Lagrangian methods for the regularization of discrete ill-posed problems

G. Landi

Abstract

In many science and engineering applications, the discretization of linear ill-posed problems gives rise to large ill-conditioned linear systems with right-hand side degraded by noise. The solution of such linear systems requires the solution of a minimization problem with one quadratic constraint depending on an estimate of the variance of the noise. This strategy is known as regularization. In this work, we propose to use Lagrangian methods for the solution of the noise constrained regularization problem. Moreover, we introduce a new method based on Lagrangian methods and the discrepancy principle. We present numerical results on numerous test problems, image restoration and medical imaging denoising. Our results indicate that the proposed strategies are effective and efficient in computing good regularized solutions of ill-conditioned linear systems as well as the corresponding regularization parameters. Therefore, the proposed methods are actually a promising approach to deal with ill-posed problems.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In many science and engineering applications it is necessary to compute an approximate solution of the linear system:

$$Ax = b, \quad A \in R^{m\times n}, \quad m \geq n, \tag{1.1}$$

where A is an ill-conditioned matrix. Troughout this work we will suppose A to be non singular. Usually, the right-hand side b is corrupted by noise, i.e:

$$b = \widetilde{b} + \eta \tag{1.2}$$

where $\widetilde{b}$ is the unknown noise-free right-hand side vector and η accounts for noise and measurement errors and it is usually not available.
Discrete ill-conditioned linear systems (1.1) arise from the discretization of continuous ill-posed problems such as Fredholm integral equations of the first kind. These integral equations are by nature ill-posed problems in the sense of Hadamard [27, 26], i.e., they have no solution or their solution, if it exists, does not depend continuously on the data b.
We would like to determine an approximation of the original solution $\widetilde{x}$ of the noise-free linear system

$$A\widetilde{x} = \widetilde{b}. \tag{1.3}$$

Unfortunately, due to the ill-conditioning of A and the noise on b, the direct solution of (1.1) is a poor approximation of the true solution $\widetilde{x}$. Therefore, to compute useful approximation of $\widetilde{x}$, it is necessary to modify the original problem (1.1) into a better conditioned one which is strictly related to (1.1) but incorporates additional information about the unknown solution $\widetilde{x}$. This strategy is referred to as *regularization* [33]. A common regularization technique consists in replacing the system (1.1) with the following equality constrained minimization problem:

$$\begin{aligned} &\text{minimize} \quad \phi(x) \\ &\text{subject to} \quad \frac{1}{2}\|Ax - b\|^2 = \frac{\sigma^2}{2}. \end{aligned} \tag{1.4}$$

where $\|\cdot\|$ denotes the Euclidean norm and σ is the variance of the noise η which is assumed to be available. The functional $\phi(x)$ is called *regularization functional*. It incorporates *a priori* information on the desired solution $\widetilde{x}$ and,

in some sense, it measures the irregularity of $\widetilde{x}$.
The quantity $\|Ax-b\|^2$ is called the *discrepancy* of the vector x. By solving the problem (1.4) we find a regularized solution x_{reg} that minimizes the regularization functional $\phi(x)$ over all the vectors $x \in \mathbb{R}^n$ with discrepancy equal to σ^2.
The regularization problem (1.4) includes classic *Tikhonov regularization* where

$$\phi(x) = \frac{1}{2}\|Lx\|^2. \tag{1.5}$$

Usually, the matrix $L \in \mathbb{R}^{p\times n}$, $p \leq n$, is the identity matrix or a discretization of the first or second order derivative operator. The so called Tikhonov regularization functional (1.5) measures the smoothness of x and penalizes the discontinuities in x.
When $\phi(x)$ is the discretization of the continuous Total Variation functional:

$$\Phi(\chi) = \int_\Omega |\nabla\chi(t)|dt, \tag{1.6}$$

where $\chi : \Omega \to \mathbb{R}$ and Ω is a bounded convex region in the d-dimensional space, the problem (1.4) is called *Total Variation regularization*. The discrete Total Variation functional allows to recover discontinuities in the vector x and therefore is well suited for image restoration. For an exhaustive analysis of the theory of regularization see, for example, [19, 33, 41, 29, 17, 19, 49, 18, 24] for Tikhonov regularization, and [25, 1, 16] for Total Variation regularization.
Actually, the optimization problem (1.4) is used to model many inverse problems in image restoration, computer tomography and seismic inversion and therefore it is a very important optimization problem in many science and engineering applications.
We stress out that the regularization problem is not usually formulated as the equality constrained problem (1.4) but as the following unconstrained problem:

$$\text{minimize} \quad \phi(x) + \frac{\lambda}{2}\|Ax-b\|^2, \quad \lambda > 0. \tag{1.7}$$

It can be proved ([17, 33, 3] and the references therein) that problems (1.4) and (1.7) are equivalent, provided that λ is the exact Lagrange multiplier of (1.4). The multiplier λ is usually called *regularization parameter* in the context of regularization and it has the crucial role of balancing the trade-off between the smoothness of the solution and the fidelity to the data b.
In the literature there exist other formulations of the regularization problem that, for A nonsingular, are equivalent to the previous ones; they are the unconstrained minimization problem:

$$\text{minimize} \quad \mu\phi(x) + \|Ax-b\|^2, \quad \mu > 0 \tag{1.8}$$

where $\mu = \frac{2}{\lambda}$ and, for the Tikhonov functional (1.5), the energy constrained minimization problem:

$$\begin{aligned} &\text{minimize} && \|Ax-b\|^2 \\ &\text{subject to} && \|Lx\|^2 = \epsilon^2 \end{aligned} \tag{1.9}$$

where ϵ is an estimate of the norm $\|L\widetilde{x}\|$ of the solution $\widetilde{x}$ of (1.3).
When dealing with the unconstrained problem (1.7), a crucial issue is the selection of a suitable approximation of the exact multiplier λ of (1.4). In fact, if

the selected value λ is too large, then the corresponding solution x_λ of (1.7) is contaminated by noise while, if λ is too small, x_λ is a poor approximation of $\widetilde{x}$. In the literature there exists a significant amount of works on the development of strategies for selecting the regularization parameter of (1.7) for the Tikhonov functional (1.5); we refer the reader to [2, 19, 23, 38] and the references therein. In practice however, given the variance σ, it is difficult to calculate a suitable value of the regularization parameter for the regularization problem (1.7) at a low computational cost.

When the norm σ of the variance is explicitly known, a common method for the selection of a suitable value of the regularization parameter is the *discrepancy principle* [2, 19, 23, 38, 37] that states that the regularization parameter λ should be chosen so that the associated solution x_λ has a discrepancy equal to the error σ^2. This approach may require the solution of many unconstrained problems (1.7) at a high computational cost. Some Newton-type methods for solving the discrepancy principle are presented in [35, 54]. In [20], a variant of the discrepancy principle is described. In [9], a numerical method based on Lanczos bidiagonalization and Gauss quadrature is introduced for computing by the discrepancy principle a suitable value of λ and the associated regularized solution x_λ. In [34], the authors determine the regularization parameter λ from a smaller dimensional problem. Other methods do not require the exact knowledge of σ but try to derive such information from the right-hand side b. They are basically heuristic methods. Very popular methods of such type are the L-curve criterion [33, 31] and the Generalized Cross Validation (GCV) criterion [33, 53, 21]. For large problems these criteria are computationally very expensive and therefore they are difficult to use for practical applications. Recently, in [8], the authors have proposed a computationally less expensive strategy based on computing an L-ribbon that contains the L-curve and its interior. For Tikhonov regularization, some strategies based on solving the quadratically constrained least squares problem (1.9) have been proposed. The algorithm presented in [11] makes use of "black box" solvers for the related unconstrained least squares problems. In [47, 46] a trust-region approach to (1.9) is discussed. In [22] the authors have presented a method for (1.9) based on Gauss quadrature; in [10] a modification of this latter method based on partial Lanczos bidiagonalization and Gauss quadrature is presented. The iterative scheme proposed in [14] is based on the works [46, 48] and requires the computation of the smallest eigenvalue and the corresponding eigenvector to determine the proper regularization parameter. In [7] a modular approach is proposed for solving the first order equation associated with (1.9). However, in spite of the numerous methods proposed in the literature, the selection of a suitable regularization parameter λ for Tikhonov regularization is still an open question (especially for large scale problems) and an active area of research. Moreover, to the best of the author' knowledge, little work has been made in developing methods for the choice of the regularization parameter for other regularization functionals, such as the Total Variation functional.

In this work we consider the formulation (1.4) of the regularization problem as a noise constrained minimization problem. Such formulation makes sense because it does not require to know a good estimate of the regularization parameter but it needs an estimate of the noise level which is achievable in most applications. In this work we propose to apply iterative Lagrangian methods [36, 5] to (1.4) in order to compute both the regularization parameter λ and the corresponding

regularized solution x_λ of (1.1). Moreover, we introduce a new regularization procedure based on iterations of the Lagrangian methods truncated according to the discrepancy principle. We have called the proposed methods Truncated Lagrangian methods. We discuss some properties of the new Truncated Lagrangian methods in the regularization framework. To evaluate the performance of the proposed approach, we have applied the Lagrangian methods and the new Truncated Lagrangian methods to a wide set of test problems, image restoration and image denoising.

The work is organized as follows. In Chapter 2 we analyze the noise constrained regularization problem (1.4) from a theoretical point of view. In Chapter 3 we present the basic Lagrangian methods and in Chapter 4 we introduce the proposed Truncated Lagrangian methods. In Chapter 5 we discuss some implementation details; in Chapter 6, we present the results of numerous numerical experiments to illustrate the performance of the discussed methods. Conclusions are given in Chapter 7.

Chapter 2

The noise constrained regularization problem

Let us consider the regularization problem formulated as a noise constrained minimization problem:

$$\begin{aligned} &\text{minimize} \quad \phi(x) \\ &\text{subject to} \quad \frac{1}{2}\|Ax-b\|^2 = \frac{\sigma^2}{2}. \end{aligned} \tag{2.1}$$

For easier notation, let $h : \mathbb{R}^n \to \mathbb{R}$ be the function

$$h(x) = \frac{1}{2}\|Ax-b\|^2 \tag{2.2}$$

and let Ω be the subset of $\mathbb{R}^n$ defined as

$$\Omega = \{x \in \mathbb{R}^n | h(x) \leq \frac{\sigma^2}{2}\}. \tag{2.3}$$

Throughout this work, we will make the following assumptions on the minimization problem (2.1).

Assumptions

A1. $\phi : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable.

A2. $\phi(x) \geq 0, \quad \phi(0) = 0.$

A3. ϕ is convex.

A4. $\text{Null}\{\phi\} \cap \Omega = \emptyset$ where $\text{Null}\{\phi\}$ is the nullspace of ϕ.

A5. $A \in \mathbb{R}^{m\times n}$, $m \geq n$, is nonsingular.

Remark 2.1. The preceding assumptions A1–A3 are verified by both the Tikhonov functional and the Total Variation functional.

In this chapter we analyze the noise constrained regularization problem (2.1) and we give results of existence and uniqueness of its solution. Such results are known for the Tikhonov and Total Variation functionals but are obtained here in the wider context of minimizing the convex functional ϕ under the equality noise constraint $h(x) = \sigma^2/2$.

Let us consider the following inequality minimization problem which is strictly related to (2.1)

$$\begin{aligned} &\text{minimize} \quad \phi(x) \\ &\text{subject to} \quad h(x) \leq \frac{\sigma^2}{2}. \end{aligned} \tag{2.4}$$

The well-known Weierstrass theorem ensures that problem (2.4) has a solution.

Theorem 2.1. *Assume assumption A.1 holds. Then* (2.4) *has a solution.*

Proof. The feasible set $\Omega \subset \mathbb{R}^n$ is compact and therefore, for the Weierstrass theorem, the continuous function ϕ achieves a minimum in Ω. □

The following theorem states that the inequality constraint of (2.4) is active at a solution.

Theorem 2.2. *Assume assumptions A1–A4 hold. If x is a solution of* (2.4)*, then*

$$h(x) = \frac{\sigma^2}{2}. \tag{2.5}$$

Proof. The feasible set Ω is convex and the objective functional ϕ is convex; then every local minimum of ϕ over Ω is a global minimum. Under the assumptions A1–A4, we have that the set of the global minima of the objective functional ϕ over all $\mathbb{R}^n$ is the null space of ϕ. If a solution x of (2.4) is in the interior of Ω, then $\nabla\phi(x) = 0$ and $x \in \text{Null}\{\phi\}$. Because of assumption A.4, we cannot have $x \in \text{Null}\{\phi\} \cap \Omega$, then we deduce that the solution x lies on the boundary of the feasible set Ω. □

The previous theorem 2.2 has the following very important consequence.

Corollary 2.1. *Under the hypothesis of the previous theorem, problem* (2.1) *is equivalent to* (2.4)*.*

Proof. The equivalence between (2.1) and (2.4) is an immediate consequence of the fact that the minimum of ϕ over Ω is reached for some x with $h(x) = \sigma^2/2$. □

The following theorem ensures the uniqueness of the solution of (2.4).

Theorem 2.3. *Assume assumptions A1–A5 hold. Then the solution x of* (2.4) *is unique.*

Proof. If x and y are solutions to (2.4) we have

$$\phi\left(\frac{x+y}{2}\right) \leq \frac{1}{2}\left(\phi(x) + \phi(y)\right) = \min \phi. \tag{2.6}$$

The last equality is true since every local minimum of ϕ is a global minimum. Moreover, since the feasible set Ω is convex, we have

$$\frac{1}{2}\left\|A\left(\frac{x+y}{2}\right)-b\right\|^2 \leq \frac{1}{2}\sigma^2 \tag{2.7}$$

with equality if and only if $Ax = Ay$. For theorem 2.2 we cannot have the strict inequality in (2.7), then $Ax = Ay$ and $x = y$ for the non singularity of A. □

The following theorem states that the Lagrange multiplier of (2.4) is strictly positive.

Theorem 2.4. *If x is the solution of* (2.4)*, let λ be the corresponding Lagrange multiplier. Then $\lambda > 0$.*

Proof. From the first-order necessary conditions for (2.4) we have

$$\begin{cases} \nabla\phi(x) + \lambda\nabla h(x) = 0, \\ \lambda\left(h(x) - \dfrac{\sigma^2}{2}\right) = 0, \\ \lambda \geq 0,\ x \in \Omega. \end{cases} \tag{2.8}$$

If $\lambda = 0$, then from the first equation of (2.8) it follows $\nabla\phi(x) = 0$. Hence, x is a global minimum of ϕ over $\mathbb{R}^n$ and $x \in \Omega \cap \text{Null}\{\phi\}$. Because of assumption A.4, x cannot belong to $\Omega \cap \text{Null}\{\phi\}$ and consequently, it must be $\lambda > 0$. □

The following theorem summarizes the previous results and shows that problem (2.1) is a well-posed problem.

Theorem 2.5. *Assume A1–A5 hold. Then the noise constrained regularization problem* (2.1) *has a unique solution x^* with positive Lagrange multiplier λ^*.*

Proof. The proof immediately follows from the equivalence between (2.1) and (2.4) (Corollary 2.1) and from the theorems 2.1, 2.3 and 2.4. □

Chapter 3

Lagrangian Methods

In this chapter we introduce the Lagrangian methods. We follow [36] for the description of the methods and we refer the reader to [36, 5, 4] for a broader treatment.
Lagrangian methods are essentially based on directly solving the nonlinear equations representing the first-order necessary conditions for the optimality of the constrained minimization problem. The first-order necessary conditions for (2.1) are expressed as:

$$\begin{aligned} \nabla_x \mathcal{L}(x, \lambda) &= 0, \\ h(x) - \frac{\sigma^2}{2} &= 0, \end{aligned} \tag{3.1}$$

where the *Lagrangian function* $\mathcal{L} : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}$ is defined by:

$$\mathcal{L}(x, \lambda) = \phi(x) + \lambda \left(h(x) - \frac{\sigma^2}{2} \right) \tag{3.2}$$

with $\lambda \in \mathbb{R}$ the Lagrange multiplier.

Equations (3.1) are usually referred to as *Lagrange equations* and represent a system of $(n + 1)$ nonlinear equations in the $(n + 1)$ variables $x_1, \ldots, x_n, \lambda$. A point x satisfying the first-order necessary conditions (3.1) for optimality is called *stationary point*. A pair (x, λ) satisfying (3.1) is called *stationary pair*.
When dealing with algorithms for constrained minimization problems, it is useful to introduce a merit function $m(x, \lambda) : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}$ to measure the progress of the algorithm towards a local minimum. Merit functions guide the algorithm by balancing the reduction of the objective function and the satisfaction of the constraints. The Lagrangian methods described in the following subsections are basically first-order and second-order methods for solving the Lagrange equations (3.1). They are characterized by a search direction and a suitable merit function which is minimized at a solution of (2.1) and serves as a descent function for the algorithm.

