Omega})$. Hence there exists a signed Borel measure $\bar{m}$ on $\bar{\Omega}$ satisfying

$$-\langle u, \bar{m} \rangle = \langle \nabla u, \bar{w} \rangle$$

for every $u \in \mathcal{C}^1(\bar{\Omega})$. Notice that $\bar{m}$ =div $\bar{w}$ in the distributional sense, but $\bar{m}$ also comprises the singular boundary measure. Let $\bar{m}$ be decomposed according to Lebesgue, that is,

$$\bar{m} = \bar{m}_a + \bar{m}_s,$$

where $\bar{m}_a$ is absolutely continuous with respect to Lebesgue measure and $\bar{m}_s$ is singular. Let $d\bar{m}_a = \bar{\phi}\, dt$, then the Lagrangian (3.1) may be rewritten in the form

$$\begin{aligned} L(u, v, e; \bar{w}, \bar{\lambda}, \bar{\mu}, \bar{\nu}) = {} & \mathcal{I}(u, v) - \langle \bar{\phi}, u \rangle - \langle \bar{m}_s, u \rangle - \langle \bar{w}, v \rangle + \langle A^t \bar{\lambda}, u \rangle \\ & - \bar{\lambda} \cdot e - \bar{\lambda} \cdot b + \langle \bar{\mu} a_0, u \rangle - \bar{\mu} b_0 + \bar{\nu}(\epsilon^2 - |e|^2). \end{aligned} \tag{3.2}$$

Notice that in contrast with (3.1), the presence of the term $\langle \bar{m}_s, u\rangle$ in formula (3.2) requires at least $u \in \mathcal{C}(\Omega)$. Assuming in the following that the program data allow for Proposition 2.2 to apply (with $t \geq n/2$), the optimal $\bar{u}$ is of class $\mathcal{C}^1(\bar{\Omega})$. In view of Proposition 3.1, this means that $(\bar{u}, \nabla\bar{u}, \bar{e}; \bar{w}, \bar{\lambda}, \bar{\mu}, \bar{\nu})$ is now a saddle point of (3.2) in the duality $\mathcal{C}(\Omega) \times \mathcal{L}^2(\Omega) \times \mathbb{R}^m \times \mathcal{L}^2(\Omega) \times \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}$.

Following the outline of convex duality theory, we shall now calculate (using (3.2)) the Fenchel conjugate $L^*(\cdot,\cdot,\cdot;\bar{w},\bar{\lambda},\bar{\mu},\bar{\nu})$ of $L(\cdot,\cdot,\cdot;\bar{w},\bar{\lambda},\bar{\mu},\bar{\nu})$ with respect to this duality. As $\bar{m}_s$ is concentrated on a Lebesgue null set, the supremum over the term $\langle \bar{m}_s, u\rangle$ may be controlled by a portion of each u living on a set with arbitrarily small Lebesgue measure, not affecting the supremum to be obtained for the remaining terms, (cf. [5] for a similar argument). The methods of Rockafellar [35] therefore lead to the following formula

$$L^*(y,z,\alpha;\bar{w},\bar{\lambda},\bar{\mu},\bar{\nu}) = \mathcal{I}^*(y+\bar{\phi}-\bar{\mu}a_0 - A^t\bar{\lambda}, z+\bar{w}) + \bar{\mu}b_0 + \bar{\lambda}\cdot b + \epsilon|\bar{\lambda}+\alpha|, \\ + \sup_{u\geq 0}\langle \bar{m}_s, u\rangle.$$

Here $A^t : \mathbb{R}^m \to \mathcal{L}^2(\Omega)$ denotes the adjoint of A, defined as $A^t\lambda = \sum_1^m \lambda_k a_k$. Now the conjugate $\mathcal{I}^*$ of the Fisher integral functional (1.5) is found to be $\mathcal{I}^*(y,z) = \int_\Omega h^*(y(x), z(x))\, dx$, where

$$h^*(y,z) = \begin{cases} 0 & \text{if } y + \frac{1}{4}|z|^2 \leq 0, \\ +\infty & \text{otherwise} \end{cases}$$

and dealing with the singular term in L^* remains.

Observe that L^* has finite value somewhere; hence $\sup_{u\geq 0}\langle \bar{m}_s, u\rangle = \sup_{u\geq 0}\big(\langle \bar{m}_s^+, u\rangle - \langle \bar{m}_s^-, u\rangle\big)$ must be finite. This implies that the positive part $\bar{m}_s^+$ must vanish. The remaining term is $\sup_{u\geq 0}\langle -\bar{m}_s^-, u\rangle = 0$. Since $L^*(0,0,0;\cdot,\cdot,\cdot,\cdot)$ is finite at $(\bar{w},\bar{\lambda},\bar{\mu},\bar{\nu})$, we must have $\bar{\phi} - \bar{\mu}a_0 - A^t\bar{\lambda} + \frac{1}{4}|\bar{w}|^2 \leq 0$ a.e. Moreover, the saddle point condition (Proposition 3.1) implies

$$\langle \bar{u}, y\rangle + \langle \nabla\bar{u}, z\rangle + \bar{e}\cdot\alpha \leq \epsilon\big(|\bar{\lambda}+\alpha| - |\bar{\lambda}|\big) \tag{3.3}$$

for all (y,z,α) for which L^* is finite, or what is the same, for which

$$y + \bar{\phi} - \bar{\mu}a_0 - A^t\bar{\lambda} + \frac{1}{4}|z+\bar{w}|^2 \leq 0 \quad \text{a.e.} \tag{3.4}$$

Suppose now we have $\bar{\phi} - \bar{\mu}a_0 - A^t\bar{\lambda} + \frac{1}{4}|\bar{w}|^2 < -\theta < 0$, (for some $\theta > 0$) on a set Σ. Choosing $y = \theta\chi_\Sigma$ gives a tuple $(y,0,0)$ admitted in (3.4); hence $\langle \bar{u}, \theta\chi_\Sigma\rangle \leq 0$. As $\bar{u}$ is strictly positive, this is possible only when Σ is a null set. This reasoning shows that

$$\bar{\phi} - \bar{\mu}a_0 - A^t\bar{\lambda} + \frac{1}{4}|\bar{w}|^2 = 0 \qquad \text{a.e.} \tag{3.5}$$

