分类号	 密级	
UDC	 编号	

中国科学院研究生院硕士学位论文

最小二乘法及其应用

文 再 文

指	导	教	师	袁亚湘(研究员,博士)
				中国科学院数学与系统科学研究院
申 :	请学	位级	别	硕士学科专业名称计算数学
论	文 提	交日	期	论文答辩日期2004.5
学	位培	养单	位	中国科学院数学与系统科学研究院
学	位授	予单	位	中国科学院研究生院
				答辩委员会主席

Least Squares and their Applications

Zaiwen Wen

Directed by Prof. Ya-xiang Yuan

Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and System Sciences, Chinese Academy of Sciences 摘要 ii

摘要

最小二乘问题是一个相当古老的课题。随着科技的发展,不仅在数学内涵的本身它被赋予了新的内容,而且被应用到很多研究领域。如神经网络、化学、物理、金融、经济、机械系统、通讯、电子电气工程、医疗成像等等中,人们都喜欢用最小二乘问题来描述模型函数与实际观测值之间的误差,以找到一组使误差最小的参数。这些新的应用,或者由于规模很大,或者由于不适定性(病态)或高度非线性,给最小二乘问题的求解带来了很大的困难。本论文试图从这些方面对最小二乘问题进行一些新的探讨。

本论文分为两个部分。第一部分(第二章,第三章,第四章)主要讨论线性最小二乘问题及其应用,重点研究不适定的线性最小二乘问题。第二部分(第五章,第六章)我们集中研究非线性最小二乘问题及其应用。

在第二章中,我们首先回顾了经典的最小二乘解。然后通过一个数值算例说明最小二乘问题的不适定性,接着分析了这种不适定性,给出了不适定性的定义。最后分别从两个不同的角度提出了克服不适定性的正则化方法:用逼近原问题的良态亏秩矩阵或良态满秩矩阵来得到近似解。这些方法是分别通过截断 SVD (TSVD) 分解算法、Tikhonov 正则化和信赖域方法来实现的。同时我们也说明了求解线性最小二乘问题的信赖域方法也是一种正则化方法。

在第三章中我们考虑了截断 SVD (TSVD) 分解算法在信号处理中的应用。这主要是通过研究傅立叶矩阵的特殊结构来实现的。我们从理论上得到了傅立叶矩阵实部和虚部的一个简单易行的奇异值分解,从而在计算速度上使得这种应用成为可能。而且由于奇异值分解充分挖掘了矩阵的性质,算法也是很稳定的。

在第四章中我们考虑了信赖域算法在图像恢复中的应用。主要是利用 Lanczos 方法与共轭梯度法的等价性,将 Lanczos 方法综合进求解信赖域子问题的截断共轭梯度法,在不增加多少计算量的基础上,有效的提高了求解的精度和速度。数值试验表明我们的这一方法在遥感图像恢复中是成功的。

在第五章中我们首先从不同角度回顾了 Gauss-Newton 法,然后在信赖域框架内分析了 Levenberg-Marquardt 方法和拟牛顿法。接着对于解非线性最小二乘的信赖域子问题,我们提出了 M 范数方法,化成一个等价的对角系统来解。此外我们概述了带约束的非线性最小二乘问题的一些解法。特别的我们还介绍了可分非线性最小二乘问题及其解法。最后分析了最小二乘问题的一些数值困难。

在第六章中,我们介绍了最小二乘法在热中子时间谱测井谱的解析中的应用。 最后在第七章中,我们对全文做了一个总结,简要的阐述了我们的研究成果并提出了 摘要 iii

一些有趣的仍未解决的问题。

关键词: 线性最小二乘,非线性最小二乘,不适定性,正则化,截断 SVD , 信赖域,Lanczos , Gauss-Newton 法,Levenberg-Marquardt 方法,拟牛顿法,M 范数方法,可分非线性最小二乘

Abstract iv

ABSTRACT

The least Square problem is an very old problem in mathematics. With the development of science and technology, Not only new ingredients are added into itself, but also people in many fields such as chemistry, physics, neural network, finance, economic, mechanical system, communication, electronic engineering, medical imaging and so on always like to use least squares to measure the discrepancies between their model and observations. These new applications, which usually are large scale problems, or highly nonlinear or ill-posed, make the traditional numerical methods difficult to single out a useful solution. This dissertation intents to discuss the least square problem from those new aspects.

The contents are divided into two parts. The first part (Chapter 2, Chapter 3 and Chapter 4) mainly discuss the linear least square problem and it's application, especially focus on the so called ill-posed linear least square problem. In the second part (Chapter 5 and Chapter 6) we concentrate on nonlinear least square problem.

In Chapter 2, we review the traditional least square method first. Then we give the definition of ill-posed problem by a simple example and analysis it. At last we present methods of regularization to overcome the ill-posedness from two different perspective: use the well-posed rank deficient matrix or the well posed full rank matrix to approximate the original problem. Those methods are realized by truncated singular decomposition method(TSVD), Tikhonov regularization and trust region method.

In Chapter 3, we apply the TSVD method to signal processing. Exploring the special structure of the Fourier matrix, we get a simple but practical SVD of the real and imaginary part of the Fourier matrix. So it is possible to develop efficient algorithm from computation. Because SVD fully utilize the property of matrix at most, this algorithm is also very robust.

In Chapter 4, we apply a trust region method to image restoration. We utilize the equivalence between Lanczos method and the conjugate gradient method, integrate Lanczos method into the truncated conjugate gradient method to solve the trust region subproblem. As a result, without increasing computation cost, we enhance the accuracy and speed of the algorithm. Numerical tests in the restoration of remote sensing prove that our method is very successful.

In Chapter 5, we examine the Gauss-Newton method from different point of views, and analyze the Levenberg-Marquardt method and Quasi-Newton method in the trust region framework. Then we present the M-norm method to solve the trust region subproblem of nonlinear least square problem, that is, we only need to solve a diagonal system deduced from the original subproblem. We also summarized the methods of solving constraint nonlinear least square problem. Particularly, we introduced the separable

Abstract v

least square problem. At last, we analysis some numerical difficulties of the least square problem.

In Chapter 6, we study a practical problem in the analysis of time-spectrum recorded in thermal neutron time spectrometry logging.

Finally, in Chapter 7, we give an overview of this thesis. Several interesting problems are also presented in this chapter.