3.1 First-Order Lagrangian method

The first-order Lagrangian method consists in a first-order iterative procedure for solving the Lagrange equations (3.1). Thus, given an iterate (x_k, λ_k), the

method is defined by the formula:

$$\begin{aligned} x_{k+1} &= x_k + \alpha_k \Delta x_k \\ \lambda_{k+1} &= \lambda_k + \alpha_k \Delta \lambda_k \end{aligned} \tag{3.3}$$

where the direction $(\Delta x_k, \Delta \lambda_k) \in \mathbb{R}^{n+1}$ is:

$$\begin{aligned} \Delta x_k &= -\nabla_x \mathcal{L}(x_k, \lambda_k), \\ \Delta \lambda_k &= h(x_k) - \frac{\sigma^2}{2}. \end{aligned} \tag{3.4}$$

The scalar step-length $\alpha_k > 0$ is such that:

$$\alpha_k = \underset{\alpha}{\operatorname{argmin}} \{ m(x_k + \alpha \Delta x_k, \lambda_k + \alpha \Delta \lambda_k) \} \tag{3.5}$$

where the merit function $m(x, \lambda)$ is defined as follows:

$$m(x, \lambda) = \frac{1}{2} \left(\|\nabla_x \mathcal{L}(x, \lambda)\|^2 + |h(x) - \frac{\sigma^2}{2}|^2 \right) - \gamma \mathcal{L}(x, \lambda) \tag{3.6}$$

for some $\gamma > 0$. It is possible to prove the convergence of the first-order Lagrangian method to a solution of the original problem (2.1) (see [36]).
In our implementation we have computed an approximate minimizer α_k of the merit function $m(x_k + \alpha \Delta x_k, \lambda_k + \alpha \Delta \lambda_k)$; i.e. we have selected the step-length α_k so that a "sufficient decrease" in the merit function has been produced along the direction $(\Delta x_k, \Delta \lambda_k)$. More specifically, we require α_k to be the first element of the sequence $\{2^{-i}\}_{i=0}^{\infty}$ satisfying the Armijo's condition:

$$\begin{aligned} &m(x_k + \alpha_k \Delta x_k, \lambda_k + \alpha_k \Delta \lambda_k) \\ &\qquad \leq m(x_k, \lambda_k) + \mu \alpha_k \left(\Delta x_k, \Delta \lambda_k \right)^t \nabla_{(x,\lambda)} m(x_k, \lambda_k) \end{aligned} \tag{3.7}$$

with $\mu = 10^{-4}$.
The iterative procedure (3.3) is terminated as soon as one of the following stopping criteria is satisfied.

$$\text{i)} \quad k \geq k_{\max} \tag{3.8a}$$

$$\text{ii)} \quad \left\| \begin{pmatrix} \nabla_x \mathcal{L}(x_k, \lambda_k) \\ h(x_k) - \frac{\sigma^2}{2} \end{pmatrix} \right\| \leq \tau_1 \left\| \begin{pmatrix} \nabla_x \mathcal{L}(x_0, \lambda_0) \\ h(x_0) - \frac{\sigma^2}{2} \end{pmatrix} \right\| \tag{3.8b}$$

$$\text{iii)} \quad \left\| \alpha_k \begin{pmatrix} \Delta x_k \\ \Delta \lambda_k \end{pmatrix} \right\| \leq \tau_2 \tag{3.8c}$$

$$\text{iv)} \quad |\alpha_k \Delta \lambda_k| \leq \tau_3 \tag{3.8d}$$

where $k_{\max}$ is the maximum number of allowed iterations and τ_1, τ_2, τ_3 are given tolerances. Criteria (3.8a)–(3.8c) are standard stopping criteria for Lagrangian methods; criterium (3.8d) requires a comment. In our experiments, we have observed that, for many problems, the value of the multiplier λ_k significantly changes only in the first iterations and then it remains almost constant, even if the corresponding iterate x_k is far from the solution. The first-order Lagrangian method has the main advantage of requiring only the computation of the first

derivatives but it is well known to converge very slowly. Therefore we propose to use the first-order Lagrangian method only to provide a good estimate λ_{k+1} of the true multiplier λ by stopping the iterative procedure when $|\lambda_{k+1} - \lambda_k|$ becomes less or equal to the tolerance τ_3. An approximation of the solution of the original problem (1.1) is then obtained by solving the unconstrained minimization problem

$$\text{minimize}_x \ \phi(x) + \frac{\lambda_{k+1}}{2}\|Ax - b\|^2. \tag{3.9}$$

with the computed value λ_{k+1} of the multiplier.
We can now outline the algorithm as follows.

ALGORITHM 3.1 *First-Order Lagrangian Method*

0. Compute an initial iterate $(x_0, \lambda_0) \in \mathbb{R}^{n+1}$; set $k = 0$.

Repeat

1. Compute:

$$\Delta x_k = -\nabla_x \mathcal{L}(x_k, \lambda_k),$$
$$\Delta \lambda_k = h(x_k) - \frac{\sigma^2}{2}.$$

2. Find the first number α_k of the sequence $\{1, \frac{1}{2}, \frac{1}{4}, \ldots, \frac{1}{2^i}, \ldots\}$ such that:

$$m(x_k + \alpha_k \Delta x_k, \lambda_k + \alpha_k \Delta \lambda_k) \leq m(x_k, \lambda_k) + \mu \alpha_k \left(\Delta x_k, \Delta \lambda_k\right)^t \nabla_{(x,\lambda)} m(x_k, \lambda_k)$$

with $\mu = 10^{-4}$.

3. Set:

$$x_{k+1} = x_k + \alpha_k \Delta x_k,$$
$$\lambda_{k+1} = \lambda_k + \alpha_k \Delta \lambda_k.$$

4. Set $k = k + 1$ and return to Step **1**.

until stopping criteria (3.8) are satisfied.

3.2 Second-Order Lagrangian method

Let us consider the nonlinear system of equations (3.1) representing the first-order necessary conditions for (2.1). An iteration of the Newton method to solve (3.1) is stated as

$$\begin{pmatrix} \nabla^2_{xx}\mathcal{L}(x,\lambda) & \nabla_x h(x) \\ \nabla_x h(x)^t & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta \lambda \end{pmatrix} = - \begin{pmatrix} \nabla_x \mathcal{L}(x,\lambda) \\ h(x) - \frac{\sigma^2}{2} \end{pmatrix} \tag{3.10}$$

where (x, λ) is the current iterate and $(\Delta x, \Delta\lambda)$ is the search direction for the next iterate.
Let us consider the merit function $m : \mathbb{R}^{n+1} \to \mathbb{R}$ defined as

$$m(x, \lambda) = \frac{1}{2}\Big(\|\nabla_x \mathcal{L}(x, \lambda)\|^2 + w|h(x) - \frac{\sigma^2}{2}|^2\Big), \tag{3.11}$$

with $w \in \mathbb{R}$ a positive parameter.
The following results show that the merit function $m(x, \lambda)$ is a well-behaved merit function for (2.1).

Theorem 3.1. *Suppose that x^* and λ^* satisfy the first-order necessary conditions for a local minimum of $m(x, \lambda)$ with respect to x and λ. Suppose also that the rank of the gradient $\nabla h(x^*)$ at x^* is full and that the hessian matrix $\nabla^2_{xx}\mathcal{L}(x^*, \lambda^*)$ is positive definite. Then, (x^*, λ^*) is a global minimum point of $m(x, \lambda)$ with value $m(x^*, \lambda^*) = 0$.*

For a proof see §14.2 of [36].

Corollary 3.1. *Under the hypothesis of theorem 3.1, (x^*, λ^*) is a stationary pair for* (2.1).

Proof. From $m(x^*, \lambda^*) = 0$ it immediately follows $\nabla\mathcal{L}(x^*, \lambda^*) = 0$ and $h(x^*) = \sigma^2/2$. □

3.2.1 The algorithm

In this section we present a second-order Lagrangian method which is essentially based on Newton iterations to solve the Lagrange equations (3.1). Besides, we show that the algorithm is well-defined.
The algorithm is formally described as follows.

ALGORITHM 3.2 *Second-Order Lagrangian Method*

0. *Data.* Compute an initial iterate $(x_0, \lambda_0) \in \mathbb{R}^{n+1}$; set $k = 0$.

Repeat

1. Compute the matrix:

$$M(x_k, \lambda_k) = \begin{pmatrix} \nabla^2_{xx}\mathcal{L}(x_k, \lambda_k) & \nabla_x h(x_k) \\ \nabla_x h(x_k)^t & 0 \end{pmatrix}. \tag{3.12}$$

2. *Computation of the search direction.*
Compute the search direction $(\Delta x_k, \Delta\lambda_k)$ by solving the linear system:

$$M(x_k, \lambda_k)\begin{pmatrix} \Delta x_k \\ \Delta\lambda_k \end{pmatrix} = -\begin{pmatrix} \nabla_x \mathcal{L}(x_k, \lambda_k) \\ h(x_k) - \frac{\sigma^2}{2} \end{pmatrix}. \tag{3.13}$$

3. *Line search.*
Find the first number α_k of the sequence $\{1, \frac{1}{2}, \frac{1}{4}, \ldots, \frac{1}{2^r}, \ldots\}$ satisfying:

$$\text{i)} \quad m(x_k + \alpha_k \Delta x_k, \lambda_k + \alpha_k \Delta \lambda_k) \leq m(x_k, \lambda_k) + \mu \alpha_k \left(\Delta x_k, \Delta \lambda_k\right)^t \nabla_{(x,\lambda)} m(x_k, \lambda_k) \quad \text{with } \mu = 10^{-4}; \tag{3.14a}$$

$$\text{ii)} \quad \lambda_k + \alpha_k \Delta \lambda_k > 0; \tag{3.14b}$$

$$\text{iii)} \quad A(x_k + \alpha_k \Delta x_k) - b \neq 0. \tag{3.14c}$$

4. *Updates.*
Set:

$$\begin{aligned} x_{k+1} &= x_k + \alpha_k \Delta x_k, \\ \lambda_{k+1} &= \lambda_k + \alpha_k \Delta \lambda_k. \end{aligned} \tag{3.15}$$

Set $k = k + 1$ and return to Step **1**.

until $\left\| \left(\nabla_x \mathcal{L}(x_k, \lambda_k), h(x_k) - \frac{\sigma^2}{2} \right) \right\| = 0$

The algorithm includes, in step 3, an inexact line search based on the Armijo's procedure (condition (3.14a)) [42] for unconstrained optimization that ensures a reasonable decrease of the merit function. In addition, condition (3.14b) forces the Lagrange multipliers to be strictly positive. From theorem 2.5 we have that (2.1) is equivalent to (2.4) and that the Lagrange multiplier of the solution of (2.1) is positive. Therefore, we can consider the presented algorithm as a sort of active set strategy applied to the inequality constrained problem (2.4) in which we treat the inequality constraint as an active constraint and we force the Lagrange multiplier to be positive on the basis of the prior knowledge on the solution provided by theorem 2.5. Finally, condition (3.14c) ensures that the matrix $M(x_k, \lambda_k)$ is nonsingular. It is a nonrestrictive condition as, obviously, if x is a solution of (2.1), then $Ax - b \neq 0$. In the following we prove that the algorithm is well-defined, i.e. the coefficient matrix $M(x_k, \lambda_k)$ of the linear system (3.13) is nonsingular and a positive step-length α_k satisfying (3.14) can be always computed. For easier presentation, in the following of this subsection, we omit the index k.

Proposition 3.1. *The direction $(\Delta x, \Delta \lambda) \in \mathbb{R}^{n+1}$ computed at step 2 of the algorithm is a descent direction for the merit function $m(x, \lambda)$ at (x, λ).*

Proof. Given $(\Delta x, \Delta\lambda)$ satisfying (3.13), we have:

$$
\begin{aligned}
&(\Delta x, \Delta\lambda)^t \nabla_{(x,\lambda)} m(x,\lambda) = \\
&= (\Delta x, \Delta\lambda)^t \begin{pmatrix} \nabla^2_{xx}\mathcal{L}(x,\lambda)\nabla_x\mathcal{L}(x,\lambda) + w\left(h(x) - \frac{\sigma^2}{2}\right)\nabla_x h(x) \\ (\nabla_x\mathcal{L}(x,\lambda))^t \nabla_x h(x) \end{pmatrix} = \\
&= \Delta x^t \nabla^2_{xx}\mathcal{L}(x,\lambda)\nabla_x\mathcal{L}(x,\lambda) + \\
&\qquad + w\left(h(x) - \frac{\sigma^2}{2}\right)\Delta x^t \nabla_x h(x) + \Delta\lambda \nabla_x h(x)^t \nabla_x \mathcal{L}(x,\lambda)) = \\
&= \left(\Delta x^t \nabla^2_{xx}\mathcal{L}(x,\lambda) + \Delta\lambda\nabla_x h(x)^t\right)\nabla_x\mathcal{L}(x,\lambda) + w\left(h(x) - \frac{\sigma^2}{2}\right)\Delta x^t \nabla_x h(x) = \\
&= -(\nabla_x\mathcal{L}(x,\lambda))^t \nabla_x\mathcal{L}(x,\lambda) - w\left(h(x) - \frac{\sigma^2}{2}\right)^2.
\end{aligned}
$$

This is strictly negative unless $\nabla_x\mathcal{L}(x,\lambda) = 0$ and $h(x) - \frac{\sigma^2}{2} = 0$. □

Remark 3.1. Notice that, if both $\nabla_x\mathcal{L}(x,\lambda) = 0$ and $h(x) - \frac{\sigma^2}{2} = 0$ then the algorithm stops and the current iterate (x,λ) is a stationary pair of (2.1).

Proposition 3.2. *Given a pair (x,λ) such that $Ax - b \neq 0$ and $\lambda > 0$, the matrix*

$$
M(x,\lambda) = \begin{pmatrix} \nabla^2_{xx}\mathcal{L}(x,\lambda) & \nabla_x h(x) \\ \nabla_x h(x)^t & 0 \end{pmatrix} \tag{3.16}
$$

is nonsingular.

Proof. Because of assumptions A3 and A5 (chapter 2, pp. 9), the hessian matrix $\nabla^2\mathcal{L}(x,\lambda) = \nabla^2\phi(x,\lambda) + \lambda A^t A$ is positive definite if $\lambda > 0$. However, the matrix $M(x,\lambda)$ is always indefinite (see §16.2 of [42]). Since $Ax - b \neq 0$, the vector $\nabla_x h(x) = A^t(Ax - b)$ is trivially a maximum rank matrix, therefore $M(x,\lambda)$ is nonsingular (see §14.1 of [36]). □

As a consequence of proposition 3.2, we have that the search direction $(\Delta x, \Delta\lambda)$ is uniquely determined. The following proposition shows that the line search procedure in step 3 is well-defined.

Proposition 3.3. *Given a pair (x,λ) such that $Ax - b \neq 0$ and $\lambda > 0$, and a search direction $(\Delta x, \Delta\lambda)$ satisfying* (3.13)*, there exists $\overline{\alpha} > 0$ such that conditions* (3.14) *are verified for any $\alpha \in]0, \overline{\alpha}]$.*

Proof. The merit function $m(x,\lambda)$ is continuous and bounded below. Since $(\Delta x, \Delta\lambda)$ is a descent direction for $m(x,\lambda)$, there exists α_1 such that $(x + \alpha\Delta x, \lambda + \alpha\Delta\lambda)$ satisfies condition (3.14a) for any $\alpha \in]0, \alpha_1]$.
As $\lambda > 0$, there exists $\alpha_2 > 0$ such that $\lambda + \alpha\Delta\lambda > 0$ for all $\alpha \in]0, \alpha_2]$; in particular, if $\Delta\lambda > 0$ then $\lambda + \alpha\Delta\lambda > 0$ for each $\alpha > 0$ and $\alpha_2 = \infty$ while, if $\Delta\lambda < 0$, $\alpha_2 = -\frac{\lambda}{\Delta\lambda}$.
Finally, let us consider the continuous function

$$
r(\alpha) = \|A(x + \alpha\Delta x) - b\|. \tag{3.17}
$$

Since $r(0) = \|Ax - b\| > 0$, there exists a neighborhood $[-\alpha_3, \alpha_3]$ of 0 such that $r(\alpha) > 0$ for any $\alpha \in [-\alpha_3, \alpha_3]$. Let $\overline{\alpha} = \inf\{\alpha_1, \alpha_2, \alpha_3\}$; then the conditions (3.14) hold for any $\alpha \in]0, \overline{\alpha}]$. □

3.2.2 Global Convergence

In this section we prove that either the algorithm stops at a solution, or any sequence $\{(x_k, \lambda_k)\}_{k\in\mathbb{N}}$ generated by the algorithm converges to the solution of (2.1).
In order to prove the global convergence result, the following additional assumption is required.

Assumption A6 The points (x_k, λ_k) generated by the algorithm lie in a compact set.

We firstly remark that, if the algorithm stops at the kth iteration, then the current iterate (x_k, λ_k) is a stationary pair of (2.1). Since ϕ and h are convex functions, the iterate x_k is a global minimum of (2.4). Suppose now that the algorithm never stops and an infinite sequence $\{(x_k, \lambda_k)\}_{k\in\mathbb{N}}$ is generated. The following result establishes the global convergence of the sequence generated by the algorithm to the solution of (2.1).

Theorem 3.2. *The sequence $\{(x_k, \lambda_k)\}_{k\in\mathbb{N}}$ generated by the algorithm converges to the solution of* (2.1).

Proof. Since the sequence $\{(x_k, \lambda_k)\}_{k\in\mathbb{N}}$ is contained in a compact, it has a convergent subsequence. Therefore, we can extract a subsequence $\{(x_k, \lambda_k)\}_{k\in\mathcal{K}}$, $\mathcal{K} \subset \mathbb{N}$, converging to a point (x^*, λ^*). Since the merit function $m(x, \lambda)$ is continuous, it follows that

$$m(x_k, \lambda_k) \longrightarrow m(x^*, \lambda^*) \quad \text{for} \quad k \in \mathcal{K}.$$

From condition (3.14a) it follows

$$m(x_{k+1}, \lambda_{k+1}) \leq m(x_k, \lambda_k) \quad \text{for} \quad k \in \mathbb{N}.$$

This means that the sequence $\{(x_k, \lambda_k)\}_{k\in\mathbb{N}}$ is a decreasing sequence and that:

$$m(x^*, \lambda^*) \leq m(x_k, \lambda_k) \quad \text{for} \quad k \in \mathcal{K}.$$

Using the convergence of the merit function m on the subsequence $\mathcal{K}$ and the definition of limit, we have that, given $\varepsilon > 0$, there exists $k_\varepsilon \in \mathcal{K}$ such that for $k > k_\varepsilon$ and $k \in \mathcal{K}$

$$m(x_k, \lambda_k) - m(x^*, \lambda^*) < \varepsilon \quad \text{for} \quad k \in \mathcal{K}.$$

Let $\ell \in \mathcal{K}$ be the element which follows k_ε in $\mathcal{K}$; thus, for all $k > \ell$, $k \in \mathbb{N}$, we have

$$m(x_k, \lambda_k) - m(x^*, \lambda^*) = \\ m(x_k, \lambda_k) - m(x_\ell, \lambda_\ell) + m(x_\ell, \lambda_\ell) - m(x^*, \lambda^*) < \varepsilon.$$

Summarizing, given $\varepsilon > 0$, there exists $\ell > 0$ such that for $k > \ell$, $k \in \mathbb{N}$

$$m(x_k, \lambda_k) - m(x^*, \lambda^*) < \varepsilon$$

that is

$$m(x_k, \lambda_k) \longrightarrow m(x^*, \lambda^*) \quad \text{for} \quad k \in \mathbb{N}.$$

By the Armijo's rule (condition (3.14a)) we have

$$m(x_{k+1}, \lambda_{k+1}) - m(x_k, \lambda_k) \leq \eta \alpha_k (\Delta_k, \Delta\lambda_k)^t \nabla_{(x,\lambda)} m(x_k, \lambda_k)$$

and taking the limit on both sides we have

$$0 \leq \lim_{k \to \infty} \eta \alpha_k (\Delta x_k, \Delta\lambda_k)^t \nabla_{(x,\lambda)} m(x_k, \lambda_k).$$