The feasibility condition (3.4) therefore becomes $y + \frac{1}{2}z\cdot\bar{w} + \frac{1}{4}|z|^2 \leq 0$ a.e. Consequently, on setting $y = -\frac{1}{2}z\cdot\bar{w} - \frac{1}{4}|z|^2$, we obtain a feasible $(y,z,0)$. Inserting this in (3.3) gives

$$\left\langle \bar{u}, -\frac{1}{2}z\cdot\bar{w}\right\rangle - \frac{1}{4}\langle \bar{u}, |z|^2\rangle + \langle \nabla\bar{u}, z\rangle \leq 0$$

for every $z \in \mathcal{L}^\infty(\Omega)$. As for small z the quadratic term becomes negligible, this readily implies

$$\left\langle -\frac{1}{2}\bar{u}\bar{w} + \nabla\bar{u}, z \right\rangle \leq 0$$

for every such z, and hence, on replacing z by $-z$, we obtain

$$\frac{1}{2}\bar{w} = \frac{\nabla\bar{u}}{\bar{u}} = \nabla(\log\bar{u}) \text{ a.e.} \tag{3.6}$$

In particular, $\bar{w}$ is seen to be a gradient field, and equation (3.5) with $\bar{\phi} = (\mathrm{div}\bar{w})_a = \big(\mathrm{div}\, \nabla(2\log\bar{u})\big)_a$ now reads as

$$2\big(\Delta(\log\bar{u})\big)_a + |\nabla(\log\bar{u})|^2 - \bar{\mu}a_0 - A^t\bar{\lambda} = 0. \tag{3.7}$$

The question which remains is when we may expect $\bar{m}_s^-$ to vanish. This is closely related to the question whether or not the implicit constraint $u \geq 0$ in our original problem (P) becomes active.

LEMMA 3.3. *The singular measure $\bar{m}_s^-$ is concentrated on the set $E = \{x \in \bar{\Omega} : \bar{u}(x) = 0\}$.*

Proof. We sketch the argument, which is of classical variational type. Observe that the infimum $V(P_r)$ of $L(\cdot,\cdot,\cdot;\bar{w},\bar{\lambda},\bar{\mu},\bar{\nu})$ is attained at $(\bar{u},\nabla\bar{u},\bar{e})$. Let $Q = \mathrm{support}(\bar{m}_s^-) \cap \{x : \bar{u}(x) > 0\}$. For a given $\epsilon > 0$ we find disjoint open intervals $I_1,\dots,I_r$ covering the support of $\bar{m}_s^-$ up to a set of $\bar{m}_s^-$-measure $< \epsilon$, and such that the whole family has Lebesgue measure $< \epsilon$. Now start modifying the function $\bar{u}$ on the small set

$$Q_\epsilon = \bigcup_{i=1}^r I_i \cap \{x : \bar{u}(x) > 0\}$$

to build a function u_ϵ, which on a relevant portion P_ϵ of Q_ϵ (i.e., $\bar{m}_s^-(Q_\epsilon \setminus P_\epsilon) \to 0$) equals $\frac{1}{2}\bar{u}$, while outside Q_ϵ still equals $\bar{u}$. The construction gives

$$\langle \bar{m}_s^-, u_\epsilon\rangle \to \frac{1}{2}\langle \bar{m}_s^-, \bar{u}\rangle.$$

As $u_\epsilon \to \bar{u}$ in the sense of Lebesgue measure, the nonsingular terms in (3.2) are not affected, that is, on passing to the limit $\epsilon \to 0$, we find

$$V(P_r) \leq L(u_\epsilon, \nabla\bar{u}, \bar{e}; \bar{w}, \bar{\lambda}, \bar{\mu}, \bar{\nu}) \to V(P_r) - \frac{1}{2}\langle \bar{m}_s^-, \bar{u}\rangle,$$

giving $\langle \bar{m}_s^-, \bar{u}\rangle = 0$, and hence $\bar{m}_s^-(Q) = 0$, as claimed. ∎

Remark. In view of Corollary 2.3, $\bar{m}_s = 0$ if either the dimension is $n = 1$ or the weights a_i are analytic functions on Ω.

Under our standing assumption on the program data, $\bar{u}$ is continuous. Then the exceptional set $E := \{x : \bar{u}(x) = 0\}$ is closed. Moreover, E is a Lebesgue null set by Corollary 3.2, and so $\Omega \setminus E$ is an open dense set. Now we compare equations (2.4) and (3.5). On $\Omega \setminus E$, the substitution $\frac{1}{2}\bar{w} = \nabla\bar{u}/\bar{u} = 2\nabla\bar{v}/\bar{v}$ may just be understood in the classical sense. The a_k being linearly independent, this implies that (up to a factor $\frac{1}{4}$) the multipliers $\bar{\lambda}_k$ coming from Proposition 2.2, and the multipliers $\bar{\lambda}_k$ coming from the convex duality of this section are identical (a fact which was to be expected, but still had to be verified). Since by Lemma 3.3, $\big(\mathrm{div}\bar{w}\big)_a = \mathrm{div}\,\bar{w}$ on $\Omega \setminus E$, equations (3.5) and (2.4) are equivalent on this set. We shall see how this piece of information may be used to obtain a convenient formulation of the dual program (∇_r).

Recall that the dual (D_r) consists in maximizing $-L^*(0,0,0;w,\lambda,\mu,\nu) = -\mu b_0 - \sum_1^m \lambda_k b_k - \epsilon|\lambda|$ over those (w,λ,μ,ν) satisfying (3.5). On replacing (3.5) by (2.4), and taking into account that the boundary condition coming along with (2.4) is the Neumann condition (2.5), we obtain the following characterization of the dual (D_r) of (P_r).

THEOREM 3.4. *Assuming* (CQ) *and the conditions on the program data from Proposition* 2.2 *(with $t \geq n/2$), the dual* (D_r) *of* (P_r) *takes the equivalent form*

$$
(\mathrm{D}_r)\qquad
\begin{array}{lll}
\textit{maximize} & -\mu b_0 - \displaystyle\sum_{k=1}^m \lambda_k b_k - \epsilon|\lambda| & \\
\textit{subject to} & \Delta v = \dfrac{1}{4}\displaystyle\sum_{k=1}^m \lambda_k a_k(x) v + \dfrac{1}{4}\mu a_0(x) v & \textit{on } \Omega, \\
& \dfrac{\partial v}{\partial n_\Omega} = 0 & \textit{on } \partial\Omega, \\
& v \geq 0, \ \displaystyle\int_\Omega a_0 v^2 dx = b_0. &
\end{array}
$$

Moreover, the optimal solution $\bar{u}$ of (P_r) *is obtained from the optimal solution* $(\bar{\lambda},\bar{\mu},\bar{v})$ *of* (D_r) *via* $\bar{u} = \bar{v}^2$.