 $\label{eq:Keywords} \textbf{Key words and phrases:} \quad \text{Least Square, ill-posed regularization, trust region, Lanczos} \ , \ \text{Gauss-Newton method, Levenberg-Marquardt method, Quasi-Newton method, M-norm, separable nonlinear least square.}$

目 录

第一章	引言	1
§1.1	背景介绍	1
§1.2	最小二乘问题	2
第一部	分 线性最小二乘法及其应用	4
第二章	线性最小二乘问题	5
$\S 2.1$	最小二乘解的数学性质	Ę
$\S 2.2$	不适定问题及其最小二乘描述	6
$\S 2.3$	一个简单的例子	8
$\S 2.4$	直接正则化方法: TSVD 正则化	10
$\S 2.5$	直接正则化方法: Tikhonov 正则化	11
$\S 2.6$	迭代正则化方法: 信赖域方法	13
第三章	时域有限信号重构的 TSVD 算法	17
§3.1	引言	17
$\S 3.2$	张量积	19
$\S 3.3$	离散傅立叶矩阵 DFT 的 SVD 分解	21
$\S 3.4$	TSVD 在信号处理中的应用	26
$\S 3.5$	数值算例	28
$\S 3.6$	讨论和结论	30
第四章	图像恢复的信赖域方法	37
$\S 4.1$	引言	37
$\S 4.2$	Toeplitz 型与 Circulant 型矩阵	38
$\S 4.3$	解信赖域子问题的截断 Lanczos 方法	41
§4.4	数值算例	45
第二部	分 非线性最小二乘法及其应用	51
第五章	非线性最小二乘问题	52

录		vi
	录	录

§5.1	无约束非线性最小二乘问题	52
$\S 5.2$	Gauss-Newton 法	53
$\S 5.3$	Levenberg-Marquardt 方法	55
$\S 5.4$	拟牛顿方法	57
$\S 5.5$	M 范数意义下的信赖域方法	59
$\S 5.6$	带约束的最小二乘问题及其解法	62
$\S 5.7$	可分非线性最小二乘问题与变量投影法	64
$\S 5.8$	数值困难	66
第六章	非线性最小二乘法在热中子测井技术中的应用	68
$\S 6.1$	中子时间谱	68
$\S 6.2$	中子时间谱分析	71
$\S6.3$	数值试验	72
$\S6.4$	讨论和总结	77
第七章	结论与展望	7 8
参考文章	献	7 9
致谢		82
攻读硕:	士学位期间部分工作	83

插图目录

3.1	DFT 矩阵实部的特征值分布	19
3.2	测试 1) 左:输入信号;右:精确频率信息	31
3.3	测试 1) 左: 带噪声, 残缺的频谱信号; 右: 重构信号	31
3.4	测试 1) 仅带噪声的重构信号	31
3.5	测试 2) 左:输入信号;右:精确频率信息	32
3.6	测试 2) 左: 带噪声, 残缺的频谱信号; 右: 重构信号	32
3.7	测试 2) 仅带噪声的重构信号	32
3.8	测试 3) 左:输入信号;右:精确频率信息	33
3.9	测试 3) 左: 带噪声, 残缺的频谱信号; 右: 重构信号	33
3.10	测试 3) 仅带噪声的重构信号	33
3.11	测试 4) 左: 输入信号; 右: 精确频率信息	34
3.12	测试 4) 左: 带噪声, 残缺的频谱信号; 右: 重构信号	34
3.13	测试 5) 左: 输入信号; 右: 精确频率信息	34
3.14	测试 5) 左: 带噪声, 残缺的频谱信号; 右: 重构图像	35
3.15	测试 6) 左: 输入信号; 右: 精确频率信息	35
3.16	测试 6) 左: 带噪声, 残缺的频谱信号; 右: 重构图像	36
4.1	真实图像:加利福利亚飞机场,Madrid 体育场和北京市鸟瞰图	48
4.2	噪声水平 $\delta = 0.001$ 下的模糊化图像	48
4.3	噪声水平 $\delta = 0.005$ 下的模糊化图像	48
4.4	噪声水平 $\delta = 0.01$ 下的模糊化图像	48
4.5	噪声水平 $\delta = 0.02$ 下的模糊化图像	48
4.6	不带预条件的恢复图像, 噪声水平 $\delta = 0.001$	49
4.7	不带预条件的恢复图像, 噪声水平 $\delta = 0.005$	49
4.8	不带预条件的恢复图像, 噪声水平 $\delta = 0.01$	49
4.9	不带预条件的恢复图像, 噪声水平 $\delta = 0.02 \dots$	49
4.10	带预条件的恢复图像, 噪声水平 $\delta = 0.001$	49
4.11	带预条件的恢复图像, 噪声水平 $\delta=0.005$	50
4.12	带预条件的恢复图像, 噪声水平 $\delta=0.01$	50
4.13		50

国日	
图目	1X
DI LI	123

6.1	中子计数率一时间关系图	73
6.2	$ar{f} - lpha - au$ 关系图	74

第一章 引言

§1.1 背景介绍

最小二乘在许多实际问题中都有很重要的应用。在过去的几十年中,它已经俨然是一个硕果累累的研究领域。人们在描述化学、物理、金融、经济中的参数模型时,通常用形如:

$$f(x) = \frac{1}{2} \sum_{j=1}^{m} [y_j - \phi(x; t_j)]^2 \quad , \tag{1.1.1}$$

的函数来度量模型与系统在不同观测点处输出的偏差,其中 y_j 为系统在时间 t_j 处的观测值, $\phi(x;t)$ 为人们描述系统的模型。然后极小化函数 (1.1.1),寻找一组参数 x,以使得模型最靠近所给的观测值。在实际中, y_j 由于试验装置或测量误差可能包含或多或少的噪声。而模型函数 $\phi(x;t)$ 可以是一个显式函数,也可以是一个无法显式积出的积分,还可以是一个常微分方程或者偏微分方程的带参数的解。通常求解(极小化)模型 (1.1.1) 也称为数据拟和 (Data fitting),参数估计 (Parameter estimation) 或回归分析 (Linear\Nonlinear Regression)。

实际上,式 (1.1.1) 中的平方和函数并不是度量模型与观测值偏差的唯一方式。我们还可以用 5 次和式:

$$\sum_{j=1}^{m} [y_j - \phi(x; t_j)]^5 \tag{1.1.2}$$

或绝对值的和:

$$\sum_{j=1}^{m} |y_j - \phi(x; t_j)| \tag{1.1.3}$$

甚至更复杂的函数来描述。尽管形如 (1.1.3) 的准则在很多场合中比较合适, 然而最小二乘准则, 正如我们在下面叙述中将要看到的一样, 在统计上有着深刻的背景。

我们稍微改变一下记号,把数据拟和问题中模型与观测值之间的偏差记为 ϵ_i ,即:

$$\epsilon_i = y_i - \phi(x; t_i)$$
.

还假设 ϵ_j 独立同分布,其方差为 σ^2 ,概率密度函数为 $g_{\sigma}(\cdot)$ (通常这是合理的)。在这个假设下,假定真实参数是 x,则一组特定的观测值 $y_j, j = 1, 2, \cdots, m$ 的似然函数定义为:

$$p(y; x, \sigma) = \prod_{j=1}^{m} g_{\sigma}(\epsilon_j) = \prod_{j=1}^{m} g_{\sigma}(y_j - \phi(x; t_j)) \quad . \tag{1.1.4}$$

如果我们知道观测值 y_j , $j=1,2,\cdots,m$ 的值,最佳的参数 x 就可以通过关于 x 极大化函数 $p(y;x,\sigma)$ 来得到。这就是所谓的参数 x 的极大似然估计。

引言 2

当假设偏差服从正态分布时,我们有:

$$g_{\sigma}(\epsilon) = \frac{1}{\sqrt{2\pi\sigma}} \exp(-\frac{\epsilon^2}{2\sigma^2})$$
.