Since $\eta \alpha_k (\Delta_k, \Delta\lambda_k)^t \nabla_{(x,\lambda)} m(x_k, \lambda_k) < 0$ for all k, in the limit we have

$$\lim_{k \to \infty} \eta \alpha_k (\Delta x_k, \Delta\lambda_k)^t \nabla_{(x,\lambda)} m(x_k, \lambda_k) \leq 0.$$

These inequalities imply

$$\lim_{k \to \infty} \eta \alpha_k (\Delta x_k, \Delta\lambda_k)^t \nabla_{(x,\lambda)} m(x_k, \lambda_k) = 0.$$

Since $\frac{1}{\alpha_k} \geq 1 > 0$ for all k, the product sequence $\frac{1}{\alpha_k} \alpha_k (\Delta x_k, \Delta\lambda_k)^t \nabla_{(x,\lambda)} m(x_k, \lambda_k)$ has limit

$$\lim_{k \to \infty} (\Delta x_k, \Delta\lambda_k)^t \nabla_{(x,\lambda)} m(x_k, \lambda_k) = 0.$$

The search direction satisfies (3.13), hence

$$(\Delta x_k, \Delta\lambda_k)^t \nabla_{(x,\lambda)} m(x_k, \lambda_k) = -\left\| \nabla_x \mathcal{L}(x_k, \lambda_k) \right\|^2 - w \left| h(x_k) - \frac{\sigma^2}{2} \right|^2;$$

consequently, taking the limit for $k \to \infty$, $k \in \mathcal{K}$, we have

$$\lim_{\substack{k \to \infty \\ k \in \mathcal{K}}} \left(-\left\| \nabla_x \mathcal{L}(x_k, \lambda_k) \right\|^2 - w \left| h(x_k) - \frac{1}{\sigma^2} \right|^2 \right) =$$

$$= -\left\| \nabla_x \mathcal{L}(x^*, \lambda^*) \right\|^2 - w \left| h(x^*) - \frac{1}{\sigma^2} \right|^2 = 0.$$

The last equation implies that the limit point (x^*, λ^*) satisfies the first-order necessary conditions of (2.1):

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0,$$
$$h(x^*) - \frac{\sigma^2}{2} = 0.$$

Next, we prove that (x^*, λ^*) is a solution of (2.1). From the condition (3.14b) we have $\lambda^* \geq 0$, therefore (x^*, λ^*) is a Kuhn-Tucker pair of (2.4). Since ϕ and h are convex functions, the point x^* is a global minimum of (2.4) and, in consequence of corollary 2.1, x^* is also a global minimum of (2.1).
To conclude the proof we have to prove the convergence of the entire sequence $\{(x_k, \lambda_k)\}_{k \in \mathbb{N}}$ to (x^*, λ^*). By contradiction assume that the sequence $\{(x^k, \lambda^k)\}_{k \in \mathbb{N}}$ does not converge to (x^*, λ^*). Therefore, it exists an index set $\mathcal{K}_1 \neq \mathcal{K}$ such that

$$(x_k, \lambda_k) \longrightarrow (\overline{x}, \overline{\lambda}) \quad \text{for} \quad k \in \mathcal{K}_1$$

and $(\overline{x}, \overline{\lambda}) \neq (x_*, \lambda_*)$. As before, we can prove that $(\overline{x}, \overline{\lambda})$ is a solution of (2.1). From theorem 2.5, we have that the solution of (2.1) is unique and therefore $(\overline{x}, \overline{\lambda}) = (x_*, \lambda_*)$. But this result is a contradiction with the fact that $(\overline{x}, \overline{\lambda}) \neq (x_*, \lambda_*)$; thus we have

$$(x_k, \lambda_k) \longrightarrow (x_*, \lambda_*) \quad \text{for} \quad k \in \mathbb{N}.$$

This concludes the proof. □

3.2.3 Implementation issues

In our implementation, the linear system (3.13) is solved inexactly by the Generalized Minimal RESidual (GMRES) method. The GMRES iterations are terminated when the relative residual of (3.13) is smaller then a tolerance τ_{GMRES} or after a maximum of $\text{maxit}_{\text{GMRES}}$ iterations. Notice that the direction $(\Delta x, \Delta\lambda)$ computed by the truncated GMRES method is not guaranteed to be a descent direction for the merit function. The GMRES method is known to converge to the solution of (3.13) within a finite number of iterations; then, for a sufficiently large number of iterations, the computed direction is a descent direction for m. In our implementation, if the computed direction is not a descent direction, we decrease the tolerance τ_{GMRES} and we allow $\text{Maxit}_{\text{GMRES}} \gg \text{maxit}_{\text{GMRES}}$ iterations to be performed. If even after $\text{Maxit}_{\text{GMRES}}$ a descent direction has not been found, we stop the algorithm. We underline that in all the performed numerical experiments this eventuality has never happened and a descent direction $(\Delta x, \Delta\lambda)$ has been always found within $\text{maxit}_{\text{GMRES}}$ iterations of the GMRES method.
In our implementation of the algorithm, the iterative procedure is terminated as soon as one of the following stopping criteria is satisfied.

$$\text{i)} \quad k \geq k_{\max} \tag{3.18a}$$

$$\text{ii)} \quad \left\| \begin{pmatrix} \nabla_x \mathcal{L}(x_k, \lambda_k) \\ \frac{1}{2} h(x_k) - \frac{1}{2}\sigma^2 \end{pmatrix} \right\| \leq \tau_1 \left\| \begin{pmatrix} \nabla_x \mathcal{L}(x_0, \lambda_0) \\ \frac{1}{2} h(x_0) - \frac{1}{2}\sigma^2 \end{pmatrix} \right\| \tag{3.18b}$$

$$\text{iii)} \quad \left\| \alpha_k \begin{pmatrix} \Delta x_k \\ \Delta \lambda_k \end{pmatrix} \right\| \leq \tau_2 \tag{3.18c}$$

where $k_{\max}$ is the maximum number of allowed iterations and τ_1, τ_2 are given tolerances.
In the following, for easier notation, we will refer to the second-order Lagrangian method simply as Lagrangian method.

Chapter 4

Truncated Lagrangian Methods

In this chapter we present a novel regularization technique, termed *Truncated Lagrangian method*, which is based on iterations of a Lagrangian method truncated according to the discrepancy principle.
Let us consider the equality constrained minimization problem:

$$\begin{aligned} &\text{minimize} \quad \phi(x) \\ &\text{subject to} \quad h(x) = 0 \end{aligned} \tag{4.1}$$

where $h(x) = \frac{1}{2}\|Ax-b\|^2$. Under the assumption A5 (chapter 2, pp. 9), the feasible set for problem (4.1) contains only the single point x^* that solves the linear system (1.1). The idea underlying the proposed method is to apply an iterative procedure to (4.1) in order to generate a succession of points $\{x_k\}$ converging to the solution x^* of (4.1) and to stop the iterations before convergence as soon as an iterate x_k satisfying the discrepancy principle has been determined. Consequently, the proposed Truncated Lagrangian method is basically an iterative regularization method [19, 33] with a Tikhonov-like prior information term.
Before describing the proposed approach, we will briefly review the discrepancy principle.

Definition 4.1. If the ill-posed problem (1.1) is consistent, i.e., $A\widetilde{x} = \widetilde{b}$ and the error norm σ is known, then an approximated solution $\widehat{x}$ to (1.1) is said to satisfy the discrepancy principle if:

$$\|A\widehat{x} - b\| \leq \rho\sigma \tag{4.2}$$

where $\rho > 1$ is an a-priori parameter.

For an iterative regularization method, stopping the iteration according to the discrepancy principle (4.2) means to stop the algorithm when the norm of the residual vector $\|Ax_k - b\|$ of the current iterate x_k satisfies:

$$\|Ax_k - b\| \leq \rho\sigma \tag{4.3}$$

where $\rho > 1$ is fixed.
In order to determine a regularized solution to (1.1), we would like to apply

one of the Lagrangian methods of the previous chapter 3 to (4.1) and to select the termination number of iterations according to the discrepancy principle. As described in the previous chapter 3, the Lagrangian methods can be viewed as methods for solving the first-order necessary optimality conditions for (4.1). Unfortunately, the first-order necessary conditions do not apply to (4.1) because its solution x^* is not a regular point, i.e. the gradient $\nabla_x h(x^*) = A^t(Ax^* - b)$ is the null vector. Therefore, in order to be able to apply the Lagrangian methods, we slightly modify problem (4.1) by adding a small positive constant ε to the right-hand side of the equality constraint. In this way we obtain the following perturbed equality constrained problem:

$$\begin{aligned} &\text{minimize} \quad \phi(x) \\ &\text{subject to} \quad h(x) = \varepsilon \end{aligned} \tag{4.4}$$

where the positive constant ε is fixed to be much smaller than $\frac{(\rho\sigma)^2}{2}$. The solution x_ε^* of (4.4) is now a regular point which satisfies, together with the corresponding multiplier λ_ε^*, the first-order necessary conditions

$$\begin{aligned} &\nabla_x \mathcal{L}(x_\varepsilon^*, \lambda_\varepsilon^*) = 0, \\ &h(x_\varepsilon^*) = \varepsilon, \end{aligned} \tag{4.5}$$

where the *Lagrangian function* $\mathcal{L} : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}$ is defined by:

$$\mathcal{L}(x, \lambda) = \phi(x) + \lambda(h(x) - \varepsilon). \tag{4.6}$$

In the following sections we will describe both the first-order Truncated Lagrangian method and the second-order Truncated Lagrangian method for computing an approximated solution to (1.1).

4.1 First-Order Truncated Lagrangian method

The first-order Truncated Lagrangian method consists in a truncated first-order iterative procedure for solving the Lagrange equations (4.5). Thus, given an iterate (x_k, λ_k), the method is defined by the formula:

$$\begin{aligned} x_{k+1} &= x_k + \alpha_k \Delta x_k \\ \lambda_{k+1} &= \lambda_k + \alpha_k \Delta \lambda_k \end{aligned} \tag{4.7}$$

where the direction $(\Delta x_k, \Delta \lambda_k)^t \in \mathbb{R}^{n+1}$ is:

$$\begin{aligned} \Delta x_k &= -\nabla_x \mathcal{L}(x_k, \lambda_k) \\ \Delta \lambda_k &= h(x_k) - \varepsilon. \end{aligned} \tag{4.8}$$

The scalar step-length $\alpha_k > 0$ is such that:

$$\alpha_k = \underset{\alpha}{\operatorname{argmin}} \{m(x_k + \alpha \Delta x_k, \lambda_k + \alpha \Delta \lambda_k)\} \tag{4.9}$$

where the merit function $m(x, \lambda)$ is defined as follows:

$$m(x, \lambda) = \frac{1}{2} \left(\|\nabla_x \mathcal{L}(x, \lambda)\|^2 + |h(x) - \varepsilon|^2 \right) - \gamma \mathcal{L}(x, \lambda) \tag{4.10}$$

for some $\gamma > 0$.
The stopping rule for the iterative procedure (4.7) is based on the discrepancy principle, i.e. the iteration (4.7) is stopped as soon as an iterate x_k satisfying the discrepancy principle (4.3) index $k > 1$ has been determined.
In our implementation we have computed an approximate minimizer α_k of the merit function $m(x_k + \alpha\Delta x_k, \lambda_k + \alpha\Delta\lambda_k)$; i.e. we have selected the step length α_k so that a "sufficient decrease" in the merit function has been produced along the direction $(\Delta x_k, \Delta\lambda_k)$. More specifically, we require α_k to be the first element of the sequence $\{2^{-i}\}_{i=0}^{\infty}$ satisfying the Armijo's condition:

$$m(x_k + \alpha_k\Delta x_k, \lambda_k + \alpha_k\Delta\lambda_k) \leq m(x_k, \lambda_k) + \mu\alpha_k \left(\Delta x_k, \Delta\lambda_k\right)^t \nabla_{(x,\lambda)} m(x_k, \lambda_k) \quad (4.11)$$

with $\mu = 10^{-4}$.
In our implementation, we have terminated the iterative procedure (4.7) not only according to the discrepancy principle:

$$\text{i)} \quad \|Ax_k - b\| \leq \rho\sigma \quad (4.12a)$$

but also when one of the following stopping criteria is satisfied:

$$\text{ii)} \quad k \geq k_{\max} \quad (4.12b)$$

$$\text{iii)} \quad \left\| \begin{pmatrix} \nabla_x \mathcal{L}(x_k, \lambda_k) \\ h(x_k) - \varepsilon \end{pmatrix} \right\| \leq \tau_1 \left\| \begin{pmatrix} \nabla_x \mathcal{L}(x_0, \lambda_0) \\ h(x_0) - \varepsilon \end{pmatrix} \right\| \quad (4.12c)$$

$$\text{iv)} \quad \left\| \alpha_k \begin{pmatrix} \Delta x_k \\ \Delta\lambda_k \end{pmatrix} \right\| \leq \tau_2 \quad (4.12d)$$

$$\text{v)} \quad |\alpha_k \Delta\lambda_k| \leq \tau_3 \quad (4.12e)$$

where $\rho > 1$ is fixed, $k_{\max}$ is the maximum number of allowed iterations and τ_1, τ_2, τ_3 are given tolerances. Conditions (4.12) represent the stopping criteria of the algorithm that can be stated as follows.

ALGORITHM 4.1 *First-Order Truncated Lagrangian Method*

0. Compute an initial iterate $(x_0, \lambda_0) \in \mathbb{R}^{n+1}$; set $k = 0$.

Repeat

1. Compute:

$$\Delta x_k = -\nabla_x \mathcal{L}(x_k, \lambda_k)$$
$$\Delta\lambda_k = h(x_k) - \varepsilon.$$

2. Find the first number α_k of the sequence $\{1, \frac{1}{2}, \frac{1}{4}, \ldots, \frac{1}{2^i}, \ldots\}$ such that:

$$m(x_k + \alpha\Delta x_k, \lambda_k + \alpha\Delta\lambda_k) \leq m(x_k, \lambda_k) + \mu\alpha_k \left(\Delta x_k, \Delta\lambda_k\right)^t \nabla_{(x,\lambda)} m(x_k, \lambda_k)$$

with $\mu = 10^{-4}$.

3. Set:

$$x_{k+1} = x_k + \alpha_k \Delta x_k,$$
$$\lambda_{k+1} = \lambda_k + \alpha_k \Delta \lambda_k.$$

4. Set $k = k + 1$ and return to Step **1**.

until stopping criteria (4.12) are satisfied.

4.2 Second-Order Truncated Lagrangian method

The second-order truncated Lagrangian method essentially consists in applying the second-order Lagrangian method of section 3.2 to the perturbed problem (4.4) and in stopping its iterations according to the discrepancy principle. That is, given the pair (x_k, λ_k), the new iterate is determined by:

$$\begin{aligned} x_{k+1} &= x_k + \alpha_k \Delta x_k \\ \lambda_{k+1} &= \lambda_k + \alpha_k \Delta \lambda_k \end{aligned} \tag{4.13}$$

where the search direction $(\Delta x_k, \Delta \lambda_k)^t \in \mathbb{R}^{n+1}$ is obtained by solving the system of equations:

$$\begin{pmatrix} \nabla^2_{xx}\mathcal{L}(x_k, \lambda_k) & \nabla_x h(x_k) \\ \nabla_x h(x_k)^t & 0 \end{pmatrix} \begin{pmatrix} \Delta x_k \\ \Delta \lambda_k \end{pmatrix} = - \begin{pmatrix} \nabla_x \mathcal{L}(x_k, \lambda_k) \\ h(x_k) - \varepsilon \end{pmatrix}. \tag{4.14}$$

The positive step-length parameter α_k is selected to satisfies conditions (3.14) where the merit function $m(x, \lambda)$ is defined as

$$m(x, \lambda) = \frac{1}{2}\Big(\|\nabla_x \mathcal{L}(x, \lambda)\|^2 + |h(x) - \varepsilon|^2 \Big). \tag{4.15}$$

According to the discrepancy principle, the iterative procedure (4.13) is terminated as soon as an iterate x_k satisfying condition (4.3) has been determined. Notice that the coefficient matrix of the linear system (4.14) is the same matrix $M(x_k, \lambda_k)$ defined by (3.12) which we have already proved to be nonsingular (proposition 3.2).
The algorithm can be now described as follows.

ALGORITHM 4.2 *Second-Order Truncated Lagrangian Method*

0. *Data.* Compute an initial iterate $(x_0, \lambda_0) \in \mathbb{R}^{n+1}$; set $k = 0$.

Repeat

1. Compute the matrix:

$$M(x_k, \lambda_k) = \begin{pmatrix} \nabla^2_{xx}\mathcal{L}(x_k, \lambda_k) & \nabla_x h(x_k) \\ \nabla_x h(x_k)^t & 0 \end{pmatrix}. \tag{4.16}$$

2. *Computation of the search direction.*
Compute the search direction $(\Delta x_k, \Delta\lambda_k)$ by solving the linear system:

$$M(x_k, \lambda_k) \begin{pmatrix} \Delta x_k \\ \Delta\lambda_k \end{pmatrix} = - \begin{pmatrix} \nabla_x \mathcal{L}(x_k, \lambda_k) \\ h(x_k) - \varepsilon \end{pmatrix}. \tag{4.17}$$

3. *Line search.*
Find the first number α_k of the sequence $\{1, \frac{1}{2}, \frac{1}{4}, \ldots, \frac{1}{2^i}, \ldots\}$ satisfying:

$$\text{i)} \quad m(x_k + \alpha_k \Delta x_k, \lambda_k + \alpha_k \Delta\lambda_k) \tag{4.18a}$$
$$\leq m(x_k, \lambda_k) + \mu\alpha_k \left(\Delta x_k, \Delta\lambda_k\right)^t \nabla_{(x,\lambda)} m(x_k, \lambda_k)$$
$$\text{with } \mu = 10^{-4};$$
$$\text{ii)} \quad \lambda_k + \alpha_k \Delta\lambda_k > 0; \tag{4.18b}$$
$$\text{iii)} \quad A(x_k + \alpha_k \Delta x_k) - b \neq 0. \tag{4.18c}$$

4. *Updates.*
Set:

$$\begin{aligned} x_{k+1} &= x_k + \alpha_k \Delta x_k, \\ \lambda_{k+1} &= \lambda_k + \alpha_k \Delta\lambda_k. \end{aligned} \tag{4.19}$$

Set $k = k + 1$ and return to Step **1**.

until $\|Ax_k - b\| \leq \rho\sigma$.