So far the arguments in this section have been for program (P_r) and its dual. Using a similar reasoning, we obtain the following result on program (P).

THEOREM 3.5. *Suppose that the above assumptions on the program data are satisfied, but with the constraint qualification* (CQ) *for* (P_r) *replaced by the following:*

$$\textit{There exists } \hat{u} \in \mathcal{C}^1(\bar{\Omega}),\ \hat{u} > 0,\ \textit{such that } A\hat{u} = b.$$

Then the dual (D) *of* (P) *has the equivalent form*

$$
(\mathrm{D})\qquad
\begin{array}{lll}
\textit{maximize} & -\mu b_0 - \displaystyle\sum_{k=1}^m \lambda_k b_k & \\
\textit{subject to} & \Delta v = \dfrac{1}{4}\displaystyle\sum_{k=1}^m \lambda_k a_k v + \dfrac{1}{4}\mu a_0(x) v & \textit{on } \Omega, \\
& \dfrac{\partial v}{\partial n_\Omega} = 0 & \textit{on } \partial\Omega, \\
& v \geq 0, \ \displaystyle\int_\Omega a_0 v^2 dx = b_0. &
\end{array}
$$

Moreover, the optimal solution $\bar{u}$ of (P) *may be recovered from the optimal solution* $(\bar{\lambda},\bar{\mu},\bar{v})$ *of* (D) *via* $\bar{u} = \bar{v}^2$.

4. Numerical aspects. In this section we obtain a formulation of the programs (D) (resp., (D_r)) which is appropriate for a numerical treatment. We adopt the assumptions about the program data made in Theorems 3.4 (resp., 3.5), assuming in addition that $a_0 \geq 0$. We start by analyzing program (D), which we reformulate slightly: replace λ_k by $-4\lambda_k$, $k = 1,\ldots,m$, and μ by -4μ, and multiply the objective with 1/4. Then the dual (D) takes the equivalent form

$$\text{(D)}\qquad \begin{aligned} &\text{maximize} && \mu b_0 + \sum_{k=1}^m \lambda_k b_k \\ &\text{subject to} && \Delta v + \sum_{k=1}^m \lambda_k a_k(x) v + \mu a_0(x) v = 0 \quad \text{on } \Omega, \\ & && \frac{\partial v}{\partial n_\Omega} = 0 \qquad\qquad\qquad\qquad\qquad \text{on } \partial\Omega, \\ & && v \geq 0, \int_\Omega a_0 v^2 dx = b_0. \end{aligned}$$

This shows that for any feasible $(\mu, \lambda_1, \dots, \lambda_m, v)$, μ is an eigenvalue for the Neumann eigenvalue problem

$$\text{(N)}\qquad \begin{aligned} & L_\lambda[v] + \mu a_0(x) v = 0 \quad \text{on } \Omega, \\ & \frac{\partial v}{\partial n_\Omega} = 0 \qquad\qquad\qquad \text{on } \partial\Omega \end{aligned}$$

with differential operator $L_\lambda[u] = \Delta u + (\sum_1^m \lambda_k a_k(x)) \cdot u$, and with v the corresponding eigenfunction, normalized to satisfy $\int_\Omega a_0 v^2 dx = b_0$. Now observe that an eigenfunction which does not change sign must belong to the smallest eigenvalue of (N), so any such μ must be the smallest eigenvalue for (N). Indeed, this is a classical fact for $a_0 \geq 0$, cf. [12, sect. 6], but may be extended to more general weights a_0 (see [9]).

Let us introduce some notation. For a self-adjoint second-order elliptic partial differential operator $L = L[\cdot]$, the smallest eigenvalue for the Neumann problem (N), which is known to be simple (cf. [21, 12]), will be denoted as $\mu_{\min}(L)$. The function $\mu_{\min}$ is known to be a concave function of the argument, see [28]. Since $\lambda \to L_\lambda[\cdot]$ is affine, we derive that $\lambda \to \mu_{\min}(L_\lambda[\cdot])$ is a concave function defined on the whole of $\mathbb{R}^m$. This function is even analytic, since the smallest eigenvalue of L_λ is simple (cf. [25, VII, Thm. 1.8] and also [12, 28]). The dual (D) may therefore be recast as the following finite-dimensional unconstrained and smooth concave program on $\mathbb{R}^m$, which is recognized as a semidefinite eigenvalue optimization program (in the sense of [28, 38]):

$$\text{(D)}\qquad \text{maximize } f(\lambda) := \mu_{\min}(L_\lambda[\cdot]) b_0 + \sum_{k=1}^m \lambda_k b_k, \qquad \lambda \in \mathbb{R}^m.$$

Given the optimal solution $\bar\lambda$ of (D), due to (3.6), the optimal solution $\bar u$ of the original program (P) is $\bar u = \bar v^2$, where $\bar v$ is the unique eigensolution belonging to the eigenvalue $\mu_{\min}(L_{\bar\lambda})$, which has been normalized to satisfy $\int a_0 \bar v^2 dx = b_0$.

The dual (D_r) of (P_r) may be treated in the same way, leading to the concave unconstrained program

$$(\mathrm{D}_r)\qquad \text{maximize } g(\lambda) := \mu_{\min}(L_\lambda[\cdot]) b_0 + \sum_{k=1}^m \lambda_k b_k - \epsilon|\lambda|, \quad \lambda \in \mathbb{R}^m.$$

The return formula for the optimal solution $\bar u$ of (P_r) in terms of the optimal $\bar\lambda$ for (D_r) is exactly the same as for (P), (D) above.

The fact that the variational problem (P) is equivalent to a smooth concave unconstrained finite-dimensional optimization problem is by itself remarkable. It must nevertheless be observed that the function $\lambda \to \mu_{\min}(L_\lambda)$ cannot be calculated explicitly, so for the numerical approach we are forced to replace the operator L_λ by a discretization L_λ^h. The error thereby introduced may be analyzed through the error term $|\mu_{\min}(L_\lambda) - \mu_{\min}(L_\lambda^h)|$.

Motivated by the image restoration problem (1.4), a natural discretization for (D_r) appears to be a finite difference approximation L_λ^h of L_λ on a rectangular grid Ω_h with mesh $h > 0$ on Ω. On the other hand, as our program (D) comprises a variety of eigenvalue optimization problems for second-order elliptic operators with a physical background, one might as well use a finite element or Rayleigh–Ritz discretization. (This was, for instance, chosen for the numerical treatment of the strongest column problem by Cox and Overton [13]). In the following we shall discuss the discretization errors associated with either method, and the sensitivity of the solution of (D) (resp., (P)) as compared to its discretized version.