将上式代入式 (1.1.4) 中,那么:

$$p(y; x, \sigma) = (2\pi\sigma^2)^{-\frac{m}{2}} \exp\left(-\frac{1}{2} \sum_{j=1}^{m} \frac{[y_j - \phi(x; t_j)]^2}{\sigma^2}\right)$$
(1.1.5)

对于任意固定的方差 σ^2 ,很明显极大化函数 p 相当于极小化平方和式 (1.1.1)。总的来说,当偏差是服从独立同分布的正态分布函数时,极大似然估计的解即为极小化平方和式 (1.1.1) 的解。

前面关于 ϵ_j 的假设很常见,但它们并不是唯一说明平方和函数的极小点具有统计意义的情形。在 [1](Nonlinear Regression)中,Seber 和 Wild 举了很多例子说明极小化形如式 (1.1.1),或者更广义的形如:

$$r^T V r$$
 , $V \in \mathbb{R}^{m \times m}$ 是对称的 , 且 $r_j(x) = y_j - \phi(x; t_j)$

的函数是从观测数据中得到参数估计的关键步骤。

§1.2 最小二乘问题

我们通常称形如

$$\min_{x} f(x) = \frac{1}{2} \sum_{j=1}^{m} r_j^2(x) = \frac{1}{2} ||r(x)||_2^2$$
(1.2.1)

的问题为最小二乘问题,其中 $r(x)=(r_1(x),r_2(x),\cdots,r_m(x))^T$, $r_j(x)$ 称为余量,是从 \mathbb{R}^n 到 \mathbb{R} 的光滑函数,如在前面提到的参数估计或数据拟和中, $r_j(x)$ 即为 $y_j-\phi(x;t_j)$ 。另外,在本文中还假定 $m\geq n$ 。

函数 f 及其导数较一般的无约束优化问题有特殊结构。引入 f(x) 的 Jacobian 矩阵:

$$J = \begin{pmatrix} \frac{\partial r_1}{\partial x_1} & \cdots & \frac{\partial r_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial r_m}{\partial x_1} & \cdots & \frac{\partial r_m}{\partial x_n} \end{pmatrix}, \tag{1.2.2}$$

则我们可以将 f(x) 的梯度和 Hessian 阵表达为:

引言 3

$$\nabla f(x) = \sum_{j=1}^{m} r_j(x) \nabla r_j(x) = J(x)^T r(x)$$
(1.2.3)

$$\nabla^2 f(x) = \sum_{j=1}^m \nabla r_j(x) \nabla r_j(x)^T + \sum_{j=1}^m r_j(x) \nabla^2 r_j(x)$$

$$= J(x)^{T}J(x) + \sum_{j=1}^{m} r_{j}(x)\nabla^{2}r_{j}(x)$$
(1.2.4)

如果函数 r_i 是线性的 (如 $\phi(x,t)$ 是线性的),则 Jacobian 矩阵是常数矩阵,因此我们有:

$$f(x) = \frac{1}{2} ||Jx + r|| \tag{1.2.5}$$

其中 r = r(0)。这就是所谓的线性最小二乘问题。线性最小二乘问题本身有很广的应用,很多模型是线性的,而且它们也是非线性最小二乘算法的子问题,因此这类问题的研究有其自身的价值。但历史的原因,求解线性最小二乘问题更多是属于数值线性代数的而不是最优化研究范畴。

如果函数 r_i 是非线性的,情况就复杂些。不过在很多应用中,计算 r_i 的一阶偏导是可能的,也就是能显式的得到 Jacobian 矩阵 J(x),这样就能根据式 (1.2.3) 就能计算出梯度。然而最小二乘问题的最显著的特点在于一旦知道了 Jacobian 矩阵,我们也就计算出了 Hessian 阵 $\nabla^2 f(x)$ 的前半部分。而且,式 (1.2.4) 中的项 $J(x)^T J(x)$ 经常比第二个和式更重要,因为模型在解的附近经常是线性的(即 $\nabla^2 r_j$ 比较小)或者余量较小(即 r_j 比较小)。大多数非线性最小二乘算法都考虑了 Hessian 阵结构上的的这一特殊性质。

本论文分为两个部分。第一部分我们将集中探讨线性最小二乘问题及其应用。重点的研究求解不适定(病态的)的线性最小二乘问题的算法,并将其应用到信号处理和图像恢复中。第二部分我们将集中探讨非线性最小二乘问题,主要利用其特殊结构,研究解非线性最小二乘的信赖域算法,并将其应用到热中子时间谱测井谱的解析中。

第一部分

线性最小二乘法及其应用

第二章 线性最小二乘问题

§2.1 最小二乘解的数学性质

设 $A \in \mathbb{R}^{m \times n} (m \ge n), b \in \mathbb{R}^m$ 。所谓的线性最小二乘问题(简称 LS 问题),是指求 $\bar{x} \in \mathbb{R}^n$ 使得:

$$||A\bar{x} - b||_2 = \min\{||Ax - b||_2 : x \in \mathbb{R}^n\}$$
 (2.1.1)

记

$$\mathcal{X}_{LS} = \{ \bar{x} \in \mathbb{R}^n : \bar{x} 满足 (2.1.1) \} ,$$

则称 \mathcal{X}_{LS} 是最小二乘问题 (2.1.1) 的解集; \mathcal{X}_{LS} 中 2 范数最小的称为最小二乘解, 记做 \mathcal{X}_{LS} , 即:

$$||x_{LS}||_2 = \min\{||\bar{x}||_2 : \bar{x} \in \mathcal{X}_{LS}\}$$
.

定理2.1.1. $x \in \mathcal{X}_{LS}$ 当且仅当 $A^{T}(Ax - b) = 0$ 。

定理2.1.2. 最小二乘问题 (2.1.1) 的解由

$$x = A^{\dagger}b + (I - A^{\dagger}A)z$$

给出,其中z表示 \mathbb{R}^n 中的任一向量;而且其最小的范数解由 $x_{LS}=A^{\dagger}b$ 给出。

其中 A^{\dagger} 为 A 的 Moore-Penrose 广义逆。可以证明 A^{\dagger} 是由 A 唯一确定的,而且若 A 的 SVD 分解为:

$$A = U \begin{pmatrix} \Sigma_r & 0 \\ 0 & \Sigma_0 \end{pmatrix} V^T \quad , \tag{2.1.2}$$

其中 $U=[u_1,u_2,\cdots,u_m]$ 和 $V=[v_1,v_2,\cdots,v_n]$ 分别是按列划分的 m 和 n 阶的正交矩阵,r=rank(A) , $\Sigma_r=diag(\sigma_1,\cdots,\sigma_r)$, $\Sigma_0=diag(\sigma_{r+1},\cdots,\sigma_n)$, $\sigma_1\geq\cdots\geq\sigma_r>\sigma_{r+1}=\cdots=\sigma_n=0$,则

$$A^{\dagger} = V \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^T \quad .$$

当然 SVD 分解提供了 x_{LS} 一个简洁的表达式和最小二乘剩余量的范数 $\rho_{LS} = ||Ax_{LS} - b||_2$,我们有下面的定理:

定理2.1.3. $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, 设 A 的 SVD 由式 (2.1.2) 给出,则

$$x_{LS} = \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} v_i \tag{2.1.3}$$

使 $||Ax-b||_2$ 极小化,且是所有极小点中 2 范数最小的,而且

$$\rho_{LS}^2 = ||Ax_{LS} - b||_2^2 = \sum_{i=r+1}^m (u_i^T b)^2.$$
(2.1.4)

求解最小二乘问题最古老而且现在仍常用的方法之一是正规化方法,这一方法的理论依据就是定理 2.1.1,即将求解问题 (2.1.1) 转换为求解法方程组:

$$A^T A x = A^T b$$
 .