4.2.1 Implementation issues

In our implementation we have terminated the iterative procedure (4.13) according to the discrepancy principle, i.e. when

$$\textbf{i)} \quad \|Ax_k - b\| \leq \rho\sigma \tag{4.20a}$$

or when one of the following stopping criteria is satisfied:

$$\textbf{ii)} \quad k \geq k_{\max} \tag{4.20b}$$
$$\textbf{ii)} \quad \left\| \begin{pmatrix} \nabla_x \mathcal{L}(x_k, \lambda_k) \\ h(x_k) - \varepsilon \end{pmatrix} \right\| \leq \tau_1 \left\| \begin{pmatrix} \nabla_x \mathcal{L}(x_0, \lambda_0) \\ h(x_0) - \varepsilon \end{pmatrix} \right\| \tag{4.20c}$$
$$\textbf{iii)} \quad \left\| \alpha_k \begin{pmatrix} \Delta x_k \\ \Delta\lambda_k \end{pmatrix} \right\| \leq \tau_2 \tag{4.20d}$$

where $\rho > 1$ is fixed, $k_{\max}$ is the maximum number of allowed iterations and τ_1, τ_2 are given tolerances.
In our implementation, the linear system (4.17) is solved by the GMRES method. The GMRES iterations are terminated when the relative residual of (4.17) is smaller then a tolerance τ_{GMRES} or after a maximum of $\text{maxit}_{\text{GMRES}}$ iterations.
In the following, we will refer to the second-order Truncated Lagrangian method simply as Truncated Lagrangian method.
In what follows this chapter we will discuss some properties of the proposed Truncated Lagrangian methods.

4.3 Properties of the Truncated Lagrangian method

For easier presentation we will consider only the second-order Truncated method; the same results can be applied to the first-order method.
The following assumption is made.

Assumption A7 The perturbation term ε is such that

$$0 < \sqrt{2\varepsilon} \ll \rho\sigma \tag{4.21}$$

where σ is the error norm and $\rho > 1$ is fixed.

The following proposition ensures that Algorithm 4.2 with the stopping rule (4.20a) terminates after a finite number k_* of iterations.

Proposition 4.1. *Assume that Assumption A1–A7 hold and that $\|b\| > \rho\sigma$ with $\rho > 1$ fixed. Furthermore, assume that the initial iterate x_0 of the Truncated Lagrangian iteration* (4.13) *is $x_0 = 0$. Then, there exists a finite stopping index $k_* \in \mathbb{R}$ such that*

$$\|Ax_{k_*} - b\| \leq \rho\sigma \tag{4.22}$$

Proof. Because of our assumptions, we have

$$\|Ax_0 - b\| = \|b\| > \rho\sigma \tag{4.23}$$

and, by virtue of the theorem 3.2, we have that

$$\lim_{k\to\infty} \|Ax_k - b\| = \sqrt{2\varepsilon} < \rho\sigma. \tag{4.24}$$

Therefore, there exists a finite index $k_* \in \mathbb{N}$ such that

$$\|Ax_k - b\| \leq \rho\sigma, \quad \forall k \geq k_*. \tag{4.25}$$

□

Remark 4.1. Assumptions A1 and A6 ensure that $\|b\| > \sigma$. Then, for ρ sufficiently small, it is $\|b\| > \rho\sigma$. Therefore, the assumption of Proposition 4.1 is not restrictive.

Remark 4.2. Proposition 4.1 is still true under the assumption that $x_0 \neq 0$ and $\|Ax_0 - b\| > \rho\sigma$.

The following proposition characterizes the approximated solution to (1.1) and the corresponding regularization parameter computed by Truncated Lagrangian method when the regularization functional $\phi(x)$ is the Tikhonov functional (1.5).

Proposition 4.2. *Assume $\phi(x) = \frac{1}{2}\|Lx\|^2$, $L \in \mathbb{R}^{p\times n}$, $p \leq n$, be the Tikhonov regularization functional. Let x_{k_*} be the approximated solution to* (1.1) *computed by the Truncated Lagrangian method with stopping rule* (4.20a) *and let λ_{k_*} be the corresponding regularization parameter. Then the pair (x_{k_*}, λ_{k_*}) is a Kuhn–Tucker pair for the equality constrained minimization problem:*

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}\|Lx - x_c\|^2 \\ &\text{subject to} \quad \frac{1}{2}\|Ax - b\|^2 = \frac{1}{2}\varrho\sigma^2 \end{aligned} \tag{4.26}$$

where ϱ is a positive parameter and $x_c \in \mathbb{R}^m$ is a vector depending on k_*.

Proof. By the definition of the stopping index k_*, the residual associated to x_{k_*} satisfies:

$$\|Ax_{k_*} - b\| \leq \rho\sigma \tag{4.27}$$

with $\rho > 1$ fixed; in particular, there exists a positive $\varsigma \in]0, 1[$ such that

$$\|Ax_{k_*} - b\| = \varsigma\rho\sigma \tag{4.28}$$

and hence

$$\frac{1}{2}\|Ax_{k_*} - b\|^2 = \frac{1}{2}(\varsigma\rho)^2\sigma^2. \tag{4.29}$$

Choosing $\varrho = (\varsigma\rho)^2$ we have

$$\frac{1}{2}\|Ax_{k_*} - b\|^2 = \frac{1}{2}\varrho\sigma^2. \tag{4.30}$$

The pair (x_{k_*}, λ_{k_*}) does not exactly satisfy the first equation of (4.5); i.e.

$$\nabla_x \mathcal{L}(x_{k_*}, \lambda_{k_*}) = \omega_* \tag{4.31}$$

where $\omega_* \in \mathbb{R}^n$ is a residual vector. Therefore, by substituting the expression

$$\nabla_x \mathcal{L}(x_{k_*}, \lambda_{k_*}) = (L^tL + \lambda_{k_*}A^tA)x_{k_*} \tag{4.32}$$

in the equation (4.31) we have that the pair (x_{k_*}, λ_{k_*}) satisfies the equations:

$$\begin{aligned} &L^tLx_{k_*} - \omega_* + \lambda_{k_*}(A^tAx_{k_*} - A^tb) = 0, \\ &\frac{1}{2}\|Ax_{k_*} - b\|^2 = \frac{1}{2}\varrho\sigma^2. \end{aligned} \tag{4.33}$$

Equations (4.33) are the first-order necessary optimality conditions for the equality constrained problem:

$$\begin{aligned} &\text{minimize} \quad x^tL^tLx - x^t\omega_* \\ &\text{subject to} \quad \frac{1}{2}\|Ax - b\|^2 = \frac{1}{2}\varrho\sigma^2 \end{aligned} \tag{4.34}$$

which is equivalent to the following equality constrained problem:

$$\begin{aligned} &\text{minimize} \quad \|Lx - x_c\|^2 \\ &\text{subject to} \quad \frac{1}{2}\|Ax - b\|^2 = \frac{1}{2}\varrho\sigma^2 \end{aligned} \tag{4.35}$$

where $x_c = \frac{1}{2}(L^t)^\dagger\omega_*$ and $(L^t)^\dagger$ is the Moore–Penrose generalized inverse of L^t. □

Chapter 5

Implementation details

In this chapter we describe some aspects of the implementation of the algorithms of the previous chapters 3 and 4 for Tikhonov and Total Variation functionals. For easier presentation, throughout this section we omit the index k.
Let us define the function $\widetilde{h}(x)$ such that:

$$\widetilde{h}(x) = \begin{cases} h(x) - \dfrac{\sigma^2}{2}, & \text{for the Lagrangian Method;} \\ \\ h(x) - \varepsilon, & \text{for the Truncated Lagrangian Method.} \end{cases} \tag{5.1}$$

In both the Lagrangian method (Chapter 3.2) and the Truncated Lagrangian method (Chapter 4.2) we have to solve a linear system

$$M(x,\lambda)\begin{pmatrix}\Delta x \\ \Delta\lambda\end{pmatrix} = -\begin{pmatrix}\nabla_x \mathcal{L}(x,\lambda) \\ \widetilde{h}(x)\end{pmatrix} \tag{5.2}$$

where the coefficient matrix is defined as

$$M(x,\lambda) = \begin{pmatrix}\nabla^2_{xx}\mathcal{L}(x,\lambda) & \nabla_x h(x) \\ \\ \nabla_x h(x)^t & 0\end{pmatrix} \tag{5.3}$$

and

$$\mathcal{L}(x,\lambda) = \phi(x) + \lambda\widetilde{h}(x).$$

5.1 Discrete formulation of the continuous Total Variation functional

In Total Variation regularization, the regularization functional ϕ is the discretization of the continuous Total Variation functional Φ defined as:

$$\Phi(\chi) = \int_\Omega |\nabla\chi(t)|dt, \tag{5.4}$$

where $\chi : \Omega \to \mathbb{R}$ and Ω is a bounded convex region in the d-dimensional space. In this section, we give a description of the discretization of $\Phi(\chi)$ by finite

difference. This discretization is quite standard for Total Variation functional; the interested reader can find a broader discussion in [13, 12, 50].
Let us consider the two-dimensional case, i.e. $\chi(t) = \chi(u, v)$ is a function of $\Omega \subset \mathbb{R}^2 \longrightarrow \mathbb{R}$. For simplicity let $\Omega = (0, 1) \times (0, 1)$. Assume to discretize the function $\chi(u, v)$ on an equispaced grid of $n_u \times n_v$ points (u_i, v_j) such that:

$$\begin{aligned} u_i &= i\Delta u, \quad i = 1, \dots, n_u, \\ v_j &= j\Delta v, \quad j = 1, \dots, n_v, \end{aligned} \tag{5.5}$$

and let x_{ij} denote the value of χ at the grid point (u_i, v_j), i.e.:

$$x_{i,j} = \chi(u_i, u_j), \quad i = 1, \dots, n_u, \, j = 1, \dots, n_v. \tag{5.6}$$

For notational simplicity, let us assume $n_u = n_v$ and let we denote

$$m = n_u = n_v, \quad n = m^2, \quad h = \frac{1}{m}. \tag{5.7}$$

The gradient of the matrix $x \in \mathbb{R}^{m \times m}$ at the (i, j)-th entry is defined by means of the forward difference operator as:

$$\nabla x_{ij} = \left(\frac{x_{(i+1)j} - x_{ij}}{h}, \frac{x_{i(j+1)} - x_{ij}}{h} \right), \quad i, j = 1, \dots, m. \tag{5.8}$$

The forward differential operator is modified according to Neumann boundary conditions for boundary pixels. We then obtain the following discretization of the continuous functional (5.4):

$$\phi(x) = h^2 \sum_{i,j=1}^{m} |\nabla x_{ij}|. \tag{5.9}$$

To overcome the difficulties due to the non-differentiability of the Euclidean norm, the functional $\phi(x)$ is usually replaced by the slightly modified functional

$$\phi_\beta(x) = h^2 \sum_{i,j=1}^{m} |\nabla x_{i,j}|_\beta, \tag{5.10}$$

where

$$|\nabla x_{i,j}|_\beta = \sqrt{|\nabla x_{i,j}|^2 + \beta^2}. \tag{5.11}$$

To apply the methods of the previous chapters 3 and 4, we need to compute an expression for the gradient $\nabla_x \phi_\beta(x)$ and the hessian $\nabla_{xx} \phi_\beta(x)$.
The gradient of $\phi_\beta(x)$ is the matrix $\nabla_x \phi_\beta(x) \in \mathbb{R}^{m \times m}$ whose (i, j)-th entry is defined as:

$$\left(\nabla_x \phi_\beta(x) \right)_{i,j} = \nabla \cdot \left(\frac{\nabla x_{i,j}}{\sqrt{|\nabla x_{i,j}|^2 + \beta^2}} \right). \tag{5.12}$$

The Lagrangian methods of chapters 3 and 4, at each iteration, require to solve a linear system (5.2) where the matrix x is assumed to be ordered row-wise in a vector also termed x. We compute the solution of (5.2) by the GMRES method that requires the computation of the action of the coefficient matrix $M(x, \lambda)$ on a vector $(p, q)^t \in \mathbb{R}^{n+1}$. We have (see (5.3)):

$$M(x, \lambda) \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} \nabla^2_{xx} \mathcal{L}(x, \lambda) p + \nabla_x h(x) q \\ \nabla_x h(x)^t p \end{pmatrix} \tag{5.13}$$

where

$$\nabla^2_{xx}\mathcal{L}(x,\lambda)p = \nabla^2_{xx}\phi_\beta(x)p + \lambda A^t Ap. \tag{5.14}$$

Thus, instead of determining a formula for the hessian $\nabla^2_{xx}\phi_\beta(x)$ of the Total Variation functional, it is sufficient to provide a formula for the computation of the matrix-vector product $\nabla^2_{xx}\phi_\beta(x)p$. Given a matrix $p \in \mathbb{R}^{m\times m}$, the discrete hessian $\nabla^2_{xx}\phi_\beta(x)$ acting on p is the $m \times m$ matrix whose (i,j)-th entry is given by:

$$\left(\nabla^2_{xx}\phi_\beta(x)p\right)_{i,j} = \nabla\cdot\left(\frac{\nabla p_{i,j}}{\sqrt{|\nabla x_{i,j}|^2+\beta^2}}\right) - \nabla\cdot\left(\frac{\nabla p_{i,j}(\nabla x_{i,j})^t}{\sqrt{|\nabla_{i,j}x|^2+\beta^2}}\nabla x_{i,j}\right). \tag{5.15}$$

The previously defined matrix (5.15) ordered row-wise can be substituted in expression (5.14) to evaluate the product $\nabla^2_{xx}\mathcal{L}(x,\lambda)p$.
Vogel and Oman [51] proposed to approximate the true hessian $\nabla^2_{xx}\phi_\beta(x)$ with the matrix obtained by dropping in (5.15) the higher order term. Therefore the action of the approximated hessian on a matrix p is defined as:

$$\left(\nabla^2_{xx}\phi_{\beta,\text{app}}(x)p\right)_{i,j} = \nabla\cdot\left(\frac{\nabla p_{i,j}}{\sqrt{|\nabla x_{i,j}|^2+\beta^2}}\right) \tag{5.16}$$

where $\nabla^2_{xx}\phi_{\beta,\text{app}}(x)$ denotes the approximated hessian. In our numerical experiments we have used both the true (5.15) and the approximated (5.16) hessian of the discrete Total Variation functional.

5.2 Computation of the matrix-vector product Ay = z when A is a BTTB matrix

Let us consider the case of A a Block Toeplitz with Toeplitz Blocks (BTTB) matrix. Such a matrix A derives, for example, from the discretization of the first kind integral equation:

$$\int_\Omega \mathcal{A}(s-t)\chi(t)dt + \eta(x) = \zeta(s). \tag{5.17}$$

This integral equation is used to model the formation process of an image $\chi(t) : \mathbb{R}^2 \to \mathbb{R}$ degraded by a gaussian spatially invariant blurring operator $\mathcal{A} : \mathbb{R}^2 \to \mathbb{R}^2$ and by gaussian noise $\eta : \mathbb{R}^2 \to \mathbb{R}$. The function $\mathcal{A}$ is termed as Point Spread Function (PSF); the function $\zeta(t) : \mathbb{R}^2 \to \mathbb{R}$ represents the measured image. The spatial coordinates $t, s \in \Omega$, with Ω a compact subset of $\mathbb{R}^2$. Assume the blurred image $\zeta(s)$ to be measured at $m\times m$ points; then, by discretizing the continuous model (5.17), we obtain the matrix equation model (1.1) where $n = m^2$ and the vectors $x, b \in \mathbb{R}^n$ represent the components of the true and the observed image, respectively, in a row-wise ordering. The coefficient matrix is BTTB; even if A is a full matrix, it is not necessary to store all its entries and the Fast Fourier Transform (FFT) can be used to efficiently evaluate the matrix-vector products involving A.
Let S be a $m \times m$ image of the PSF partitioned into four blocks S_{ij}:

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \tag{5.18}$$

where $S_{22}(1,1)$ is the centre of the PSF. Let $S_e \in \mathbb{R}^{2m\times 2m}$ be the extension of S defined as:

$$S_e = \begin{pmatrix} S_{22} & 0 & S_{21} \\ 0 & 0 & 0 \\ S_{12} & 0 & S_{11} \end{pmatrix}. \tag{5.19}$$

where the null matrix $0 \in \mathbb{R}^{m\times m}$.
Let $Y \in \mathbb{R}^{m\times m}$ partitioned into four blocks:

$$Y = \begin{pmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{pmatrix} \tag{5.20}$$

and let $Y_e \in \mathbb{R}^{m\times m}$ be the extended array such that:

$$Y_e = \begin{pmatrix} Y_{22} & 0 & Y_{21} \\ 0 & 0 & 0 \\ Y_{12} & 0 & Y_{11} \end{pmatrix}. \tag{5.21}$$

Let $\mathsf{extract}$ be the operator extracting from a $2m \times 2m$ matrix a $m \times m$ submatrix made by the elements in the first m rows and columns. Moreover let $\mathsf{reshape}$ represent the operator converting a $n = m^2$ vector in a $m \times m$ array via filling it by columns and let $\mathsf{fft2}$ and $\mathsf{ifft2}$ denote the two-dimensional Discrete FFT (2DFFT) and the two-dimensional Discrete Inverse FFT (2DIFFT), respectively. Then, the product $z := Ay$ with $y = \mathsf{reshape}(Y, n, 1)$, is obtained as follows:

$$z = \mathsf{reshape}\Big(\mathsf{extract}\Big(\mathsf{ifft2}\big(\mathsf{fft2}(S_e).*\mathsf{fft2}(Y_e)\big)\Big), n, 1\Big) \tag{5.22}$$

where $.*$ is the component-wise product. More details on the computation of the products involving A can be found in [30].
Therefore, even if the size of linear systems (5.2) arising in image restoration is large, the matrix vector products can be performed via FFTs at a reduced computational cost and iterative methods, such as the GMRES method, are effectively usable for solving such linear systems.