LEMMA 4.1. *Let $L_\lambda[u] = \Delta u + \sum_{k=1}^m \lambda_k a_k(x)u$, and let $\phi(\lambda) = \mu_{\min}(L_\lambda[\cdot])$ be the minimal eigenvalue for the Neumann problem* (N) *above. Then ϕ is analytic and the first-order partial derivatives are*

$$\frac{\partial \phi(\lambda)}{\partial \lambda_i} = \int_\Omega a_i(x) e_\lambda(x)^2 dx, \tag{4.1}$$

where e_λ is the normalized eigenfunction belonging to the eigenvalue $\mu_{\min}(L_\lambda)$.

Proof. Since the smallest eigenvalue of L_λ is simple, the analyticity of ϕ is a direct consequence of Theorem VII. 1.8 in [25]. To prove (4.1) we shall need an approximation argument.

Let L_λ^h be a finite difference approximation for L_λ with mesh $h > 0$. For h small enough, the first eigenvalue $\phi_h(\lambda) := \mu_{\min}(L_\lambda^h)$ of the matrix L_λ^h is simple. Now observe that the eigenfunctions belonging to L_λ^h are defined on the Euclidean space $\mathcal{L}_h^2$ with weighted norm $|e|_0 = h^{n/2}\big(\sum_{x\in\Omega_h} e(x)\big)^{1/2}$, so we have

$$\frac{\partial \phi_h(\lambda)}{\partial \lambda_i} = (h^{n/2} e_\lambda^h) \cdot \frac{\partial L_\lambda^h}{\partial \lambda_i}(h^{n/2} e_\lambda^h) \tag{4.2}$$

according to a well-known formula in finite dimensions (cf. [28]). Here the eigenvectors e_λ^h must satisfy $|e_\lambda^h|_0 = \|h^{n/2} e_\lambda^h\|_2 = 1$.

Now observe that $\partial L_\lambda^h / \partial \lambda_i$ is the diagonal matrix with entries $a_i(x)$, $x \in \Omega_h$, the discretization of $a_i(x)$. Therefore the right-hand side of (4.2) equals

$$\sum_{x\in\Omega_h} h^n a_i(x) e_\lambda^h(x)^2,$$

which converges to the right-hand side of (4.1) as $h \to 0$. (For details concerning the finite difference discretization see, e.g., [21, sect. 11.2]). $\square$

Let us now consider the case of a finite-element or Rayleigh–Ritz approximation of problems (D), (D_r). Let $V = W^{1,2}(\Omega)$, $U = \mathcal{L}^2(\Omega)$ be the Gelfand pair associated with the Neumann problem (N), and let V_h be an approximating sequence of finite-dimensional subspaces of V. We assume that (cf. [21, sect. 11.2])

$$\mathrm{dist}(u, V_h) \le C\|u\|_{1+s} h^s \qquad \text{for every } u \in W^{1+s,2}(\Omega), \tag{4.3}$$

which means that the basis functions are polynomials of degree $s \ge 1$ on the elements; cf. [11, Thm. 3.2.1]. Further, as in previous sections, we assume linear independence and smoothness of the weight functions $a_0, a_1, \ldots, a_m$, and for convenience we let $a_0 \equiv 1$. We now work on a bounded domain Ω with $\mathcal{C}^{0,1}$ boundary.

Let L_λ^h be the operator belonging to the approximation (N_h) of the variational problem (N) on the subspace V_h. Let $f(\lambda) = \mu_{\min}(L_\lambda)b_0 + \sum_1^m \lambda_k b_k$ be the objective of (D), and let the objective belonging to the approximate program (D_h) be $f_h(\lambda) = \mu_{\min}(L_\lambda^h)b_0 + \sum_1^m \lambda_k b_k$ to be maximized over $\mathbb{R}^m$.

Let the optimal solutions of (D), (D_h) be $\bar\lambda, \lambda^h$, respectively. Let $\bar v$ (resp., v^h) be the positive eigensolutions belonging to $\mu_{\min}(L_{\bar\lambda})$ (resp., $\mu_{\min}(L_{\lambda^h}^h)$), both normalized to satisfy $\int_\Omega \bar v^2 dx = b_0 = \int_\Omega (v^h)^2 dx$. Then we consider $u^h = (v^h)^2$ as an approximation of $\bar u$, and our interest will be in the error $\|\bar u - u^h\|_{W^{1,1}}$.

Theorem 4.2. *Under the above assumptions on the program data, consider a finite element approximation V_h of V satisfying* (4.3). *Then we have the following statements.*

1. *For every compact set $\Lambda \subset \mathbb{R}^m$, there exists a constant C (depending only on s and Λ) such that the estimates*

$$|f(\lambda) - f_h(\lambda)| \le Ch^{2s} \quad \textit{and} \quad |\nabla f(\lambda) - \nabla f_h(\lambda)| \le Ch^s$$

   are satisfied for all $h > 0$ and $\lambda \in \Lambda$.
2. *The optimal solutions converge: $\lambda_h \to \bar\lambda$, $v^h \to \bar v$ (in $W^{1,2}$), and $u^h \to \bar u$ (in $W^{1,1}$), and the rate of convergence is $\|\bar u - u^h\|_{W^{1,1}} = \mathcal{O}(h^s)$.*

Proof. a) It is a known classical fact that the smallest eigenvalue of L_λ^h converges to the smallest eigenvalue of L_λ as $h \to 0$; cf. [21, 11.2.10]. This implies $f_h \to f$ pointwise and hence uniformly on any compact set Λ, the latter since f, f_h are concave.

Let us now fix a convex compact Λ with the global maximum $\bar\lambda$ of f in its interior. (Notice here that the maximum $\bar\lambda$ of f is unique since for any other maximum $\tilde\lambda$, the corresponding $\tilde\mu = \mu_{\min}(L_{\tilde\lambda})$ and $\tilde v$ would satisfy the Euler–Lagrange equation (2.4) with boundary condition (2.5), and by Theorem 3.5 would produce an optimal solution for $(\tilde P)$. This gave $\bar v = \tilde v$, and hence also $(\bar\lambda, \bar\mu) = (\tilde\lambda, \tilde\mu)$ by the pseudo-Haar assumption). As $f_h \to f$ uniformly on Λ, the global maxima λ^h of the f_h, which are unique for h small enough, must belong to Λ for h small enough. Again, this strongly relies on the concavity of f, f_h.

b) Since L_λ is a second-order operator, the smallest eigenvalue $\mu_{\min}(L_\lambda)$ is simple (cf. [12, sect. 6]) and hence so is the smallest eigenvalue $\mu_{\min}(L_\lambda^h)$ for every $h < h_0$ (h_0 depending only on Λ) and all $\lambda \in \Lambda$.