当 rank(A) = n 时, A^TA 是对称正定方程组,因而其解是唯一的,可用 cholesky 分解来求得。但此时解的相对误差依赖于条件数的平方。另一个常用的方法是正交化方法,由于矩阵的 2 范数是正交不变的,因此对于任意的正交矩阵 $Q \in \mathbb{R}^{m \times m}$,问题 (2.1.1) 等价于:

$$||Q^T(A\bar{x}-b)||_2 = \min\{||Q^T(Ax-b)||_2 : x \in \mathbb{R}^n\}$$
.

我们可以适当选取正交矩阵 Q,这样就使原问题转化成一个较为容易求解的最小二乘问题。实现正交化的基本方法有 Householder 方法,Givens 方法等等。

对于最小二乘问题通常的这些数值解法,一般大学教程 [2, 3] 里都有详尽的介绍,我们在这里就不再详细的展开讨论。虽然最小二乘问题很经典,但当数据有扰动的时候,特别当矩阵 A 成病态的时候,即条件数很大,甚至亏秩的时候,用传统的这些方法要得到一个满意的解就会有相当大的困难,甚至和精确解相差很远。这时最小二乘问题是一个不适定问题,需要用特殊的技巧,如正则化方法来求解。在接下来的内容里,我们将重点介绍克服这些不适定性的正则化方法。

§2.2 不适定问题及其最小二乘描述

病态问题这个概念最早可以追溯到上世纪初的 Hadamard。Hadamard 实质上定义了一个问题是**不适定的(ill-posed**)如果其解不唯一或者它不是关于数据的连续函数 — 即数据的任意微小的扰动能使解产生任意大的扰动。Hadamard 认为不适定问题只是"人造的 (artificial)",它们不能描述现实的物理系统。但事实上近年来,从科学和工程领域 [4, 5, 20, 21, 22] 却涌现出了大量的不适定问题。

不适定问题的一个经典的例子是带第一类平方可积积分核的 Fredholm 积分方程:

$$\int_{a}^{b} K(s,t)f(t) dt = g(s) \quad , \quad c \le s \le d \quad , \tag{2.2.1}$$

其中右端项 q 和积分核 K 是给定的, f 是要求的解。如果解 f 加上一个微小扰动:

$$\Delta f(t) = \epsilon \sin(2\pi pt)$$
 , $p = 1, 2, \dots$, ϵ 为常数 ,

则相应的右端项g的扰动为:

$$\Delta g(s) = \epsilon \int_a^b K(s,t) \sin(2\pi pt) dt$$
 , $p = 1, 2, \cdots$.

由 Riemann-Lebesgue 引理可以得到当 $p \to \infty$ 时 $\Delta g \to 0$ 。因此,p 足够大时, $\|\Delta f\|/\|\Delta g\|$ 可以变得任意大。这表明 (2.2.1) 是一个不适定问题。

严格的讲,不适定问题必须是无穷维的,否则 $\|\Delta f\|/\|\Delta g\|$ 有界,尽管它可以变得非常大。但是,一些有穷维的离散问题有无穷维问题的类似性质,如对高频扰动高度敏感。因此,很自然的可以称之为离散不适定问题或数值不适定问题(见文献 [21, 22])。

下面我们给出不适定问题的一个一般的定义:

定义2.2.1. 设 F,U 均为度量空间, ρ_F,ρ_U 分别是它们所对应空间上的度量,算子 $A: F \to U$ 把 F 映照到 U,则问题或方程:

$$Az = u$$
 , $z \in F$, $u \in U$ (2.2.2)

称为适定的,如果它们同时满足下述三个条件:

- 1. $\forall u \in U$, 都存在 $z \in F$ 满足方程 (2.2.2) (解的存在性);
- 2. 设 $u_1, u_2 \in U$,若 z_1, z_2 分别是方程 (2.2.2) 对应于 $u_1 \neq u_2$ 的解,则 $z_1 \neq z_2$ (**解的唯一**性);
- 3. 解相对于空间偶 (F,U) 而言是稳定的 (解的稳定性), 即: $\forall \epsilon > 0, \exists \delta(\epsilon) > 0$, 只要

$$\rho_U(u_1, u_2) \le \delta(\epsilon) \qquad (u_1, u_2 \in U)$$

便有:

$$\rho_F(z_1, z_2) < \epsilon$$
 $(Az_1 = u_1, Az_2 = u_2)$

反之, 若上述三个条件中, 至少有一个不能满足, 则成为是不适定的。

在定义 2.2.1 中,算子 A 可为积分算子、微分算子或矩阵 (有限秩算子),这在 [22] 中有更详细的讨论。

离散不适定问题的通常的表现形式是由不适定问题离散化之后得到的线性方程组系统和线性最小二乘问题,即如果 Fredholm 积分方程 (2.2.1) 用有限元类型的方法来离散,则会产生形如

$$Ax = b$$
 , $A \in \mathbb{R}^{n \times n}$ (2.2.3)

或

$$\min_{x} \|Ax - b\|_2 \quad , \quad A \in \mathbb{R}^{m \times n} \quad , \quad m > n \quad , \tag{2.2.4}$$

的系统。其中矩阵 A 的元素 a_{ij} 和右端项 b 的元素 b_i 如下:

$$a_{ij} = \int_a^b \int_c^d K(s,t)\phi_i(s)\psi_j(t) \ ds \ dt \quad , \quad b_i = \int_c^d \phi_i(s)g(s) \ ds \quad ,$$

其中 ϕ_i, ψ_j 是有限元方法中用到的基函数。对于这类问题,由于第一类积分方程的求解是不适定的,故由它直接导出的离散线性系统 (2.2.3) 或 (2.2.4) 即使可解,也将是病态的,并且随着离散步长的缩小而不断加剧。

下面我们利用 SVD 分解对离散系统的不适定性做进一步的说明。设 A 的 SVD 分解由式 (2.1.2) 给出,则可以得出关系式:

$$Av_i = \sigma_i u_i$$

$$||Av_i||_2 = \sigma_i$$

$$i = 1, \dots, n .$$

 $||Av_i||_i = \sigma_i$ 相对于 $||A||_2 = \sigma_1$ 比较小,意味着 A 的列之间存在着很大程度上的线性相关。也就是说小的奇异值 σ_i 表明 A 接近亏秩,而与 σ_i 对应的向量 v_i 在数值上为 A 的零空间向量。从 A 的这个性质可以得出离散不适定系统通常是高度病态的,其零空间由符号改变得很厉害的向量张成。

这里我们再给出离散不适定系统的一个更直观的定义:

定义2.2.2. 线性方程系统 (2.2.3) 和线性最小二乘问题 (2.2.4) 称之为离散不适定问题, 如果它们同时满足以下两条准则:

- 1. 最大非零奇异值与最小非零奇异值的比例非常大,
- 2. A 的奇异值最终蜕化为 0。

准则 1 说明矩阵 A 是病态的,即其解对扰动很敏感。准则 2 表明没有一个适定的,数值秩很好的矩阵与之"很接近"。

有意思的是,虽然不适定问题是"病态的",但这并不意味着我们就不能计算得到有意义的解。然而,正是这种"病态性"使得数值代数 [2, 3] 中求解问题 (2.2.3) 和 (2.2.4) 方法,如 LU, Cholesky 或 QR 方法不能直接用来计算这类不适定问题。必须用其它的技巧来保证计算得到的解有意义。这正是正则化方法的目标所在。

考虑最小二乘问题:

$$\min_{x} ||Ax - b||_{2} , \quad \sharp + A = \begin{pmatrix} 0.16 & 0.10 \\ 0.17 & 0.11 \\ 2.02 & 1.29 \end{pmatrix} , \quad b = \begin{pmatrix} 0.27 \\ 0.25 \\ 3.33 \end{pmatrix} .$$

在这里,右端项 b 是精确解 $x^* = (1 \ 1)^T$ 对应的右端项 b^* 加上一个微小的扰动 δ 后产生的,即:

$$b = \begin{pmatrix} 0.16 & 0.10 \\ 0.17 & 0.11 \\ 2.02 & 1.29 \end{pmatrix} \begin{pmatrix} 1.00 \\ 1.00 \end{pmatrix} + \begin{pmatrix} 0.01 \\ -0.03 \\ 0.02 \end{pmatrix} .$$

这个问题的难点在于矩阵 A 成病态,其条件数为 $1.1 \cdot 10^3$ 。这就意味着数值计算过程中解对数据的微小扰动都很敏感。实际上,如果我们用通常的最小二乘法,如 A 的 QR 分解来求解,其解 x_{LSO} 为:

$$x_{LSQ} = \begin{pmatrix} 7.01 \\ -8.40 \end{pmatrix} \quad .$$

这个解显然与我们所需要的相差很远,必须通过其它的方式来得到更好的近似。矩阵 $A = (a_1 \ a_2)$ 的条件数很大表明 A 的列 a_1 和 a_2 几乎线性相关。可以考虑用 $(a_1 \ 0)$ 或 $(0 \ a_2)$ 来替换 A 求解。相应的解称为基本解,分别为:

$$x_B^1 = \begin{pmatrix} 1.65 \\ 0.00 \end{pmatrix} \quad , \quad x_B^2 = \begin{pmatrix} 0.00 \\ 2.58 \end{pmatrix} \quad .$$

尽管相应的余量很小:

$$||Ax_B^1 - b||_2 = 0.031 \quad ||Ax_B^2 - b||_2 = 0.036$$

但基本解却与精确解 $x^* = (1 \ 1)^T$ 没有多少共同之处。

一般的最小二乘解 x_{LS} 的主要困难之处在于其范数明显的大于精确解的范数。因此可以在原问题的基础上,对解添加约束使其范数不超过某个值 α ,即求解:

$$\min_{x} \|Ax - b\|_2 \quad , \quad \|x\| \le \alpha \quad .$$

上述问题的解 x_{α} 非线性依赖于参数 α 的选取。如 α 取 0.1,1,1.385,10 时,解分别为:

$$x_{0.1} = \begin{pmatrix} 0.08 \\ 0.05 \end{pmatrix}$$
 , $x_1 = \begin{pmatrix} 0.84 \\ 0.54 \end{pmatrix}$, $x_{1.385} = \begin{pmatrix} 1.17 \\ 0.74 \end{pmatrix}$, $x_{10} = \begin{pmatrix} 6.51 \\ -7.60 \end{pmatrix}$.

可以看到,如果 α 选取恰当,我们就可以得到精确解的较好近似,如 $x_{1.385}$. 但值得注意的是,选取 α 必须非常小心,而且什么是好的 α 并不明显。

上面的例子虽然很小,但突出了离散的不适定问题的三个主要困难:

- 1. 矩阵 A 的条件数很大。
- 2. 用一个从 A 得到的条件数好的矩阵替代 A 并不意味着能得到一个好的解。
- 3. 在添加额外约束条件时必须谨慎。

§2.4 直接正则化方法: TSVD 正则化

不适定系统 (2.2.3) 和 (2.2.4) 的主要困难在于它们被 A 的小奇异值簇大大的削弱了。一个很自然的正则化方法是用一个良态的亏秩矩阵 B 来近似 A。所谓良态的亏秩矩阵是指这个矩阵的条件数,即最大的非零奇异值与最小的非零奇异值的比值:

$$cond(B) \stackrel{\text{def}}{=} \|B\| \|B^{\dagger}\| = \frac{\sigma_1}{\sigma_{rank(B)}}$$

不是很大。下面我们通过 SVD 分解来找这样的一个近似矩阵。

SVD 分解深刻的揭露了矩阵的结构。设 A 的 SVD 由式 (2.1.2) 给出,则可以得到以下丰富信息:

$$null(A) = span\{v_{r+1}, \cdots, v_n\}$$

$$||A||_F^2 = \sigma_1^2 + \cdots + \sigma_n^2$$

$$range(A) = span\{u_1, \cdots, u_r\}$$

$$A = \sum_{i=1}^r \sigma_i u_i v_i^T$$

$$||A||_2 = \sigma_1$$

$$||A||_2 = \sigma_1$$

$$||A||_2 = \sigma_1$$

$$||A||_2 = \sigma_1$$

同时由以下定理可以刻画给定矩阵与比其秩低的矩阵的靠近程度。

定理2.4.1. ([2] Golub) 设 $A \in \mathbb{R}^{m \times n}$ 的 SVD 由式 (2.1.2) 给出,如果 k < r = rank(A) 且

$$A_k = \sum_{i=1}^k \sigma_i u_i v_i^T \quad , \tag{2.4.1}$$

则:

$$\min_{rank(B)=k} ||A - B||_2 = ||A - A_k||_2 = \sigma_{k+1}$$
(2.4.2)

证明. 由于 $U^T A_k V = diag(\sigma_1, \dots, \sigma_k, 0, \dots, 0)$, 从 $rank(A_k) = k$ 和

$$U^T(A - A_k)V = diag(0, \dots, 0, \sigma_{k+1}, \dots, \sigma_p)$$

可知 $||A - A_k||_2 = \sigma_{k+1}$ 。

现假定对某一 $B \in \mathbb{R}^{m \times n}$ 满足 rank(B) = k。则存在单位正交向量 x_1, \dots, x_{n-k} 使得 $null(B) = span\{x_1, \dots, x_{n-k}\}$ 。利用维数可知:

$$span\{x_1, \dots, x_{n-k}\} \cap span\{v_1, \dots, v_{k+1}\} \neq \{0\}$$

令 z 是这一交集中的一个 2 范数单位向量。利用 Bz=0 和

$$Az = \sum_{i=1}^{k+1} \sigma_i(v_i^T z) u_i \quad ,$$

我们有:

$$||A - B||_2^2 \ge ||(A - B)z||_2^2 = ||Az||_2^2$$

= $\sum_{i=1}^{k+1} \sigma_i^2 (v_i^T z)^2$.