5.3 Solution of the Newton linear system when A is a BTTB matrix and $\phi(x) = \|I\|^2$

Let us consider the linear system (5.2). We have already observed that the hessian matrix $\nabla^2_{xx}\mathcal{L}(x,\lambda)$ is nonsingular. If it is easily invertible, i.e. its inverse $\big(\nabla^2_{xx}\mathcal{L}(x,\lambda)\big)^{-1}$ can be easily computed, then the solution of the linear system (5.2) can be obtained by the following explicit formula (§14.1 of [36]):

$$\begin{aligned} \Delta x &= -\Big(\nabla^2_{xx}\mathcal{L}\Big)^{-1}\Big(\nabla_x h\Delta\lambda + \nabla_x\mathcal{L}\Big), \\ \Delta\lambda &= \Big((\nabla_x h)^t\nabla^2_{xx}\mathcal{L}\nabla_x h\Big)^{-1}\Big((\nabla_x h)^t(\nabla^2_{xx}\mathcal{L})^{-1}\nabla_x\mathcal{L} - \widetilde{h}\Big). \end{aligned} \tag{5.23}$$

Notice that, since $\nabla_x h$ is a vector of dimension n, then the inversion of the term $(\nabla_x h)^t\nabla^2_{xx}\mathcal{L}\nabla_x h$ is simply the inversion of a scalar. In order to be able to effectively use formula (5.23), we need an expression for the inverse of the

hessian $\nabla^2_{xx}\mathcal{L}(x,\lambda)$.
Assume that A is a BTTB matrix and the regularization functional is the Tikhonov functional (1.5) with L equal to the identity I. Then:

$$\nabla^2_{xx}\mathcal{L}(x,\lambda) = I + \lambda A^t A. \tag{5.24}$$

Observe that, even if A is BTTB, A^tA could not be BTTB. It is possible to approximate A^tA by a BTTB matrix T as described in [43, 44]. Following the notation of the previous section, the matrix-vector products involving T can be computed as follows. Let y be a n-dimensional vector and let

$$\widehat{T} = |\mathsf{fft2}(S_e)|.\widehat{}\,2 \tag{5.25}$$

where $|\cdot|$ and $.\widehat{}$ are the component-wise absolute value and squaring of the $2m \times 2m$ matrix S_e. Then:

$$Ty = \mathsf{reshape}\Big(\mathsf{extract}\Big(\mathsf{ifft2}\big(\widehat{T}. * \mathsf{fft2}(Y_e)\big)\Big), n, 1\Big) \tag{5.26}$$

where $Y = \mathsf{reshape}(y, m, m)$. By substituting T in (5.24), we obtain the following approximation of the hessian of the Lagrangian

$$\nabla^2_{xx}\mathcal{L}_{\text{app}}(x,\lambda) = I + \lambda T. \tag{5.27}$$

The approximated hessian $\nabla^2_{xx}\mathcal{L}_{\text{app}}(x,\lambda)$ can be explicitly inverted by use of FTTs. By replacing the true hessian (5.24) with the approximated one (5.27), we obtain the Modified Lagrangian method and the Modified Truncated Lagrangian methods where the search direction is computed by solving the modified linear system

$$\begin{pmatrix} \nabla^2_{xx}\mathcal{L}_{\text{app}}(x,\lambda) & \nabla_x h(x) \\ \nabla_x h(x)^t & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} = - \begin{pmatrix} \nabla_x \mathcal{L}(x,\lambda) \\ \widetilde{h}(x) \end{pmatrix} \tag{5.28}$$

via the explicit formula:

$$\begin{aligned} \Delta x &= -\Big(\nabla^2_{xx}\mathcal{L}_{\text{app}}\Big)^{-1}\Big(\nabla_x h \Delta\lambda + \nabla_x \mathcal{L}\Big), \\ \Delta\lambda &= \Big((\nabla_x h)^t \nabla^2_{xx}\mathcal{L}_{\text{app}} \nabla_x h\Big)^{-1}\Big((\nabla_x h)^t (\nabla^2_{xx}\mathcal{L})^{-1}\nabla_x\mathcal{L} - \widetilde{h}\Big). \end{aligned} \tag{5.29}$$

Chapter 6

Numerical Results

In the following, for easier notation, we generically refer to all the methods described in chapters 3 and 4 as *Lagrangian-type methods*.
In this section we report the results obtained from several numerical experiments performed in order to evaluate the performance of the Lagrangian-type methods. In our experiments we have used as regularization functional the Tikhonov functional and the Total Variation functional. We present three sets of experiments. The first set (section 6.1) consists of one-dimensional numerical experiments. The second and the third sets consist of two-dimensional experiments: an application to image deblurring (section 6.2) and some applications to image denoising (section 6.3). The numerical experiments have been executed on a Pentium IV PC using Matlab 6.5. In each numerical experiment we have selected, as initial iterate (x_0, λ_0) of the considered Lagrangian-type methods, the vector:

$$(x_0, \lambda_0) = (\mathbf{0}, 1) \in \mathbb{R}^{n+1} \tag{6.1}$$

where $\mathbf{0}$ is the null vector of $\mathbb{R}^n$. In all the experiments, we have fixed the following value for the parameter ρ of the discrepancy principle (criterium (4.20a)):

$$\rho = 1 + \text{eps} \tag{6.2}$$

where eps is the machine precision.
The values of the tolerances for the stopping criteria are different for each set of experiments and are indicated in the following.

6.1 One-dimensional test problems

Aim of the one-dimensional experiments is to compare the Lagrangian methods (chapter 3) and the Truncated Lagrangian methods (chapter 4) with some well-known regularization methods.
To test the proposed methods when applied to Tikhonov regularization (1.5), we have considered some one-dimensional test problems from the Regularization Tools package by P. C. Hansen [33, 32]. This package includes a collection of test problems which are widely used in the literature.
The considered test problems from [32] are the following: phillips, baart, shaw and heat. They all come from the discretization of a Fredholm equation of the

first kind:

$$\int_a^b K(s,t)f(t)dt = g(s); \tag{6.3}$$

the problem is to compute a regularized solution of the linear system (1.1) where $A \in \mathbb{R}^{100\times 100}$ comes from the discretization of the kernel $K(s,t)$ on a grid of 100 points. The right-hand side vector $b \in \mathbb{R}^{100}$ is given by:

$$b = \widetilde{b} + \eta \tag{6.4}$$

where $\widetilde{b} \in \mathbb{R}^{100}$ derives from the discretization of $g(s)$ and $\eta \in \mathbb{R}^{100}$ is a white noise vector with noise level:

$$\delta := \frac{\|\eta\|}{\|\widetilde{b}\|} = 10^{-3} \tag{6.5}$$

for the baart, shaw and heat test problems and $\delta = 10^{-2}$ for the phillips test problem. For these test problems the exact solution $\widetilde{x}$ is known and we can compute the relative error of the approximated solution. In our experiments, we have used the identity matrix ($L = I$) or the discretization of the second order derivative operator ($L = \nabla^2$) as regularization matrix L. For a comparison with the proposed approach, we have considered the solution obtained by Tikhonov method with the regularization parameter λ determined by the GCV method, the L-curve method and heuristically. Besides, we have considered the Conjugate Gradient for Least Squares (CGLS) method [6] with iterations stopped according to the discrepancy principle for a comparison of the Lagrangian-type methods with a well established iterative regularization method [19, 33, 28]. For this set of test problems, the selected values of the tolerances of the stopping criteria (3.8), (3.18), (4.12) and (4.20) of the Lagrangian-type methods are the following:

$$\begin{aligned} &k_{\max} = 50, \quad \tau_1 = 10^{-4}, \quad \tau_2 = 10^{-8}, \quad \tau_3 = 10^{-2}, \\ &w = 10^{10}, \quad \text{maxit}_{\text{GMRES}} = 150, \quad \tau_{\text{GMRES}} = 10^{-6}. \end{aligned}$$

In tables 7.1, 7.2, 7.3 and 7.4 we have presented the results obtained for the considered test problems. In these tables we have reported the value of the regularization parameter λ and the relative error of the corresponding solution x_λ obtained by Tikhonov method with GCV criterium (Tikh + GCV), L-curve criterium (Tikh + L-curve) and with an heuristic criterium for the choice of the regularization parameter (Tikh + heuristic). In the heuristic criterium, several experiments have been performed in order to determine a good regularization parameter. For such computation we have used the codes from [32]. For the CGLS method, we have reported the relative error of the computed solution and the number of performed iterations. For the second-order and the first-order Lagrangian and Truncated Lagrangian methods (L, 1-stL, TL and 1-stTL) we have indicated the computed value of λ, the relative error of x_λ and the number of inner and outer iterations. In particular, the first-order methods have been used only to compute an approximation of the regularization parameter: the parameter λ computed by the first-order methods is substituted in the Tikhonov unconstrained problem

$$\underset{x}{\text{minimize}} \quad \|Lx\|^2 + \frac{\lambda}{2}\|Ax - b\|^2 \tag{6.6}$$

to calculate a regularized solution x_λ of the original linear system (1.1). Therefore, in the last two rows of the tables 7.1, 7.2, 7.3 and 7.4, the value of λ is the value computed by the first-order methods, while the relative error is the relative error of the solution x_λ of the Tikhonov unconstrained problem (6.6). In the last column of the tables, we have indicated the type of matrix L used in each experiment. The solutions obtained by all the considered methods are plotted in figures 7.1, 7.2, 7.3 and 7.4 (continuous line) together with the exact solutions (dotted line).
From the numerical results shown in the tables and from the plots in the figures, we have observed that the Lagrangian method and the Truncated Lagrangian method always give results of very good quality. The Truncated Lagrangian method has a less computational cost compared with the traditional Lagrangian method (except for the baart test problem) and it has a smaller relative error for the phillips and baart test problems. Obviously, Tikhonov regularization with heuristically determined regularization parameter always provides solution of very good quality. On the other hand, the GCV and L-curve criteria are not always reliable: in some cases they are effective in computing a good value of λ but in other cases they are unsuccessful. Instead, in each test problem the Lagrangian-type methods provide at the same time a good approximation of λ and a regularized solution whose relative error is comparable with the relative errors of the Tikhonov solution with heuristic criterium. The first-order Lagrangian and Truncated Lagrangian methods are not always successful in computing a good regularization parameter but, if it is the case, they require a really little amount of computational work.
We remark that even if for three test problems (phillips, shaw and heat) the CGLS method gives a very good solution with little computational cost, for the test problem baart it gives a very poor approximation to the true solution. It is evident that, for such test problem, in order to compute a good solution, it is necessary to add to the original problem (1.1) the information deriving from the regularization functional $\|Lx\|^2$ with $L = \nabla^2$. The term $\|Lx\|^2$ gives additional information on the regularity of the solution $\widetilde{x}$ and this information can not be recovered by iterative regularization methods such as the CGLS method. On the other hand, for the test problem shaw, Tikhonov regularization does not reproduce very well the true solution: even if the approximated solutions by Tikhonov method (Tikh + GCV, Tikh + L-curve and Tikh + heuristic) have a smaller relative error, they are not visually satisfactory. Instead, the Lagrangian methods compute a good approximate solution for all the considered test problems. From these considerations, we can state that the Truncated Lagrangian method is an *iterative regularization method with a Tikhonov-like prior information term*. The terminate iteration number k and the regularization parameter λ are selected according to the discrepancy principle. In some sense therefore, the proposed truncated method joins the properties of the classic Tikhonov regularization with those of the regularization by iteration.
For the Total Variation functional, we have tested the proposed approach on two test problems of function denoising, i.e. A in (1.1) is the identity matrix. Figure 7.5 shows the noise-free functions (figures 7.5(a) and 7.5(b)) and the noisy ones (figures 7.5(c) and 7.5(d)) for the considered two test problems that we have called box and towers, respectively. For this set of test problems, the selected values of the tolerances of the stopping criteria (3.8), (3.18), (4.12) and

(4.20) of the Lagrangian-type methods are the following:

$$k_{\max} = 50, \quad \tau_1 = 10^{-1}, \quad \tau_2 = 10^{-8}, \quad \tau_3 = 10^{-2},$$
$$w = 1, \quad \text{maxit}_{\text{GMRES}} = 100, \quad \tau_{\text{GMRES}} = 10^{-2}.$$

The level of the noise is $\delta = 0.75$ for the test problem box and $\delta = 0.5$ for the test problem towers. For a numerical comparison with the proposed approach, we have considered the Fixed Point Iteration method of Vogel [51, 52, 50, 15] and the Newton method [51]. Both the cited methods compute a Total Variation regularized solution of (1.1) by solving the first-order Euler-Lagrange equations of the unconstrained minimization problem (1.7), therefore they require a suitable value of the regularization parameter. In tables 7.5 and 7.6 we have compared the numerical results for the test problems box and towers obtained with the following methods: Fixed Point (FP) and Newton (N) methods, Lagrangian and Truncated Lagrangian methods with exact hessian (5.15) (L, TL) and with approximated hessian (5.16) (L (app. hess.), TL (app. hess.)) and first-order Lagrangian and Truncated Lagrangian methods (1-stL, 1-stTL). In these tables we have reported the relative error of the computed solution x_λ, the value of the corresponding regularization parameter λ and the number of outer and inner iterations. For the Fixed Point method and the Newton method, the value of λ has been heuristically determined and the used inner solver is the Conjugate Gradient method. As in the previous test problem, the first-order Lagrangian and Truncated Lagrangian methods have been used only to compute an estimate of the regularization parameter λ. The corresponding regularized solution x_λ has been determined by applying the Fixed Point method the unconstrained problem (1.7) with λ approximated by the first-order methods. Therefore, for the first-order Lagrangian methods, the number of outer iterations refers to the number of iterations of the first-order Lagrangian method for providing an estimate of λ. The relative error is the error of the solution computed by Fixed Point method with the computed approximation of λ. In the column of the inner iterations, we have indicated the number of iterations of the inner solver of the Fixed Point method and, between braces, the number of outer iterations. Figures 7.6 and 7.7 show the computed solutions (continuous line) and the exact solution (dotted line). From the figures we observe that the considered methods provide reconstructions of different quality. In fact, the Fixed Point and Newton reconstructions greatly depend on the value of the regularization parameter and for the box test problem the obtained reconstructions are not visually satisfactory, even if they have a smaller relative error. On the other hand, the solutions obtained by Lagrangian-type methods are always satisfactory and have relative errors comparable to the relative errors of the solutions given by Fixed Point and Newton methods. Besides, from the tables we observe that Lagrangian-type methods require more computational work but they provide at the same time both a suitable value of the regularization parameter λ and the regularized solution x_λ.

We conclude this section dedicated to the one-dimensional test problems with a summarizing comment. The experiments have been performed with the purpose of evaluating the behavior of the Lagrangian-type methods and compare them with other well-established methods. For all the considered one-dimensional test problems, the Lagrangian and Truncated Lagrangian methods give results of actually good quality that are always comparable with those of the methods

known in literature. Their computational cost is reasonable considered that, given an estimate of the variance σ, they provide both the regularization parameter λ and the approximate solution x_λ.

6.2 Image deblurring test problem

In this section, we present the results on an image deblurring test problem consisting in restoring the image of a satellite in space degraded by Gaussian blur and noise. The image to be restored is a computer simulation of a field experiment showing a satellite as taken from a ground based telescope. The set of data of this test problem was developed at the US Air Force Phillips Laboratory, Laser and Imaging Directorate, Kirtland US Air Force Base, New Mexico [45, 40], and has been widely used in the literature to test image deblurring algorithms. The data set is available from:
`http://www.mathcs.emory.edu/~nagy/RestoreTools/index.html` [39]; it consists of the true and blurred satellite images (figures 7.8(a) and 7.8(b)) of 256×256 pixels. The blurring matrix A is an ill-conditioned $256^2 \times 256^2$ BTTB matrix and it is implicitly defined by a Point Spread Function. Therefore, the Fast Fourier Transform can be used for matrix vector multiplications with A (see section 5.2). For large size problem, such as image restoration problems where the image to be restored is usually represented by 256×256 pixels, the L-curve and the GCV criteria are practically unusable for determining an estimate of λ because of their extremely high computational cost. Therefore, in large size applications, λ is usually chosen heuristically on the basis of preliminary experiments and experience. Our goal is to show that, even for large size applications, Lagrangian-type methods can be actually used for automatic selection of the regularization parameter. In this experiment, we have tested and compared the Lagrangian method and the Truncated Lagrangian method. As regularization functional, we have considered both Tikhonov functional (with L equal to the identity matrix) and Total Variation functional with exact (5.15) and approximated (5.16) hessian. For this test problem, the selected values of the tolerances of the stopping criteria (3.18) and (4.20) of the Lagrangian-type methods are the following:

$$k_{\max} = 50, \quad \tau_1 = 10^{-3}, \quad \tau_2 = 10^{-8},$$
$$w = 1, \quad \text{maxit}_{\text{GMRES}} = 30, \quad \tau_{\text{GMRES}} = 10^{-2},$$

for both Tikhonov and Total Variation functional. For Tikhonov functional, we have considered two different implementations of the aforementioned methods. The first one is the "standard" implementation described in section 3.2 while the second one is the implementation of the Modified Lagrangian method based on the approximation of the product A^tA by a circulant matrix as described in section 5.3 (in this case $w = 10^{14}$). In figure 7.8 the reconstructed images are displayed. Figures 7.8(c), 7.8(e) and 7.8(g) show the images obtained with the Lagrangian method; figures 7.8(d), 7.8(f) and 7.8(h) depict the reconstructions given by the Truncated Lagrangian method. Table 7.7 reports the value of the computed λ, the relative error of x_λ and the number of inner and outer iterations for the considered methods.
From the displayed images and the numerical results reported in the table, it is

evident that Lagrangian-type methods behave very well on this test problem; they gives satisfactory reconstructions at a quite low computational cost: only 37 matrix-vector products are required. Notice that the reconstructions by Tikhonov and Total Variation functionals are surprisingly very similar (figures 7.8(c)- 7.8(d) and 7.8(g)- 7.8(h)).