By Proposition 2.2, the eigensolutions are in $W^{1,1+s}(\Omega)$. Using this, another known result tells us that (with a constant C_L independent of h) the smallest eigenvalue $\mu_{\min}(L)$ belonging to a Neumann boundary problem for a V-coercive second-order partial differential operator L satisfies estimates of the form

$$|\mu_{\min}(L) - \mu_{\min}(L^h)| \le C_L h^{2s} \quad \text{and} \quad \|e(L) - e(L^h)\|_V \le C_L h^s \tag{4.4}$$

if $e(L), e(L^h)$ are the corresponding normalized eigensolutions; cf. [21, sect. 11]. The difficulty is now that we are dealing with a family L_λ, $\lambda \in \Lambda$, of such operators and will need a uniform estimate of the type (4.4).

Following carefully the arguments in sections 11.2.2 and 11.2.3 of [21], where the estimates (4.4) are obtained for single eigenvalues of general V-coercive elliptic operators, it can in fact be shown that the constant C_{L_λ} for the L_λ is of the form $\mathcal{O}(|\lambda|)$. Therefore, on a compact set Λ, the uniform estimates

$$|\mu_{\min}(L_\lambda) - \mu_{\min}(L_\lambda^h)| \le Ch^{2s} \quad \text{and} \quad \|e(L_\lambda) - e(L_\lambda^h)\|_V \le Ch^s \tag{4.5}$$

are satisfied. We do not present the individual steps of the derivation of (4.5) here, since this may be done by tracing the constants in the above reference, while otherwise does not provide any new insight. (Notice, however, that the arguments in [21] are for single eigenvalues, and a similar result for say the finite element approximation of the strongest column problem [13] would require a more sophisticated reasoning.)

Notice that the first formula in (4.5) gives the first part of statement 1. The second part of item 1 now follows readily from the representation (4.1) of the partial derivatives of $\lambda \to \mu_{\min}(L_\lambda)$, which gives $\partial f(\lambda)/\partial\lambda_i = \int_\Omega a_i e(L_\lambda)^2 dx + b_i$ and $\partial f_h(\lambda)/\partial\lambda_i = \int_\Omega a_i e(L_\lambda^h)^2 dx + b_i$. This proves item 1.

c) The second formula of item 1 implies $|\nabla f(\lambda^h)| = |\nabla f(\lambda^h) - \nabla f_h(\lambda^h)| \le Ch^s$ for h small enough, and in particular, $\nabla f(\lambda^h) \to 0$. Now by the continuity of ∇f, every convergent subsequence λ^{h_i} of $\lambda^h \in \Lambda$ must have a limit λ satisfying $\nabla f(\lambda) = 0$. Since the global maximum $\bar\lambda$ of f is unique, we deduce that $\lambda^h \to \bar\lambda$.

Now consider any sequence $e^h := e(L^h_{\lambda^h})$ of normalized eigenfunctions corresponding to $\mu_{\min}(L^h_{\lambda^h})$. Since the embedding $W^{1,2} \to \mathcal{L}^2$ is compact, we find a subsequence $e^{h_i} \to e$ for some $e \in \mathcal{L}^2$. Reasoning exactly as in [21, 11.2.11], we can argue that in fact $e \in W^{1,2}$ and convergence is in the Sobolev norm. In particular, e must be a normalized eigenfunction of $L_{\bar\lambda}$, and since there is only one such function, $e(L_{\bar\lambda})$, we deduce that the entire sequence converges to this normalized eigenfunction in $W^{1,2}$-norm. (Again, this part of the reasoning makes strong use of the fact that the smallest eigenvalue is simple.) We deduce that $v^h = b_0^{1/2} e(L^h_{\lambda^h}) \to \bar v = b_0^{1/2} e(L_{\bar\lambda})$, which gives the second part of statement 2. The third part of item 2 is clear from $u^h = (v^h)^2$ and $\bar u = \bar v^2$.

d) We finally obtain the rate of convergence in statement 2. Starting out with

$$\|e(L^h_{\lambda^h}) - e(L_{\bar\lambda})\|_V \le \|e(L^h_{\lambda^h}) - e(L_{\lambda^h})\|_V + \|e(L_{\lambda^h}) - e(L_{\bar\lambda})\|_V,$$

we observe that the first term on the right-hand side is $\mathcal{O}(h^s)$ by (4.5). For the second term, observe that $|\lambda^h - \bar\lambda| = \mathcal{O}(h^{2s})$, so the claim follows when we observe that $\lambda \to e_\lambda = e(L_\lambda)$ is Lipschitz in a neighborhood of $\bar\lambda$. The latter follows again from the fact that the eigenvalues $\mu_{\min}(L_\lambda)$ are simple, whence the corresponding eigenfunctions $e(L_\lambda)$ depend smoothly on λ; cf. [25]. This shows $\|v^h - \bar v\|_V = b_0^{1/2}\|e(L^h_{\lambda^h}) - e(L_{\bar\lambda})\|_V = \mathcal{O}(h^s)$, and hence $\|u^h - \bar u\|_{W^{1,1}} = \mathcal{O}(h^s)$. □

Remarks. 1) An analogous result may be derived for the relaxed program (P_r) and its dual (D_r). The same rate of convergence is obtained, since the argument is based on properties of the function $\mu_{\min}$. We do not present the detailed statement.

2) Similar results for (D), (D_r) could be obtained for a finite difference approximation. The analogues of (4.4) are now

$$|\mu_{\min}(L) - \mu_{\min}(L^h)| \le C_L h \quad \text{and} \quad \|e(L) - e(L^h)\|_1 \le C_L h \tag{4.6}$$

(see [21, sect. 11]), and again one has to show that the constants behave like $C_{L_\lambda} = \mathcal{O}(|\lambda|)$ for the operators L_λ we have in mind. This is once again a problem of tracing constants in a classical derivation of (4.6), as given, e.g., in [21, sect. 11]. (And again, the situation is expected to be more complicated for multiple eigenvalues.)