定理得证。 □

定理 2.4.1 表明 $\mathbb{R}^{m \times n}$ 中的满秩矩阵的集合是开的而且是稠密的。同时可知 A 的最小奇异值是从 A 到所有秩亏矩阵集合的 2 范数距离。更进一步的, A_k 是到 A 最近的秩为 k 的亏秩矩阵,其 2 范数距离是 σ_{k+1} 。

这样我们就差不多找到了一个亏秩矩阵 A_k 与 A 充分接近,但还需给定条件保证这个矩阵是良态的。取 $k(\epsilon)$ 满足要求:

$$k(\epsilon) = rank(A_{k(\epsilon)}) = \min_{\|A - B\|_2 < \epsilon} rank(B) \quad , \tag{2.4.3}$$

即:

$$\sigma_1 \ge \cdots \ge \sigma_{k(\epsilon)} > \epsilon \ge \sigma_{k(\epsilon)+1} \ge \cdots \sigma_n$$
.

只要取 ϵ 合适,这样构造的 $A_{k(\epsilon)}$ 就是我们要找的良态矩阵。

定义2.4.2. 设 $A \in \mathbb{R}^{m \times n}$ 的 SVD 分解由式 (2.1.2) 给出,给定 ϵ ,取 $k(\epsilon)$ 满足式 (2.4.3),则

$$A_{k(\epsilon)} = \sum_{i=1}^{k(\epsilon)} \sigma_i u_i v_i^T \tag{2.4.4}$$

称为 A 的截断 SVD 分解 (TSVD)。在不引起混淆的情况下,记 $A_{k(\epsilon)}$ 为 A_k 。

截断 SVD 分解实质上是通过控制 ϵ 过虑掉对矩阵性质影响较大的小奇异值。在实际中, ϵ 一般可以取为观测数据的精度。

有了 A 的截断 SVD 分解 $A_{k(\epsilon)}$, 我们就将问题 (2.2.4) 转化为求解:

$$\min_{x} \|A_{k(\epsilon)}x - b\|_2$$
 , (2.4.5)

其最小二乘解为:

$$x_{k(\epsilon)} = A_{k(\epsilon)}^{\dagger} b = \sum_{i=1}^{k(\epsilon)} \frac{u_i^T b}{\sigma_i} v_i \quad . \tag{2.4.6}$$

这就是我们所说的TSVD 正则化方法。该方法在实际中取得了很好的计算效果,我们将在第三章中介绍它在数值信号处理中的一个应用。

§2.5 直接正则化方法: Tikhonov 正则化

在上一节中, 我们用一个亏秩的良态矩阵 B 来近似 A, 从而达到正则化的效果。同

样,我们也可以用一个满秩的良态矩阵来逼近 A,在这一节中,我们将介绍另一个经典的正则化方法: Tikhonov 正则化。

正则化的基本思想在于:由于系统高度病态,或观测数据不精确,解被误差淹没掉了,因此必须添加关于解的其它一些进一步的信息以规范问题,从而识别出一个有意义的,稳定的解。虽然原则上解有很多类型的附加信息可以利用,但正则化离散不定问题的主要方法是要求解的二范数或适当的半范数要小。也可以将解的一个初始估计 x* 包含在约束中,即极小化:

$$\Omega(x) = ||L(x - x^*)||_2 \quad . \tag{2.5.1}$$

这里,矩阵 L 一般取为单位矩阵 I_n 或第 $p \times n$ 阶微分算子的 $p \times n$ 离散近似,即 L 是一个满秩的带状矩阵。在一些情况下,可以把约束写成 Sobolev 范数的形式:

$$\Omega(x)^{2} = \alpha_{0}^{2} \|(x - x^{*})\|_{2}^{2} + \sum_{i=1}^{q} \alpha_{i}^{2} \|L_{i}(x - x^{*})\|_{2}^{2} ,$$

这里 L_i 是第 i 阶微分算子的近似。如果记 L 为矩阵 $\alpha_0^2 I_n + \sum_{i=1}^q \alpha_i^2 L_i^T L_i$ 的 Cholesky 分解因子,这里的 Ω 经常就可以写成 (2.5.1) 的形式。通过约束 Ω ,我们就可以控制正则解的光滑性。

在引进约束 Ω 后,就必须放弃线性系统 (2.2.3) 中 Ax 等于 b 的要求,转而寻找一个解使得极小化 $\Omega(x)$ 和极小化余量范数 $\|Ax - b\|_2$ 达到相对平衡。基本准则是使得正则解有小的范数 (半范数) 和恰当小余量,而且距离没有扰动过的原问题的未知解不远。这个原则对最小二乘问题 (2.2.4) 同样适用。

最常用,最知名的正则化方法是 Tikhonov 正则化方法。Tikhonov 正则化方法定义正则 解 x_{λ} 为问题:

$$\min_{x} \|Ax - b\|_{2}^{2} + \lambda^{2} \|Lx\|_{2}^{2}$$
 (2.5.2)

的最优解,即余量和约束 Ω 的组合形式的最优解,其中 λ 为正则化参数,它控制极小化 Ω 相对于极小化余量的权重。同时 λ 也控制正则解 x_{λ} 对 A 和 b 的扰动的敏感度。实际上,求解最小二乘问题 (2.5.2) 相当于求解最小二乘问题:

$$\min_{x} \quad \left\| \begin{pmatrix} A \\ \lambda L \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|_{2} , \qquad (2.5.3)$$

如果 λ 选取得当,可以使矩阵:

$$\left\| \begin{pmatrix} A \\ \lambda L \end{pmatrix} \right\|$$

是一个列满秩良态矩阵。从而达到寻找近似解的目的。

此外极小化问题 (2.5.2) 也相当于求解线性代数方程组:

$$(A^T A + \lambda^2 L^T L)x = A^T b \quad . \tag{2.5.4}$$

相对于原问题 (2.2.4) 的法方程组,方程 (2.5.4) 中多了一个修正项 $\lambda^2 L^T L$,通过这个修正项,达到改善法方程组的条件数和谱(特征值)的分布的目的。因此正则化参数 λ 对正则解的性质影响极大,必须谨慎的选择。如果 λ 太小,则对问题的谱的改善没有起到什么作用,即解的不稳定性仍然存在;反之,如果 λ 太大,则我们得到的新的问题是可以稳定的求到解了,但该问题与原问题相差甚远,是一个相当糟糕的逼近。所以最优的正则参数的选取应当兼顾这两个情况。通常,参数 λ 的选取有先验和后验两种。关于后验选取正则参数,在书 [22] 中详细的阐述。关于先验选取,主要是依据经验进行,通常我们都取 $\lambda \in (0,1)$ 。