6.3 Image denoising test problems

In this section we present the results obtained on a set of image denoising test problems. Aim of this set of test problems is to compare the behaviour of the Lagrangian and the Truncated Lagrangian methods when applied to image denoising. For this set of test problems, we have considered the Tikhonov ($L = \nabla^2$) and Total Variation regularization functionals.
For Tikhonov functional, the selected values of the tolerances of the stopping criteria (3.18) and (4.20) of the Lagrangian-type methods are the following:

$$k_{\max} = 50, \quad \tau_1 = 10^{-1}, \quad \tau_2 = 10^{-8},$$
$$\text{maxit}_{\text{GMRES}} = 50, \quad \tau_{\text{GMRES}} = 10^{-2}.$$

For Total Variation functional, the tolerances of the stopping criteria (3.18) are:

$$k_{\max} = 50, \quad \tau_1 = 10^{-1}, \quad \tau_2 = 10^{-8},$$
$$\text{maxit}_{\text{GMRES}} = 50, \quad \tau_{\text{GMRES}} = 10^{-2},$$

while the tolerances of the stopping criteria (4.20) are:

$$k_{\max} = 50, \quad \tau_1 = 10^{-2}, \quad \tau_2 = 10^{-8},$$
$$\text{maxit}_{\text{GMRES}} = 50, \quad \tau_{\text{GMRES}} = 10^{-4}.$$

In the first test problem, the exact satellite image of section 6.2 is used. Figure 7.9 depicts the exact, the noisy ($\delta = 0.5$) and the reconstructed images.
In the second test problem, the well-known Lena image is exposed with noise ($\delta = 0.09$). Figure 7.10 shows the original and noisy images together with the obtained reconstructions. In the next test problem, a real Magnetic Resonance (MR) image (figure 7.11(a)) is exposed with noise of level $\delta = 0.2$ (figure 7.11(b)). The reconstructions are displayed in figure 7.11. The last test problem concerns a blocky image degraded by noise ($\delta = 0.12$). This image is available at the url: `http://www.math.uib.no/~mariusl/`. The original, the noisy and the reconstructed images are shown in figure 7.12. Tables 7.8, 7.9, 7.10 and 7.11 show the relative error of the approximated solution x_λ, the computed value of the regularization parameter λ and the number of the performed iterations for these four test problems. As expected, the Total Variation functional is more suited for image denoising than Tikhonov functional. In fact, from the figures and the relative error values shown in the tables, it is evident that the Total Variation functional performs better than Tikhonov functional in removing noise from the images; the images obtained by Tikhonov regularization have a higher relative error and exhibit the presence of noise. Besides, the images obtained with exact and approximated hessian of the Total Variation functional are comparable in quality and computational cost.

As far as concerns the performance of the proposed Lagrangian-type methods, we have noticed that for this set of test problems the Lagrangian method and the new Truncated Lagrangian method perform differently.
For the satellite image, the Truncated Lagrangian method does not completely filter out the noise but it conserves all the details of the image.
The Lena and the MR images have many regions with gradual image variation and therefore they are not well suited for Total Variation regularization. As expected, the Total Variation functional transforms smooth regions into piecewise constant regions but the so called staircase effect is more evident in the images reconstructed by the Lagrangian method. Tikhonov functional does not completely remove the noise and tends to blur the images, however the images reconstructed by the Truncated Lagrangian method are less blurred than those obtained by the Lagrangian method.
Finally, for the blocky image, the images reconstructed by the two methods are visually very similar but the reconstructions given by the Truncated Lagrangian method have a smaller relative error (for the Total Variation functional).
From the reported results, it is evident that the proposed Lagrangian-type methods are efficient and effective when applied to image denoising test problems. On the whole, the considered methods have a contained computational cost: on average, they require no more than 150 matrix-vector products to provide a good denoised image and the corresponding regularization parameter. In particular, the new Truncated Lagrangian method filters out the noise very well and, at the same time, the new method is able to retain the fine details of the images as well as to preserve its smooth regions. It is well-known that both the Total Variation and the Tikhonov functionals have some drawbacks: the first one suffers from a staircase effect in regions with gradual image variation, while the second one blurs the details of the images. The results show that such defects are less evident in the images reconstructed by the Truncated Lagrangian method: the images are less flat or less blurred if obtained by the Truncated Lagrangian method rather than the Lagrangian method. Furthermore, the computational cost of the Truncated Lagrangian method is considerably smaller than those of the classic Lagrangian method. Therefore, the proposed Truncated Lagrangian method appears to be a promising technique for image denoising.

Chapter 7

Conclusions

In this work, we have considered the noise constrained formulation of the problem of the regularization of discrete ill-posed linear systems. We have analyzed the noise constrained regularization problem from a theoretical point of view and we have characterized its solution. We have analyzed the behavior of the Lagrangian method when applied to the noise constrained regularization problem. Moreover, we have introduced a novel regularization technique based on truncated iterations of the Lagrangian method. The numerous performed experiments have shown that the considered methods are effective and efficient in computing both a regularized solution of the linear system and the corresponding regularization parameter.
The analyzed Lagrangian-type methods require the error norm to be known. In our future work we intend to investigate the use of the L-curve criterium for an a-posteriori selection of the terminate index of the Truncated Lagrangian iterations.
Finally, we remark that the Lagrangian-type methods do not require neither the objective functional $\phi(x)$ nor the constraint $h(x)$ to be linear. Therefore, the proposed regularization approach based on iterations of Lagrangian-type methods is actually attractive for nonlinear regularization.

Method	Rel. Error	λ	outer it.	inner it.	L
Tikh + GCV	4.6218e-002	6.9079e+001	-	-	I
Tikh + L-curve	1.0841e-001	2.3731e+002	-	-	I
Tikh + heuristic	2.8913e-002	2.0000e+001	-	-	I
CGLS	2.6626e-002	-	5	-	-
L	2.9282e-002	2.2757e+001	13	144	I
TL	2.8318e-002	2.3920e+001	9	91	I
1st-L	3.1604e-002	8.7457e+000	4	0	I
1st-TL	3.1603e-002	8.7467e+000	4	0	I

Table 7.1: Test problem: phillips

Method	Rel. Error	λ	outer it.	inner it.	L
Tikh GCV	3.6452e-002	4.2810e-001	-	-	∇^2
Tikh + L-curve	3.5122e-002	3.1871e-001	-	-	∇^2
Tikh + heuristic	3.3225e-002	1.0000e-001	-	-	∇^2
CGLS	1.6602e-001	-	3	-	-
L	5.7800e-002	7.8523e-003	7	695	∇^2
TL	3.2199e-002	7.9387e-002	14	1402	∇^2
1st-L	4.6671e-002	1.6604e+000	5	-	∇^2
1st-TL	4.6671e-002	1.6604e+000	5	-	∇^2

Table 7.2: Test problem: baart

Method	Rel. Error	λ	outer it.	inner it.	L
Tikh + GCV	3.8349e-002	5.0772e+004	-	-	I
Tikh + L-curve	4.1591e-002	1.2939e+005	-	-	I
Tikh + heuristic	3.7763e-002	6.6667e+004	-	-	I
CGLS	4.7917e-002	-	8	-	-
L	4.3629e-002	1.3208e+004	22	182	I
TL	4.5340e-002	7.0868e+003	18	140	I
1st-L	1.6595e-001	3.5276e+001	8	-	I
1st-TL	1.6595e-001	3.5276e+001	8	-	I

Table 7.3: Test problem: shaw

Method	Rel. Error	λ	outer it.	inner it.	L
Tikh + GCV	5.1613e-002	1.6777e+006	-	-	I
Tikh + L-curve	2.5826e-001	2.8474e+007	-	-	I
Tikh + heuristic	4.0180e-002	6.6667e+005	-	-	I
CGLS	4.5846e-002	-	21	-	-
L	5.6880e-002	1.6934e+005	21	446	I
TL	7.0751e-002	9.8693e+004	18	341	I
1st-L	9.6490e-001	1.1095e+000	2	-	I
1st-TL	9.6490e-001	1.1095e+000	2	-	I

Table 7.4: Test problem: heat

Method	Rel. Error	λ	outer it.	inner it.
FP	8.1876e-002	1.0000e+002	36	973
N	8.3591e-002	1.0000e+002	86	5153
L	9.2849e-002	9.9545e+001	23	2300
TL	8.3654e-002	9.7977e+001	26	2563
L (app. hess.)	7.8005e-002	9.9368e+001	22	2200
TL (app. hess.)	8.3070e-002	9.7596e+001	12	1200
1st-L	8.2771e-002	1.0000e+002	41	769 (24 outer it.)
1st-TL	8.2754e-002	1.0001e+002	42	764 (24 outer it.)

Table 7.5: Test problem: box

Method	Rel. Error	λ	outer it.	inner it.
FP	1.1656e-001	1.0000e+002	29	815
N	1.1392e-001	1.0000e+002	74	4081
L	1.1855e-001	9.9359e+001	22	2165
TL	1.1850e-001	9.8638e+001	24	2400
L (app. hess.)	1.2221e-001	9.9128e+001	25	2500
TL (app. hess.)	1.2154e-001	9.7574e+001	30	3000
1st-L	1.1610e-001	1.0000e+002	50	658 (23 outer it.)
1st-TL	1.1568e-001	1.0001e+002	50	660 (23 outer it.)

Table 7.6: Tets problem: towers

Method	Rel. Error	λ	outer it.	inner it.	reg.
L	3.6764e-001	6.5536e+004	2	37	Tikh
TL	3.6764e-001	6.5536e+004	2	37	Tikh
L (fft)	4.1395e-001	8.0689e+002	13	-	Tikh
TL (fft)	3.7907e-001	1.8922e+003	9	-	Tikh
L	3.6730e-001	6.5536e+004	2	37	TV
TL	3.6730e-001	6.5536e+004	2	37	TV
L (app. hess.)	3.6730e-001	6.5536e+004	2	37	TV
TL (app. hess.)	3.6730e-001	6.5536e+004	2	37	TV

Table 7.7: Test problem: deblurring a satellite image

Method	Rel. Error	λ	outer it.	inner it.	reg.
L	2.3898e-001	6.6269e-001	6	99	Tikh
TL	2.6711e-001	9.4191e-001	2	19	Tikh
L	1.6666e-001	4.3674e-002	11	147	TV
TL	2.6879e-001	4.5496e-001	2	19	TV
L (app. hess.)	1.6903e-001	4.4778e-002	10	121	TV
TL (app. hess.)	2.4949e-001	4.8883e-001	2	19	TV

Table 7.8: Test problem: denoising a satellite image

Method	Rel. Error	λ	outer it.	inner it.	reg.
L	6.5752e-002	3.1413e+000	8	98	Tikh
TL	6.6226e-002	6.2345e+000	8	97	Tikh
L	5.3317e-002	9.5027e-002	12	118	TV
TL	5.3088e-002	1.2105e-001	4	48	TV
L (app. hess.)	5.3160e-002	9.4114e-002	8	55	TV
TL (app. hess.)	5.3450e-002	1.2368e-001	4	48	TV

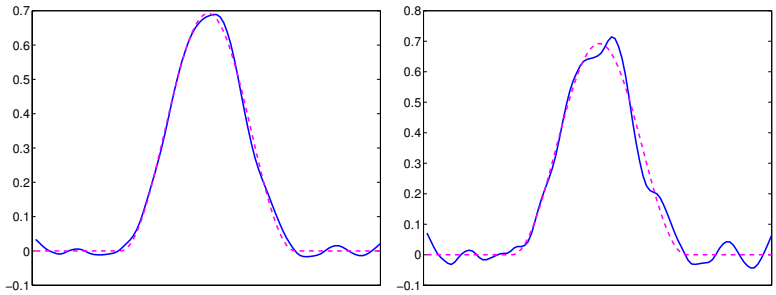
Table 7.9: Test problem: denoising the Lena image

Method	Rel. Error	λ	outer it.	inner it.	reg.
L	1.3437e-001	2.2992e+000	8	120	Tikh
TL	1.2767e-001	6.2838e+000	6	76	Tikh
L	1.2179e-001	1.0308e-001	12	139	TV
TL	1.2232e-001	2.4368e-001	3	33	TV
L (app. hess.)	1.2181e-001	1.0255e-001	7	56	TV
TL (app. hess.)	1.2024e-001	2.5827e-001	3	33	TV

Table 7.10: Test problem: denoising a MR image

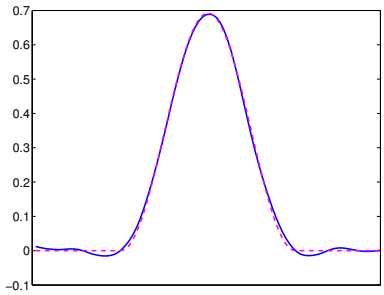
Method	Rel. Error	λ	outer it.	inner it.	reg.
L	1.1170e-001	9.0030e+000	8	85	Tikh
TL	1.1534e-001	1.2004e+001	7	77	Tikh
L	8.1718e-002	1.4433e-001	11	95	TV
TL	7.0494e-002	2.2867e-001	4	53	TV
L (app. hess.)	8.1966e-002	1.4526e-001	8	57	TV
TL (app. hess.)	7.2142e-002	2.5129e-001	4	52	TV

Table 7.11: Test problem: denoising a blocky image

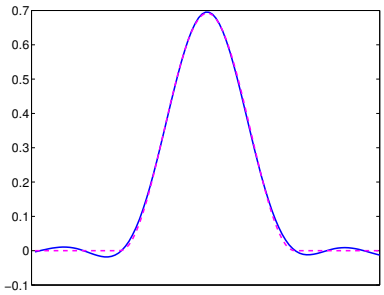


(a) Tikhonov with GCV

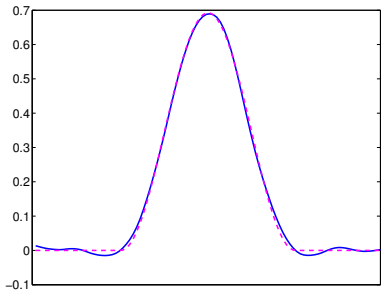
(b) Tikhonov with L-curve



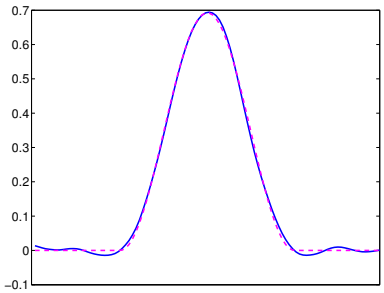
(c) "Heurisytic" Tikhonov



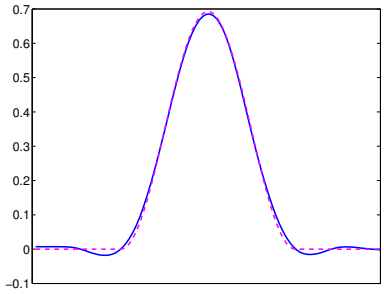
(d) CGLS



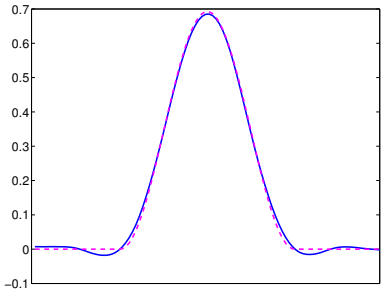
(e) Lagrangian method



(f) Truncated Lagrangian method

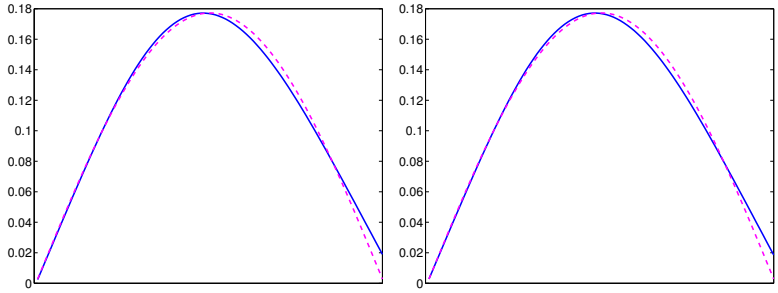


(g) First-order Lagrangian method



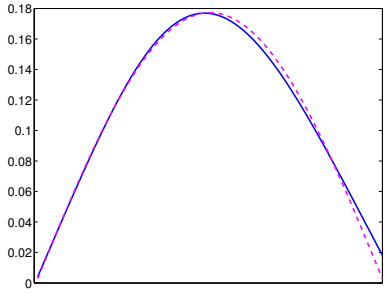
(h) First-order Truncated Lagrangian method

Figure 7.1: Test problem: phillips. Computed solutions (continuous line) and exact solution (dotted line).

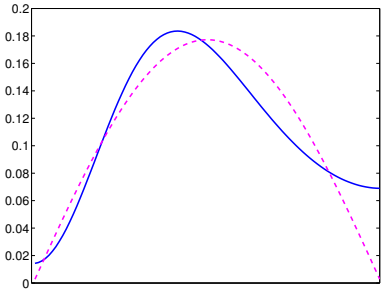


(a) Tikhonov with GCV

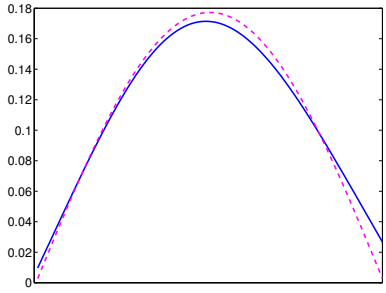
(b) Tikhonov with L-curve



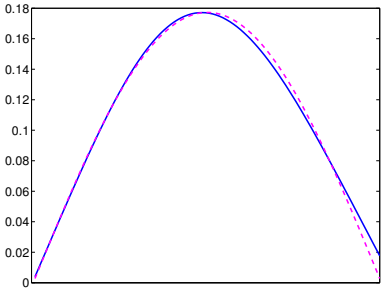
(c) "Heurisytic" Tikhonov



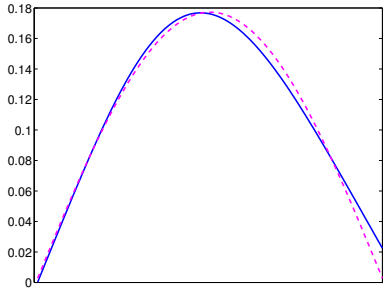
(d) CGLS



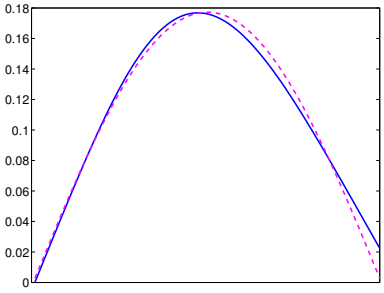
(e) Lagrangian method



(f) Truncated Lagrangian method

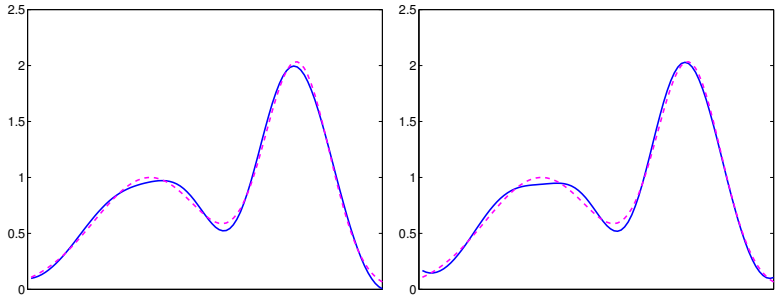


(g) First-order Lagrangian method



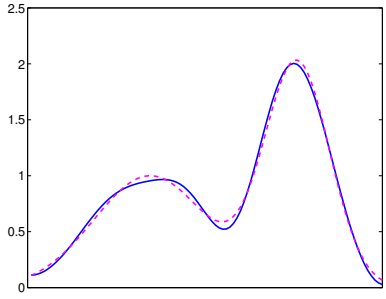
(h) First-order Truncated Lagrangian method

Figure 7.2: Test problem: baart. Computed solutions (continuous line) and exact solution (dotted line).