Let us now consider the discretization of programs (D), (D_r). For simplicity we keep the notation $\Delta + \sum_{k=1}^m \lambda_k a_k = \Delta + \sum_{k=1}^m \lambda_k A^k$ for the discretized form of the operator L_λ. Switching from the conventional operator notation (N) to the usual matrix notation, we have $\mu_{\min}(L_\lambda) = -\mu_{\max}(L_\lambda)$, where $\mu_{\max}(L_\lambda)$ now denotes the largest eigenvalue of the matrix L_λ. Passing to the more convenient convex form, the dual (D) is now a semidefinite program

$$
\text{(D)} \qquad \begin{aligned} &\text{minimize} && \alpha b_0 - \sum_{k=1}^{m} \lambda_k b_k \\ &\text{subject to} && \alpha I - \Delta - \sum_{k=1}^{m} \lambda_k A^k \geq 0, \ \alpha \in \mathbb{R}, \ \lambda \in \mathbb{R}^m, \end{aligned}
$$

which can be solved using the duality theory as presented in [28, Thms. 12.1, 12.2] or [38], references to which we refer the reader for concise information on the subject.

Since the tolerance model (1.4) was based on the Euclidean norm, (D_r) does not directly lead to a semidefinite program. Solving (D_r) as it stands requires techniques from eigenvalue optimization as presented, e.g., in Overton [33] and Overton and Womersley [34].

Let us show how to obtain a semidefinite programming formulation for the restoration problem (1.4). The way to proceed is to replace the tolerance model (P_r), resp., its dual (D_r) by a penalty-type approach

$$
(\mathrm{P}_p) \qquad \begin{aligned} &\text{maximize} && \mathcal{I}(u) + \frac{C}{2}|Au - b|^2 \\ &\text{subject to} && \int_\Omega u \, dx = b_0. \end{aligned}
$$

The same duality scheme as applied in previous sections leads to the concave dual

$$
(\mathrm{D}_p) \qquad \text{maximize } h(\lambda) = \mu_{\min}(L_\lambda[\cdot]) b_0 + \sum_{k=1}^{m} \lambda_k b_k - \frac{2}{C}|\lambda|^2
$$

based on the same return formula (cf. Theorems 3.4, 3.5). The important fact is now that (P_r) and (P_p) are equivalent in the sense that, for a fixed $\epsilon > 0$, the solution $\bar{u}$ of (P_r) with tolerance level ϵ also solves (P_p) for a certain constant $C = C(\epsilon)$, and conversely, if $\hat{u}$ solves (P_p) with a given constant $C > 0$, then it also solves (P_r) with tolerance $\epsilon = \epsilon(C) = |A\hat{u} - b|$. This may be checked by writing down the necessary optimality conditions for both programs.

A disadvantage of $(\mathrm{P}_p), (\mathrm{D}_p)$ as compared to $(\mathrm{P}_r), (\mathrm{D}_r)$ is that it is harder to obtain a default value for the constant $C > 0$ than for the tolerance ϵ, the dependence $C = C(\epsilon)$ not being explicit. On the other hand, it is certainly more convenient to deal with (D_p) as far as eigenvalue optimization is concerned. Indeed, we obtain a semidefinite programming formulation for (D_p) by introducing a new variable t together with the constraint $t \geq |\lambda|^2 = \lambda^t \lambda$. The latter translates into $tI \geq \lambda\lambda^t$, which is recognized as a semidefinite constraint (see [38]). Indeed, (D_p) appears as

$$
(\mathrm{D}_p) \qquad \begin{aligned} &\text{minimize} && \alpha b_0 - \sum_{k=1}^{m} \lambda_k b_k + \frac{2}{C} t \\ &\text{subject to} && \alpha I - \Delta - \sum_{k=1}^{m} \lambda_k A^k \geq 0, \\ & && \begin{pmatrix} tI & \lambda \\ \lambda^t & 1 \end{pmatrix} \geq 0, \quad \lambda \in \mathbb{R}^m, \ t, \alpha \in \mathbb{R}. \end{aligned}
$$

While the tolerance model had to be replaced by the penalty-type model in order to deal with the Euclidean norm, the original form of (D_r) is suitable if in (1.4)

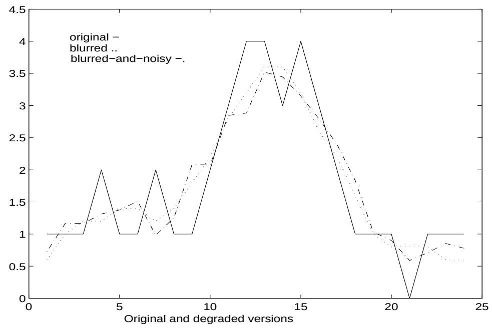

FIG. 1.

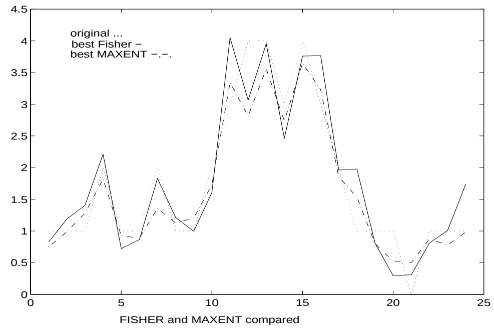

FIG. 2.

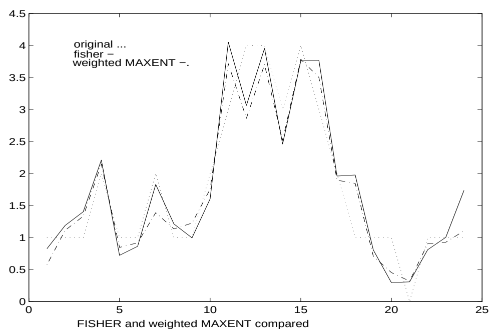

FIG. 3.

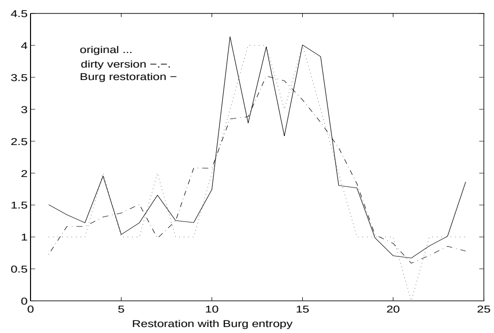

FIG. 4.

polyhedral norms like the one or infinity norm are used. In fact, starting with any norm $|Au-b|_*$ in (1.4) gives the dual norm expression $|\lambda|$ in (D_r). Now if, for instance, $|\cdot|$ in (D_r) denotes the one norm, a semidefinite programming formulation is obtained by introducing new variables λ_k^+, λ_k^- (satisfying $\lambda_k = \lambda_k^+ - \lambda_k^-$ and $\lambda_k^\pm \geq 0$):

$$(\mathrm{D}_r) \qquad \begin{array}{ll} \text{minimize} & \alpha b_0 - \displaystyle\sum_{k=1}^m \left(\lambda_k^+ - \lambda_k^-\right) b_k + \sum_{k=1}^m \left(\lambda_k^+ + \lambda_k^-\right) \\ \text{subject to} & \alpha I - \Delta - \displaystyle\sum_{k=1}^m \left(\lambda_k^+ - \lambda_k^-\right) A^k \geq 0, \\ & \lambda_k^+ \geq 0,\ \lambda_k^- \geq 0,\ \alpha \in \mathbb{R}. \end{array}$$

Similarly, when $|\cdot|$ in (D_r) denotes the infinity norm, the same decomposition leads to a term $\max_k \left(\lambda_k^+ + \lambda_k^-\right)$, which may be transformed to the semidefinite form as for instance in [38, sect. 1].