由于问题 (2.5.3) 和问题 (2.5.4) 的系数矩阵是良态的,因此可以用线性代数中的经典方法来求解。例如对于问题 (2.5.4),系数矩阵 $A^TA + \lambda^2 L^TL$ 正定,设其 Cholesky 分解为:

$$A^T A + \lambda^2 L^T L = \bar{L} D \bar{L}^T \quad ,$$

其中 \bar{L} 为下三角矩阵, 其主对角线的元素全为 1, D 为对角线矩阵。于是问题的解可以通过解下述线性方程组:

$$v = L^{-1}A^{T}b$$
$$x = L^{-T}D^{-1}v$$

得到。

§2.6 迭代正则化方法: 信赖域方法

信赖域算法是 80 年代开始兴起的一种优化算法,该方法在确保问题的全局收敛的同时又要求问题在局部的快速收敛性。比如我们考虑如下的优化问题:

$$\min_{x \in \mathbb{R}^n} f(x) \quad . \tag{2.6.1}$$

信赖域的基本思想是每一次迭代在信赖域 $\{s|||s|| \leq \Delta_k\}$ 内解原优化问题的一个逼近问题(常称为子问题)得到试探步 s_k ,该试探步使得 $x_k + s_k$ 是在以 x_k 为中心的广义球 $\{x_k + s_k\}$ 上的"最好的点",因此信赖域法不进行线搜索,而是直接令 $x_{k+1} = x_k + s_k$ 或者 $x_{k+1} = x_k$ 。

下面我们考虑用信赖域方法求解不适定的最小二乘问题:

$$\min_{x} f(x) = \frac{1}{2} ||Ax - b||_{2}^{2} , \qquad (2.6.2)$$

其中函数 f(x) 在点 x_k 处的梯度和 Hessian 阵分别可以显式的计算为:

$$g_k = \nabla f = A^T A x_k - A^T b$$
 , $B_k = A^T A$.

在这里,因为 f(x) 本身是二次的,故信赖域子问题可取为:

$$\min_{s \in \mathcal{S}} f(x_k + s) = \frac{1}{2} ||A(x_k + s) - b||_2^2$$
s.t. $||s|| < \Delta_k$. (2.6.3)

由于 $f(x_k + s) = f(x_k) + g^T s + \frac{1}{2} s^T B_k s$, 因此 (2.6.3) 相当于求解子问题:

$$\min_{s \in \mathcal{S}} \phi_k(s) = g^T s + \frac{1}{2} s^T B_k s$$
s.t. $||s|| \le \Delta_k$. (2.6.4)

设 s_k 是子问题 (2.6.4) 的解,为了检测该试探步是否值得信赖,我们定义逼近模型的子问题的下降量

$$Pred_k = \phi_k(0) - \phi_k(s_k)$$

为模型的预估下降量。而目标函数的真实下降量为:

$$Ared_k = f(x_k) - f(x_k + s_k)$$
.

这样就可以用真实下降量与预估下降量的比值:

$$r_k = \frac{Ared_k}{Pred_k} \tag{2.6.5}$$

来作为一个标准。粗略的说, r_k 越大,说明目标函数的下降越多,新的点 $x_k + s_k$ 也就越好,故 x_{k+1} 可以取为 $x_k + s_k$,而且 Δ_{k+1} 也可以考虑扩大。否则 r_k 越小,例如 $r_k < 0$,这时 $f(x_k + s_k) > f(x_k)$,故 x_{k+1} 应该取为 x_k , Δ_{k+1} 也应小于 Δ_k 。

但对于二次型问题,我们发现 $r_k \equiv 1$ 。事实上,通过解问题 (2.6.3) 或 (2.6.4),函数值 $f(x_k + s_k)$ 至少不会比 $f(x_k)$ 更差。因此不管 s_k 的好坏 (函数下降量的多少),信赖域半径总会增大。这样我们定义新的比值:

$$r_k = \frac{f(x_k + s_k)}{f(x_k)} \tag{2.6.6}$$

或

$$r_k = \frac{\|\nabla f(x_k + s_k)\|}{\|\nabla f(x_k)\|}$$
 (2.6.7)

来衡量是否接受信赖域试探步以及是否调整信赖域半径。这样做实际上也利用了问题的偏差大小,从而更好的描述了问题的本质。

基于以上分析,我们把求解不适定线性最小二乘问题的信赖域方法描述如下:

算法2.6.1. 求线性最小二乘问题的信赖域算法

步 θ : 初始化。给定初始点 x_0 和信赖域半径 Δ_0 。选取信赖域参数 $\eta_1, \eta_2, \gamma_1, \gamma_2, \gamma_3$ 使 得 $0 < \eta_1 \le \eta_2 < 1$ 和 $0 < \gamma_1 \le 1 \le \gamma_2 \le \gamma_3$ 。计算 $f(\mathbf{x_0})$ 并令 k := 0。

步 1: **求步长**。若满足停机条件,则停机;否则求解子问题 (2.6.4) 得到步长 s_k ,使得 "目标函数模型充分下降"。

步 2: 是否接受试探步。计算 $f(\mathbf{x_k} + s_k)$, 并由 (2.6.6) 或 (2.6.7) 来计算 r_k , 令

步 3: 更新信赖域半径。令

$$\Delta_{k+1} \in \begin{cases} [\gamma_2 ||s_k||, \gamma_3 \Delta_k], & \text{若} \quad r_k \leq \eta_2, \\ [\gamma_1 \Delta_k, \gamma_2 \Delta_k], & \text{否则}. \end{cases}$$

步 4: 计算 g_{k+1} , 并令 k := k+1; 转步 1。

在以上算法中参数 η_i , γ_i 的选取可由用户来定,一种典型的取法为 $\eta_1 = 0.6$, $\eta_2 = 0.6$, $\gamma_1 = 0.8$, $\gamma_2 = 1.5$, $\gamma_3 = 2$, 在算法的第二步中,我们用偏差原理来作为停机准则:当迭代点第一次出现:

$$||Ax_k - b|| \le \delta$$

时,终止迭代。其中 δ 为大于零的常数,不同的问题 δ 的取值不尽相同,在实际应用问题中, δ 通常被称为误差水平。因此 δ 的取值范围应当在 (0,1) 区间中。

对于子问题 (2.6.4), 我们有以下结论:

定理2.6.2. s_k 是子问题 (2.6.4) 的解当且仅当存在 $\lambda_k > 0$ 使得:

$$(B_k + \lambda_k I)s_k = -g_k$$

$$\|s_k\|_2 \le \Delta_k$$

$$\lambda_k (\Delta_k - \|s_k\|_2) = 0$$

$$(2.6.8)$$

并且 $B_k + \lambda_k I$ 是半正定矩阵。

下面我们来考虑怎样求解子问题 (2.6.4)。假定 $B_k + \lambda_k I$ 正定,则 s_k 可以通过求解:

$$\lambda_k (\Delta_k - \|(B_k + \lambda_k I)^{-1} g_k\|_2) = 0$$
$$\|(B_k + \lambda_k I)^{-1} g_k\|_2 \le \Delta_k \quad , \quad \lambda_k \ge 0$$

来得到 λ_k ,然后置 $s_k = (B_k + \lambda_k I)^{-1} g_k$ 。如果 B_k 正定且 $B_k^{-1} g_k \leq \Delta_k$,则 $s_k = B_k^{-1} g_k$ 。否则 $\lambda_k > 0$,只需求解:

$$\psi(\lambda) = \frac{1}{\|(B_k + \lambda_k I)^{-1} g_k\|_2} - \frac{1}{\Delta_k} = 0 \quad . \tag{2.6.9}$$