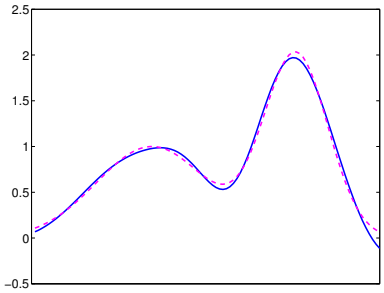


(a) Tikhonov with GCV

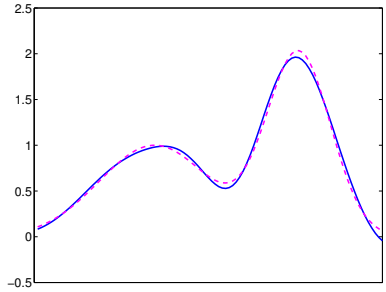
(b) Tikhonov with L-curve



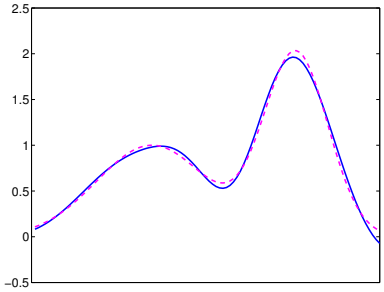
(c) "Heurisytic" Tikhonov



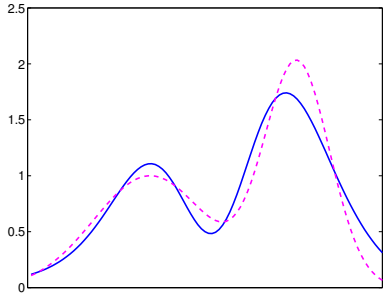
(d) CGLS



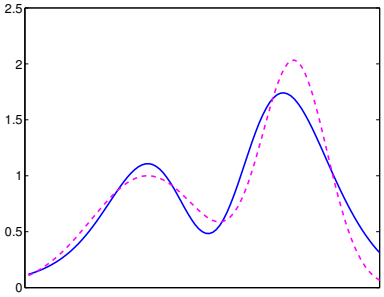
(e) Lagrangian method



(f) Truncated Lagrangian method

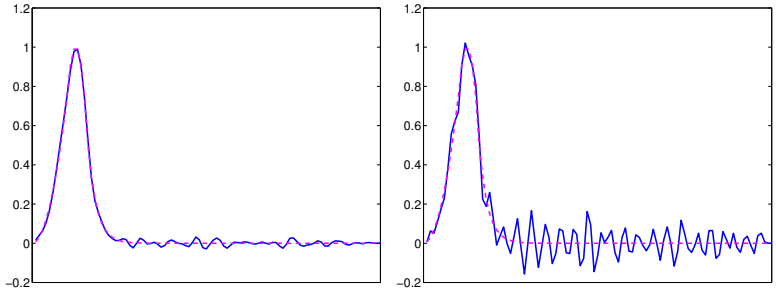


(g) First-order Lagrangian method



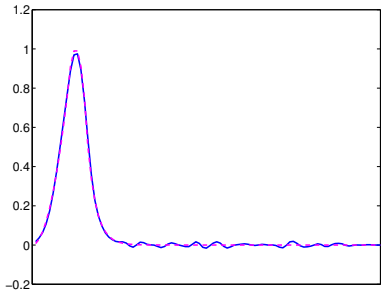
(h) First-order Truncated Lagrangian method

Figure 7.3: Test problem: shaw. Computed solutions (continuous line) and exact solution (dotted line).

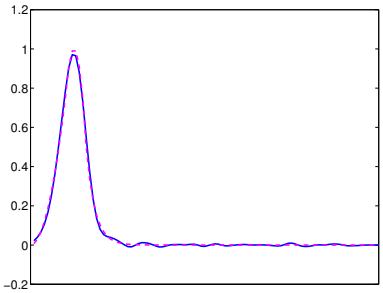


(a) Tikhonov with GCV

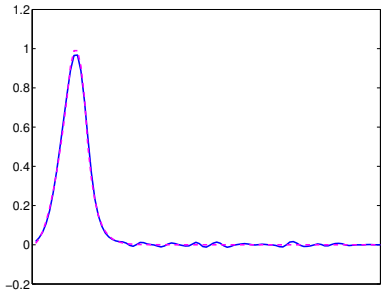
(b) Tikhonov with L-curve



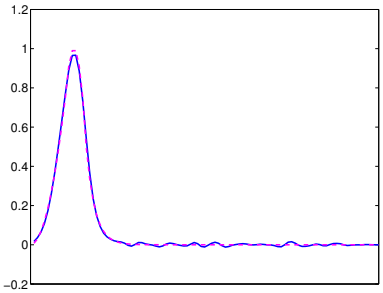
(c) "Heurisytic" Tikhonov



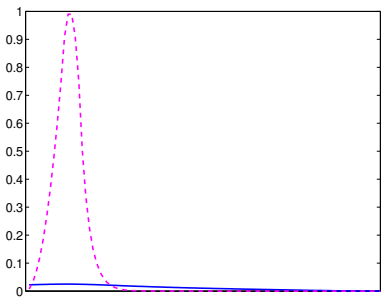
(d) CGLS



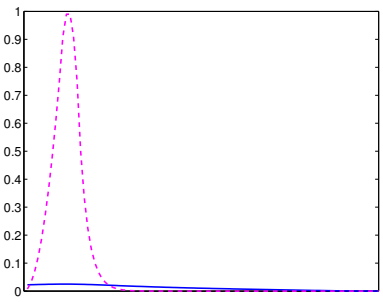
(e) Lagrangian method



(f) Truncated Lagrangian method

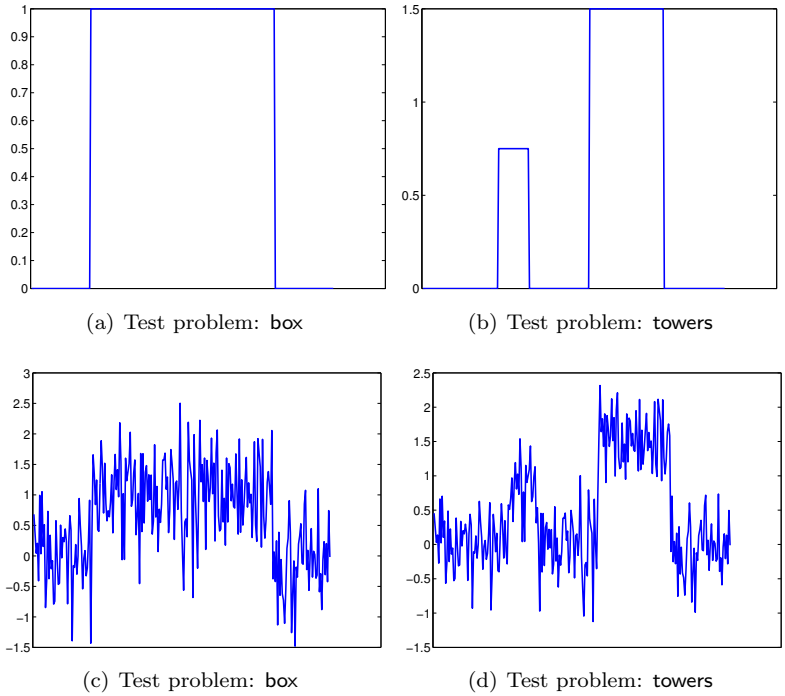


(g) First-order Lagrangian method



(h) First-order Truncated Lagrangian method

Figure 7.4: Test problem: heat. Computed solutions (continuous line) and exact solution (dotted line).



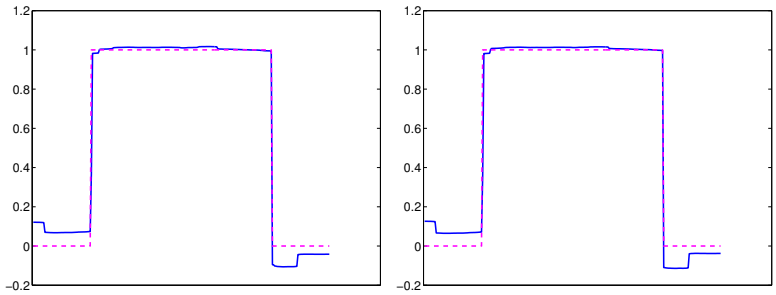
(a) Test problem: box

(b) Test problem: towers

(c) Test problem: box

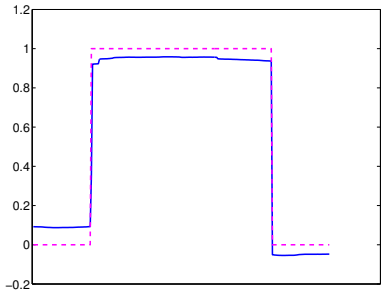
(d) Test problem: towers

Figure 7.5: Exact signals (top line) and noisy signals (bottom line).

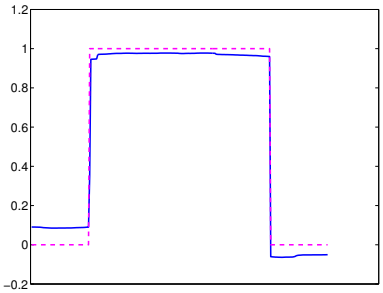


(a) Fixed Point Iteration

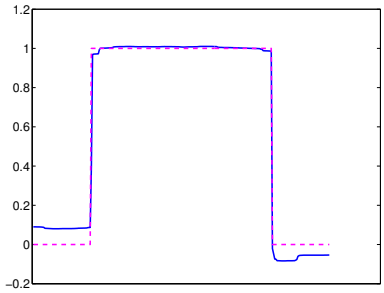
(b) Newton method



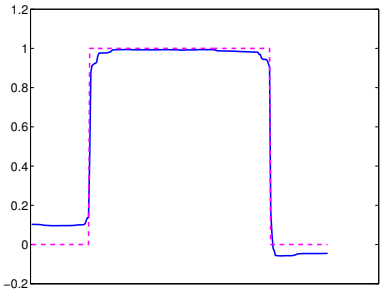
(c) Lagrangian method



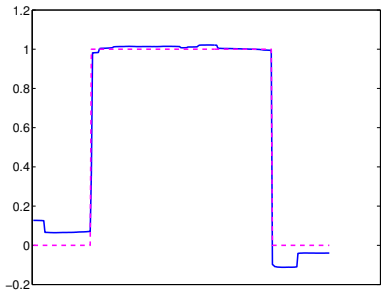
(d) Truncated Lagrangian method



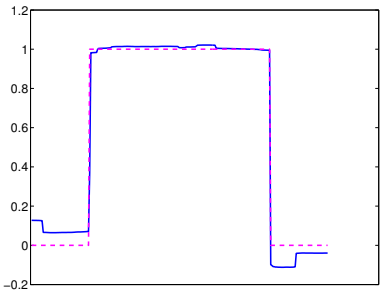
(e) Lagrangian method with approximated hessian



(f) Truncated Lagrangian method with approximated hessian

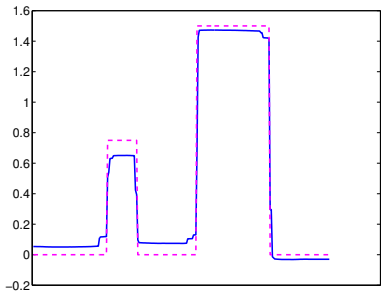


(g) First-order Lagrangian method

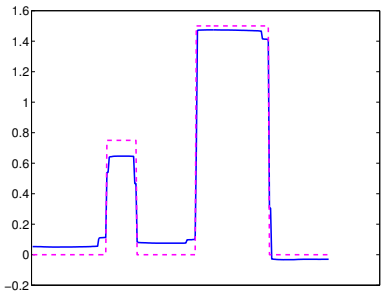


(h) First-order Truncated Lagrangian method

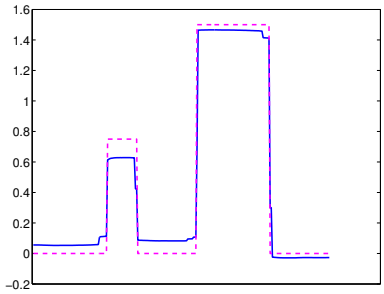
Figure 7.6: Test problem: box. Computed solutions (continuous line) and exact solution (dotted line).



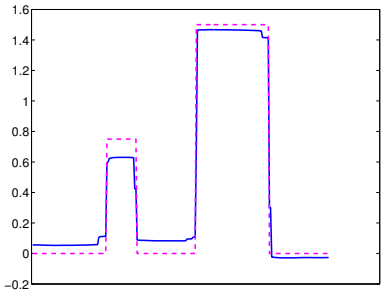
(a) Fixed Point Iteration



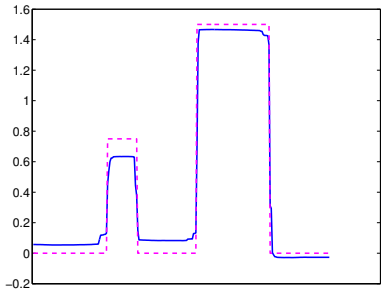
(b) Newton method



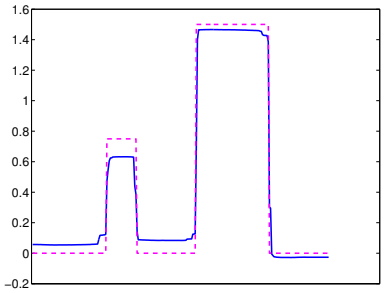
(c) Lagrangian method



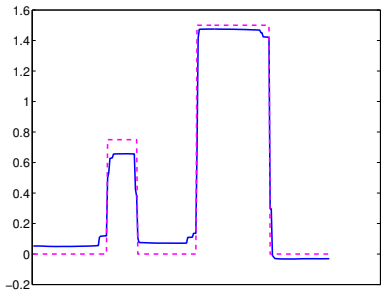
(d) Truncated Lagrangian method



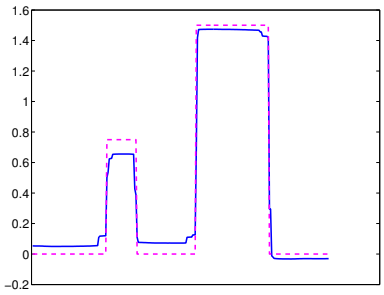
(e) Lagrangian method with approximated hessian



(f) Truncated Lagrangian method with approximated hessian

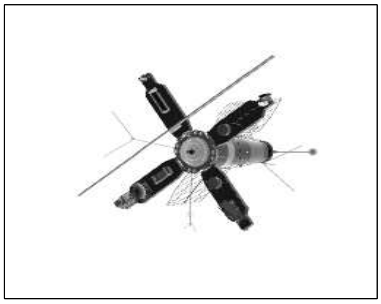


(g) First-order Lagrangian method

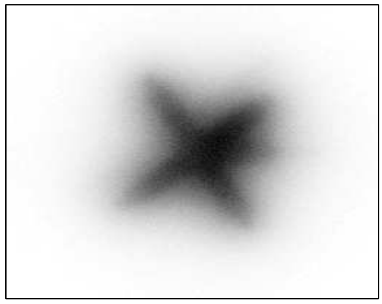


(h) First-order Truncated Lagrangian method

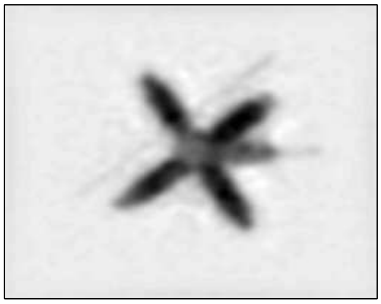
Figure 7.7: Test problem: towers. Computed solutions (continuous line) and exact solution (dotted line).



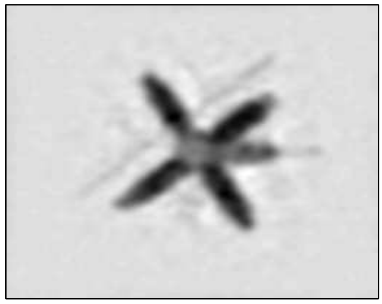
(a) Exact image



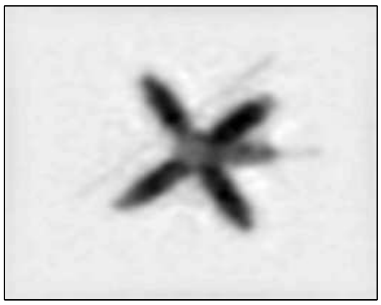
(b) Noisy and blurred image



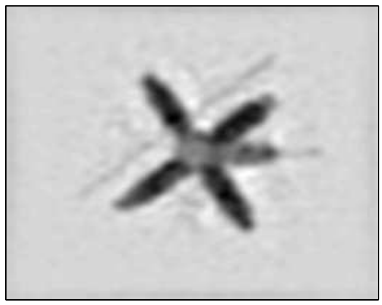
(c) Tikhonov regularization: Lagrangian method



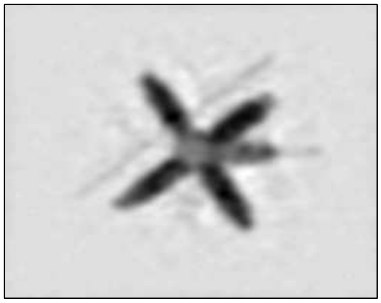
(d) Tikhonov regularization: Truncated Lagrangian method



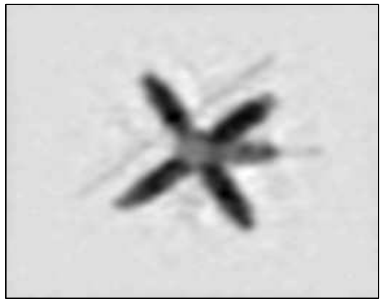
(e) Tikhonov regularization: Lagrangian method (circulant matrix approximation)



(f) Tikhonov regularization: Truncated Lagrangian method (circulant matrix approximation)

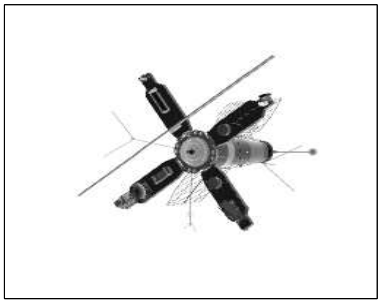


(g) Total Variation regularization: Lagrangian method

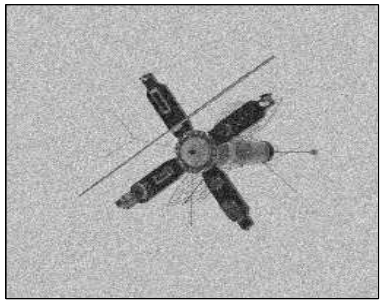


(h) Total Variation regularization: Truncated Lagrangian method

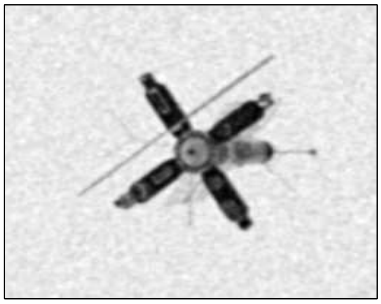
Figure 7.8: Test problem: deblurring a satellite image.