5. Experiments. This last section presents an experiment where we compared various restoration techniques.

Consider the setting (1.3), where a signal $u(t)$ has been blurred through convolution with a kernel $q(t)$ and by adding a random noise term. As a simulation, the piecewise linear original signal u on $[0, 1]$ displayed in Figure 1 (continuous line) was convolved by a pill-box blur $q(t)$ of variance .5, and white noise of variance $\sigma^2 = .4$ was added, leading to a signal-to-noise ratio of $16dB$. The degraded signal v was sampled at 25 equidistant points. The blurred (resp., blurred-and-noisy) images are also shown in Figure 1 as the dotted (resp., broken) line.

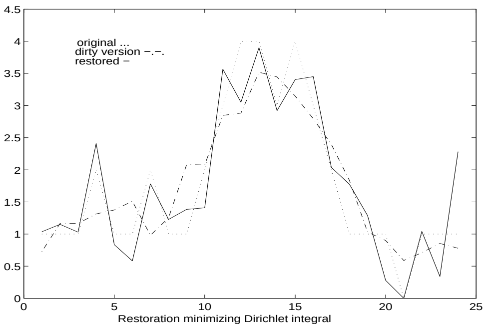

Fig. 5.

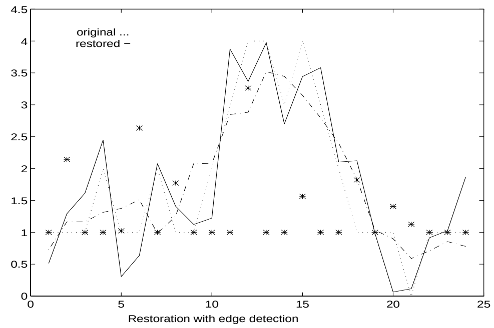

Fig. 6.

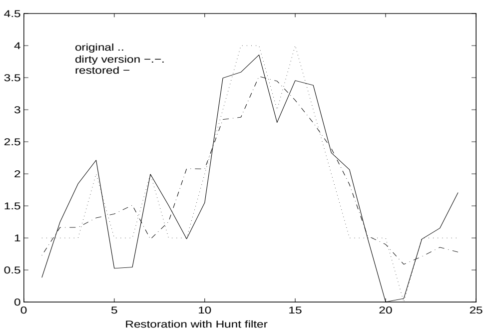

Fig. 7.

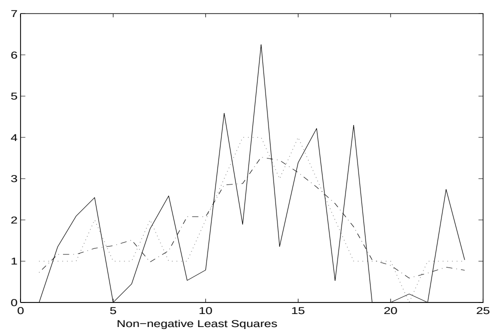

Fig. 8.

The Fisher restoration was obtained on the basis of model (D_p), using a discretization with $h = 1/50$, and the penalty constant was found experimentally. Figure 2 shows the Fisher restoration (continuous line), compared with a maximum entropy restoration (broken line) based on the model discussed, e.g., in [3, 10, 18, 22] and displayed against the original (dotted line). Figure 3 shows a comparison of the same Fisher restoration with a minimum cross-entropy method, where the dirty signal was used as a priori. As can be seen, with this weighting, the maximum entropy technique can catch up better with the Fisher restoration. For both entropy restorations, the tolerance model (1.4) was used, with the tolerance found experimentally, starting with the default as suggested by Example 1.2.

Figure 4 shows a restoration obtained by Burg's entropy$-\int \log u(x)\,dx$, displayed against the original and dirty versions. Not unexpectedly, Burg's method produced too spiky results, which agrees with experiments reported, e.g., in [6, 16]. Figure 5 shows the restoration obtained by minimizing the Dirichlet integral $\int |\nabla u(x)|^2 dx$ subject to the constraints (1.4) which for this example works surprisingly well, but generally has a tendency to produce restorations lacking in contrast. Figure 6 shows how this could be improved using a method along the lines of Geman and Geman [19], which combines deconvolution and edge detection. In its 1D discretized form it consists of minimizing a functional of the type

$$f(u, w) = \sum_{k=1}^{n-1} \frac{(u_{k+1} - u_k)^2}{w_k} + \alpha \sum_{k=1}^{n-1} (w_k - 1)$$

subject to the constraints (1.4) and $w_k \geq 1$. Here the variables w_k play the role of switching parameters. The state $w_k = 1$ corresponds to the switch being on, so that the gap $(u_{k+1} - u_k)^2$ is fully penalized. Any value $w_k > 1$ is meant to indicate a structural gap at position k in the unknown image u (or at least a tendency for such), the philosophy being that in this case the penalty for a jump at position k should be gradually released or even switched off. Switching off is paid for through the second term, with the control parameter α ruling the trade-off between the two. Notice that while Geman and Geman [19] use real switches, leading to a discrete variable with states on and off, we prefer a continuous version (using dimmers instead of switches). The states of the variables w_k in Figure 6 are represented by the stars.

Figure 7 shows a restoration obtained by Hunt's filter [23], which is based on minimizing the integral $\int |\Delta u(x)|^2 dx$ and which gives a linear inverse filter if the penalty approach is used. The final Figure 8 was included to show how a direct inversion of (1.3) using least squares badly fails due to the relatively high noise contribution.

Acknowledgments. We thank the editor, Stephen Boyd, and the two anonymous referees for their valuable comments. Helpful discussions with Jean-Baptiste Hiriart-Urruty (Toulouse), Adrian Lewis (Waterloo), and Martin Wursthorn (Stuttgart) during the preparation of this paper are gratefully acknowledged.

REFERENCES

[1] R. Adams, *Sobolev Spaces*, Academic Press, New York, 1975.

[2] D. L. Angwin and H. Kaufman, *Nonhomogeneous image identification and restoration procedures*, Springer Ser. Inform. Sci., 23 (1991), pp. 177–208.

[3] J. M. Borwein and A. S. Lewis, *Duality relationships for entropy-like minimization problems*, SIAM J. Control Optim., 29 (1991), pp. 325–338.

[4] J. M. Borwein and A. S. Lewis, *Partially finite convex programming* I, II, Math. Programming, 57 (1992), pp. 15–48 and 49–84.

[5] J. M. Borwein, A. S. Lewis, and D. Noll, *Maximum entropy reconstruction using derivative information* I*: Fisher information and convex duality*, Math. Oper. Res., 21 (1996), pp. 442–468.

[6] J. M. Borwein, A. S. Lewis, M. N. Limber, and D. Noll, *Maximum entropy reconstruction using derivative information* II*: Computational results*, Numer. Math., 69 (1995), pp. 243–256.

[7] J. M. Borwein, M. N. Limber, and D. Noll, *Fast heuristic methods for function reconstruction using derivative information*, Appl. Anal., 58 (1995), pp. 241–261.

[8] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed., Springer-Verlag, New York, 1991.

[9] K. J. Brown and S. S. Lin, *On the existence of positive eigenfunctions for an eigenvalue problem with indefinite weight functions*, J. Math. Anal. Appl., 75 (1980), pp. 112–120.

[10] S. F. Burch, S. F. Gull, and J. K. Skilling, *Image restoration by a powerful maximum entropy method*, Computer Vision, Graphics, and Image Processing, 23 (1983), pp. 113–128.

[11] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, North–Holland, Amsterdam, 1978.

[12] R. Courant, *Methoden der mathematischen Physik* I, Springer Verlag, Berlin, 1931.

[13] S. J. Cox and M. L. Overton, *Optimal design of a column against buckling*, SIAM J. Math. Anal., 23 (1992), pp. 287–325.

[14] B. Dacorogna, *Direct Methods in the Calculus of Variations*, Springer-Verlag, Berlin, 1989.

[15] J. L. Doob, *Probability and statistics*, Trans. Amer. Math. Soc., 36 (1934), pp. 759–775.

[16] G. J. Erickson and C. R. Smith, eds., *Maximum-Entropy and Bayesian Methods in Science and Engineering*, Kluwer, Dordrecht, The Netherlands, 1988.

[17] R. A. Fisher, *Theory of statistical estimation*, Philos. Trans. Roy. Soc. London, 222 (1930), pp. 309–368.

[18] B. R. Frieden, *Restoring with maximum likelihood and maximum entropy*, J. Opt. Soc. Amer., G2 (1972), pp. 511–518.

[19] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images*, IEEE Trans. Pattern Analysis and Machine Intelligence, 6 (1984), pp. 721–741.
[20] B. Ginsburg and A. Ioffe, *The maximum principle in optimal control of systems governed by semilinear equations*, in Proc. IMA Workshop on Nonsmooth Analysis and Geometric Methods in Deterministic Optimal Control, B. Mordukhovich and B. Sussman, eds., Springer-Verlag, 1995.
[21] W. Hackbusch, *Elliptic Differential Equations: Theory and Numerical Treatment*, Springer-Verlag, New York, 1992.
[22] R. M. Haralick, E. Østevold, and X. Zhuang, *A differential equation approach to maximum entropy image reconstruction*, IEEE Trans. Acoust. Speech Signal Process., 35 (1987), pp. 208–218.
[23] B. R. Hunt, *Deconvolution of linear systems by constrained regression and its relationship to the Wiener theory*, IEEE Trans. Automat. Control, 23 (1972), pp. 703–705.
[24] R. W. Johnson and J. E. Shore, *Which is the better entropy expression for speech processing: $-s \log s$ or $\log s$?*, IEEE Trans. Acoust., Speech Signal Proc., 32 (1984), pp. 129–137.
[25] T. Kato, *Perturbation Theory for Linear Operators*, 2nd ed., Springer-Verlag, New York, 1984.
[26] A. K. Katsaggelos and K.-T. Lay, *Maximum likelihood identification and restoration of images using the expectation-maximization algorithm*, Springer Series in Information Sciences, 23 (1991), pp. 143–176.
[27] A. S. Lewis, *The convergence of entropy-based approximations for moment problems*, Optimization, 28 (1994), pp. 383–395.
[28] A. S. Lewis and M. L. Overton, *Eigenvalue optimisation*, Acta Numerica, (1996).
[29] M. K. V. Murthy and G. Stampacchia, *A variational inequality with mixed boundary conditions*, Israel J. Math., 13 (1972), pp. 188–224.
[30] D. Noll, *Rates of convergence for best entropy estimates*, Stat. Decisions, 13 (1995), pp. 141–165.
[31] D. Noll, *Restoration of degraded images with maximum entropy*, J. Global Optim., 10 (1997), pp. 91–103.
[32] D. Noll, *Consistency of a nonlinear deconvolution method with applications in image restoration*, Adv. Math. Sci. Appl., 7 (1997), pp. 789–808.
[33] M. L. Overton, *Large scale optimisation of eigenvalues*, SIAM J. Optim., 2 (1992), pp. 88–120.
[34] M. L. Overton and R. S. Womersley, *Second derivatives for optimizing eigenvalues of symmetric matrices*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 697–718.
[35] R. T. Rockafellar, *Integrals which are convex functionals* I, II, Pacific J. Math., 24 and 39 (1968), pp. 525–539 and 439–469.
[36] R. N. Silver, *Quantum Statistical Inference*, IEEE Report, Kluwer, Dordrecht, The Netherlands, 1992.
[37] H. J. Trussell, *The relationship between image restoration by the maximum a posteriori method and a maximum entropy method*, IEEE Trans. Acous. Speech Signal Proc., 28 (1980), pp. 114–117.
[38] L. Vandenberghe and S. Boyd, *Semidefinite programming*, SIAM Rev., 38 (1996), pp. 49–95.
[39] W. P. Ziemer, *Weakly Differentiable Functions*, Springer-Verlag, New York, 1989.
[40] V. Zivojnovic, *Higher-order statistics and Huber's robustness*, in IEEE Signal Processing Workshop on Higher-Order Statistics, Spain, 1995.