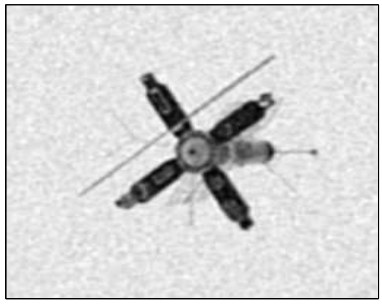
(a) Exact image



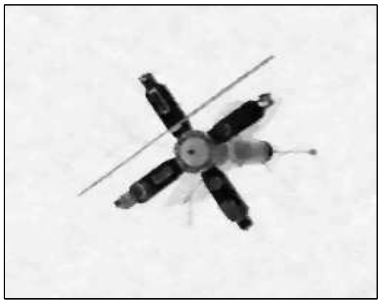
(b) Noisy image



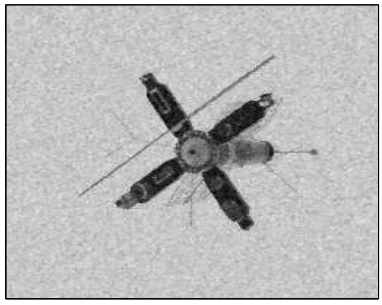
(c) Tikhonov regularization: Lagrangian method



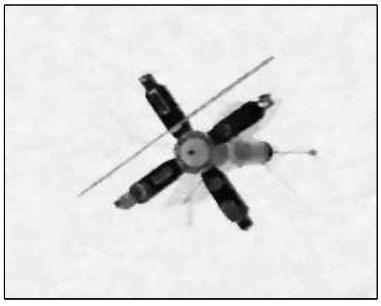
(d) Tikhonov regularization: Truncated Lagrangian method



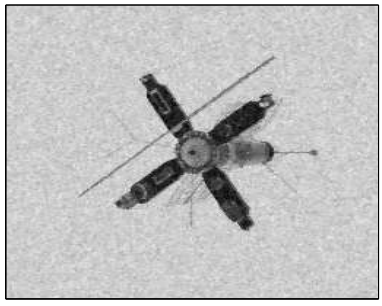
(e) Total Variation regularization: Lagrangian method



(f) Total Variation regularization: Truncated Lagrangian method



(g) Total Variation regularization: Lagrangian method (approximated hessian)



(h) Total Variation regularization: Truncated Lagrangian method (approximated hessian)

Figure 7.9: Test problem: denoising a satellite image.

(a) Exact image

(b) Noisy image

(c) Tikhonov regularization: Lagrangian method

(d) Tikhonov regularization: Truncated Lagrangian method

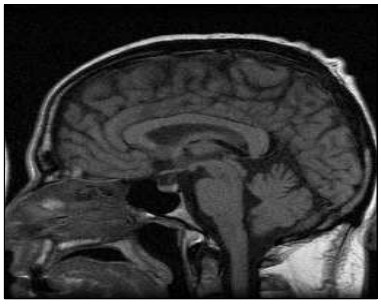
(e) Total Variation regularization: Lagrangian method

(f) Total Variation regularization: Truncated Lagrangian method

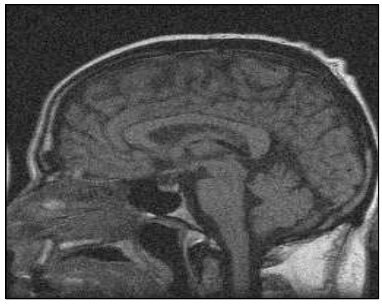
(g) Total Variation regularization: Lagrangian method (approximated hessian)

(h) Total Variation regularization: Truncated Lagrangian method (approximated hessian)

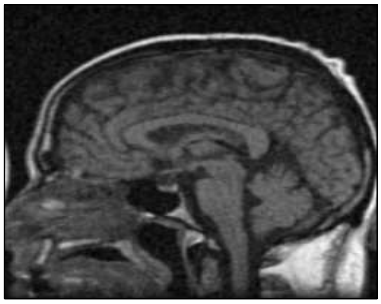
Figure 7.10: Test problem: denoising the Lena image.



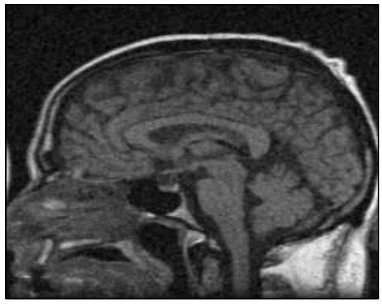
(a) Exact image



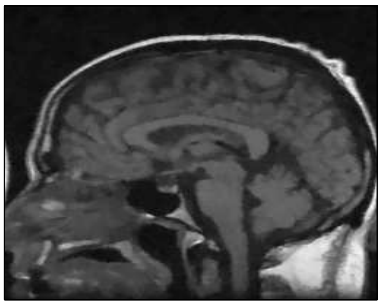
(b) Noisy image



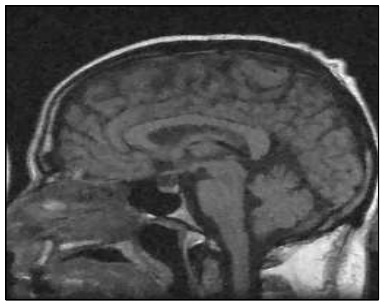
(c) Tikhonov regularization: Lagrangian method



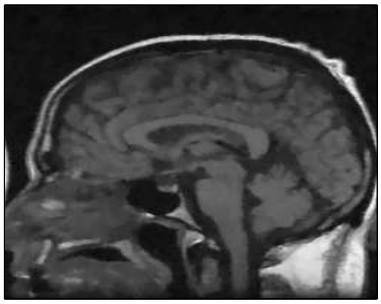
(d) Tikhonov regularization: Truncated Lagrangian method



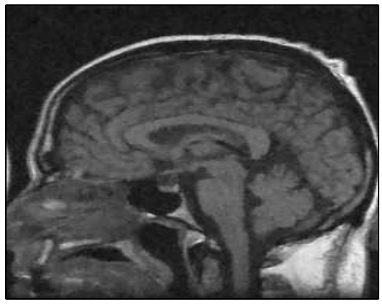
(e) Total Variation regularization: Lagrangian method



(f) Total Variation regularization: Truncated Lagrangian method

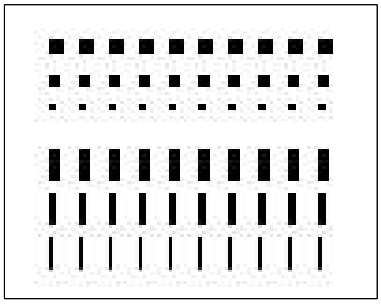


(g) Total Variation regularization: Lagrangian method (approximated hessian)

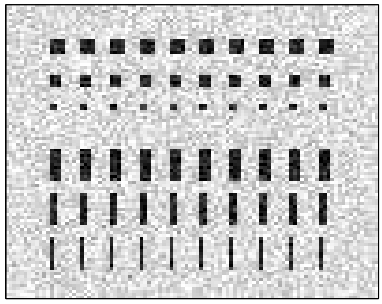


(h) Total Variation regularization: Truncated Lagrangian method (approximated hessian)

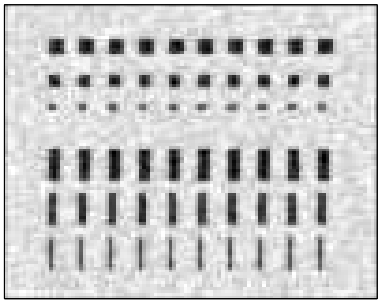
Figure 7.11: Test problem: denoising a MR image.



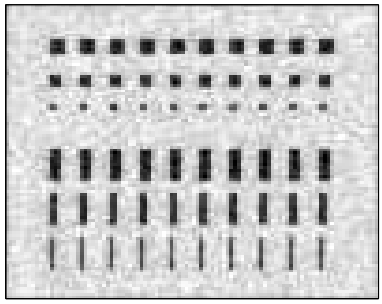
(a) Exact image



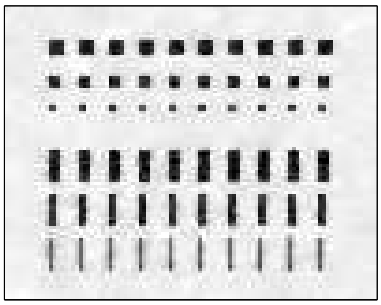
(b) Noisy image



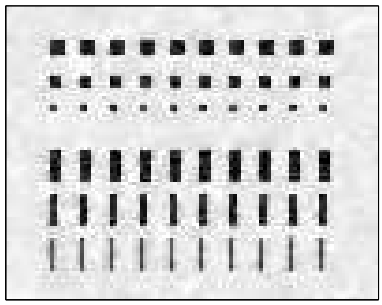
(c) Tikhonov regularization: Lagrangian method



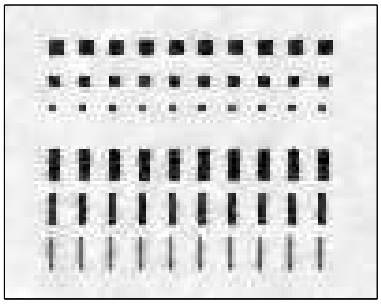
(d) Tikhonov regularization: Truncated Lagrangian method



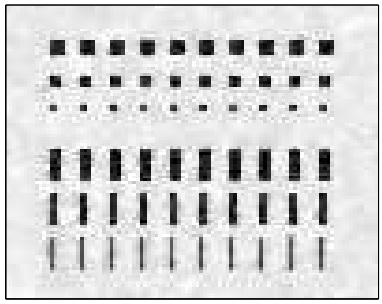
(e) Total variation regularization: Lagrangian method



(f) Total Variation regularization: Truncated Lagrangian method



(g) Total Variation regularization: Lagrangian method (approximated hessian)



(h) Total Variation regularization: Truncated Lagrangian method (approximated hessian)

Figure 7.12: Test problem: denoising a blocky image.

Bibliography

[1] R. Acar and C. R. Vogel. Analysis of total variation penalty methods. *Inv. Prob.*, 10:1217–1229, 1994.

[2] J. Baumeister. *Stable Solution of Inverse Problems.* Vieweg, Braunschweig, 1986.

[3] M. Bertero and P. Boccacci. *Introduction to Inverse Problems in Imaging.* IOP Publishing, Bristol, 1998.

[4] D. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods.* Academic Press, 1982.

[5] D. Bertsekas. *Nonlinear Programming.* Athena Scientific, (2nd Edition), 1999.

[6] A. Bjorck. *Numerical methods for least squares problems.* SIAM, Philadelphia, 1996.

[7] P. Blomgren and T. F. Chan. Modular solvers for image restoration problems using the discrepancy principle. *Numerical Linear Algebra with Applications*, 9(5):347–358, 2002.

[8] D. Calvetti, G. Golub, and L. Reichel. Estimation of the L-curve via Lanczos bidiagonalization. *BIT*, 39:603–619, 1999.

[9] D. Calvetti and L. Reichel. Tikhonov regularization of large linear problems. *BIT*, 43:263–283, 2003.

[10] D. Calvetti and L. Reichel. Tikhonov regularization with a solution constraint. *SIAM J. Sci. Comput.*, 26(1):224–239, 2004.

[11] T. Chan, J. Olkin, and D. Cooley. Solving quadratically constrained least squares using black box solvers. *BIT*, 32(3):481–495, 1992.

[12] T. F. Chan, R. H. Chan, and H. M. Zhou. Continuation method for total variation denoising problems. *Proceedings to the SPIE Symposium on Advanced Signal Processing: Algorithms, Architectures, and Implementations*, 2563:314–325, July 1995. Ed. F. Luk.

[13] T. F. Chan, G. H. Golub, and P. Mulet. A nonlinear primal-dual method for total variation-based image restoration. *SIAM J. Sci. Comput.*, 20(6):1964–1977, 1999.

[14] D. Dementiev and J. G. Nagy. Bound constrained regularization for ill-posed problems. *BIT*, 40(4):1–4, 2000.

[15] D. Dobson and C.R. Vogel. Convergence of an iterative method for total variation denoising. *SIAM J. Numer. Anal.*, 34:1779–1791, 1997.

[16] D. C. Dobson and F. Santosa. Recovery of blocky images from noisy and blurred data. *SIAM J. Appl. Math.*, 56:1181–1198, 1996.

[17] L. Eldén. Algorithms for the regularization of ill-conditioned least squares problems. *BIT*, 17:134–145, 1977.

[18] H. W. Engl. Regularization methods for the stable solution of inverse problems. *Surveys on Mathematics for Industry*, 3:71–143, 1993.

[19] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*. Kluwer Academic Publishers, Dordrecht, 1996.

[20] H. Gfrer. An *a posteriori* parameter choice for ordinary and iterated tikhonov regularization of ill-posed problems leading to optimal convergence rates. *Math. Comput.*, 49:507–522, 1987.

[21] G. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21:215–223, 1979.

[22] G. H. Golub and U. von Matt. Quadratically constrained least squares and quadratic problems. *Numer. Math.*, 59:561–580, 1991.

[23] C. W. Groetsch. *The theory of Tikhonov regularization for Fredholm equations of the first kind*. Pitman, Boston, 1983.

[24] C. W. Groetsch. *Inverse Problems in the Mathematical Sciences*. Vieweg, Bruanschweig, 1993.

[25] S. Gutman. Identification of discontinuous parameters in flow equations. *SIAM J. Control Optim.*, 28:1049–1060, 1990.

[26] J. Hadamard. Lectures on Cauchy's problem in linear partial differential equations. *New Haven, CT: Yale University Press*.

[27] J. Hadamard. *Bull. Univ. Princeton*, 13(49), 1902.

[28] M. Hanke. *Conjugate Gradient Type Methods for Ill-Posed Problems*. Pitman Research Notes in Mathematics, Longman House, Harlow, Essex, 1995.

[29] M. Hanke and P. C. Hansen. Regularization methods for large-scale problems. *Surveys on Mathematics for Industry*, 3:253–315, 1993.

[30] M. Hanke and G. Nagy. Restoration of atmospherically blurred images by symmetric indefinite conjugate gradient techniques. *Inv. Prob.*, 12:157–173, 1996.

[31] P. C. Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Rev.*, 34(4):561–580, 1992.

[32] P. C. Hansen. Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems. *Numer. Alg.*, 6:1–35, 1994.

[33] P. C. Hansen. *Rank-Deficient and Discrete Ill-Posed Problems.* SIAM, Philadelphia, 1998.

[34] M. E. Kilmer and D. P. O'Leary. Choosing regularization parameters in iterative methods for ill-posed problems. *SIAM J. Matrix Anal. Appl.*, 22:1204–1221, 2001.

[35] K. Kunisch and Z. Jun. Iterative choices of regularization parameter in linear inverse problems. *Inv. Prob.*, 14:1247–1264, 1998.

[36] D. G. Luenberger. *Linear and nonlinear programming.* Kluwer Academic Publishers , (2nd Ed.), 2003.

[37] V. A. Morozov. On the solution of functional equations by the method of regularization. *Soviet Math. Dokl.*, 7:414, 1966.

[38] V. A. Morozov. *Methods for solving incorrectly posed problems.* Springer-Verlag, New York, 1984.

[39] J. G. Nagy, K. Palmer Lee, and L. Perrone. Iterative methods for image restoration: a Matlab object oriented approach. *Numer. Alg.*, 36:73–93, 2003.

[40] J. G. Nagy, R. J. Plemmons, and T. C. Torgersen. Iterative image restoration using approximate inverse preconditioners. *IEEE Trans. Image Proc.*, 5(7):1151–1162, 1996.

[41] A. Neumaier. Solving ill-conditioned and singular linear systems: a tutorial on regularization. *SIAM Rev.*, 40:636–666, 1998.

[42] J. Nocedal and S. J. Wright. *Numerical optimization.* Springer-Verlag New York, Inc., 1999.

[43] K. Riley. *Two-Level Preconditioners for Regularized Ill-Posed Problems.* PhD thesis, July 1999.

[44] K. Riley and C. R. Vogel. Two-level preconditioners for ill-conditioned linear systems with semidefinite regularization. Submitted to Journal of Computational and Applied Mathematics.

[45] M. C. Roggemann and B. Welsh. *Imaging through torbolence.* CRC Press, Boca Raton, Florida, 1996.

[46] M. Rojas. A large-scale trust-region approach to the regularization of discrete ill-posed problems. Ph.d. thesis, technical report 98-19, Department of Computational and Applied Mathematics, Rice University,, 1998.

[47] M. Rojas and D. C. Sorensen. A trust-region approach to the regularization of large-scale discrete forms of ill-posed problems. *SIAM J. Sci. Comput.*, 26(3):1843–1861, 2002.

[48] D. Sorensen. Minimization of a large-scale quadratic function subject to a spherical constraint. *SIAM J. Optim.*, 7:141–161, 1997.

[49] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-Posed Problems*. John Wiley, New York, 1977.

[50] C. R. Vogel. *Computational Methods for Inverse Problems*. SIAM, 2002.

[51] C. R. Vogel and M. E. Oman. Iterative methods for total variation denoising. *SIAM J. Sci. Comput.*, 17:227–238, 1996.

[52] C. R. Vogel and M. E. Oman. Fast, robust total variation–based reconstruction of noisy, blurred images. *IEEE Trans. on Image Processing*, 7:813–824, 1998.

[53] G. Wahba. Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.*, 14:651–667, 1977.

[54] Y. Wang and T. Xiao. Fast realization algorithms for determining regularization parameters in linear inverse problems. *Inv. Prob.*, 17:281–291, 2001.