很多情况下 $\psi(\lambda)$ 近似一个线性函数,因此可以用牛顿法来求解:

$$\lambda^{+} = \lambda - \frac{\psi(\lambda)}{\psi'(\lambda)}$$

$$= \lambda - \frac{g_{k}^{T}(B_{k} + \lambda_{k}I)^{-3}g_{k}}{\|(B_{k} + \lambda_{k}I)^{-1}g_{k}\|_{2}^{3}} \left[\frac{1}{\|(B_{k} + \lambda_{k}I)^{-1}g_{k}\|_{2}} - \frac{1}{\Delta_{k}} \right]$$

事实上,我们不需要精确求解信赖域子问题 (2.6.4),可以用共轭梯度法来求解,生成的迭代点列如下:

$$s_{k+1} = s_k + \alpha_k d_k \quad , \tag{2.6.10}$$

$$d_{k+1} = -g_k + \beta_k d_k \quad , \tag{2.6.11}$$

其中:

$$\alpha_k = -g_k^T d_k / d_k^T B_k d_k \quad , \tag{2.6.12}$$

$$\beta_k = \|g_{k+1}\|_2^2 / \|g_k\|_2^2 . (2.6.13)$$

算法2.6.3. 截断共轭梯度法

步 1: 给定初始点 $s_0, \epsilon > 0$,计算 $g_0 = g(s_0)$ 并置 $k = 1, d_0 = -g_0$ 。

步 2: 如果 $||A(x_k + s_k) - b|| \le \tau ||Ax_k - b||, 0 < \tau < 1$,停机,输出 $s_k^* = s_k$; 计算 $d_k^T B_k d_k$: 若 $d_k^T B_k d_k \le 0$,转步 4; 依式 (2.6.12) 计算 α_k 。

步 3: 若 $||s_k + \alpha_k d_k|| \ge \Delta_k$, 转步 4; 依式 (2.6.10) 计算 s_{k+1} 并计算 g_{k+1} ; 依式 (2.6.11) 计算 β_k 并依式 (2.6.11) 计算 d_{k+1} k := k+1, 转步 2;

步 4: 计算 $r_k^* \ge 0$ 使其满足 $||s_k + r_k^* d_k|| = \Delta_k$ 令 $s_k^* = s_k + r_k^* d_k$,停机

王彦飞在 [28] 中证明了带信赖域技巧的截断共轭梯度法式一种正则化方法。事实上, 信赖域子问题 (2.6.3) 的 Lagrangian 函数为:

$$L(s,\lambda) = \frac{1}{2} ||A(x_k + s) - b||_2^2 + \lambda(||s|| - \Delta_k)$$
 (2.6.14)

因此极小化 $L(s,\lambda)$ 实际上和 Tikhonov 正则化 (2.5.2) 是等价的。因此信赖域方法从这个角度可以看成是迭代的正则化方法,每个信赖域子问题都等价于求解一个 Tikhonov 正则化问题。但信赖域方法具有自动调节信赖域参数,即正则化参数的功能,因此现在来越来越引起人们的关注。

我们将在第四章中介绍信赖域方法在图像复原中的一个应用。

第三章 时域有限信号重构的 TSVD 算法

§3.1 引言

所有的科学仪器当看成一个线性系统时有共同的性质:有一个输入 f,一个输出 F,再带一个加法型噪声 n。离散傅立叶变换 (DFT) 是一类点扩散函数的代表。通常时域信号和频域信号通过一个有限带宽的系统时,信息经常要损失掉。例如一个离散的时间有限时域信号 f 通过一个离散傅立叶系统,该过程可以描述为:

$$\mathcal{F}f = F, (3.1.1)$$

其中 F 是频域信号,可以通过观测或试验部分的得到(例如,一个透镜就代表一个傅立叶系统,当时域信号通过这个透镜时,我们就可以得到频率谱信息)。上面这个过程如果用线性代数方程组可以表示为:

$$\sum_{n=0}^{N-1} w_N^{kn} f_{n+1} = F_{k+1}, \tag{3.1.2}$$

其中,

$$f = [f_1, f_2, \dots, f_N]^T,$$

 $F = [F_1, F_2, \dots, F_N]^T,$

但 F_i , $i = 1, 2, \dots, N$ 只有部分信息能知道或近似的知道。在 (3.1.2) 中, $w_N = \exp(-i(2\pi/N))$ 。实际上,(3.1.2) 写成矩阵的形式为:

$$\mathcal{F}f = F \quad \text{or} \quad \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & w_N & \cdots & w_N^{N-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & w_N^{N-1} & \cdots & w_N^{(N-1)^2} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{bmatrix} = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_N \end{bmatrix}. \tag{3.1.3}$$

二维的情形可以类似的产生,这时傅立叶系统描述为:

$$\mathcal{F}f = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} w_{MN}^{lm,kn} f_{(m+1)(n+1)} = F_{(l+1)(k+1)}, \tag{3.1.4}$$

其中,

$$w_{MN}^{lm,kn} = \exp(-i2\pi(lm/M + kn/N)),$$

 $l = 0, 1, \dots, M - 1; \quad k = 0, 1, \dots, N - 1.$

 $\mathcal{F} \in \mathbb{R}^{MN \times MN}$, $f \in \mathbb{R}^{MN}$, $F \in \mathbb{R}^{MN}$, 值得注意的是, \mathcal{F} 是一个张量积, 因此我们可以利用张量积的性质来简化计算。在本章的第二节中, 我们将集中给出张量积的一些有用的性质。

很多文献 [6, 8, 7, 10, 13, 18, 19, 12] 都讨论了这个问题的应用和方法。在 [10] 和 [13] 中作者考虑了磁共振成像问题,他们的方法是求解超定方程组 (3.1.1)。这一类型的问题通常包括选取观测点,因为得到观测值 F 中的元素非常耗时,并且常常很危险,因此只能是有限的。在他们的文章中,引入了解的最小二乘估计。文献 [18] 和 [19] 中,介绍了线性系统的 SVD 方法。其中在 [18] 中,超分辨率问题的解是基于傅立叶矩阵 $\mathcal F$ 的 SVD 分解。而在 [19] 中,作者利用 SVD 分解求解衍射问题的有限带宽信号的重构。并且讨论了带 SVD 的 Miller 正则化方法与最小二乘估计差别。在 [8, 12] 中,提出了傅立叶雷达问题的谱估计算法。而在 [7],提出了 SAR 图像重构的非二次正则化方法。文献 [20] 明确提出了基于复 Fourier 矩阵的截断奇异值分解算法 (TSVD),并将之应用于信号的重构问题中。所有这些应用都包括解一个线性的傅立叶系统。

通过离散傅立叶变换可以重构整个频率谱信息 F。仅仅知道部分的频域和时域信息,我们可以决定时域信号 f 和频率谱 F。很显然 f 可以通过求解线性系统 (3.1.1) 得到。通常用的方法是快速傅立叶变换 FFT。通过 FFT,时域信号转换为频域上的信号,然后运算在频域上进行;最后逆快速傅立叶变换 IFFT 将频域信号转换为时域上的信号。但有时直接计算更加合适,而不必要进行时域和频域的转换。而且,标准的 FFT 只有当在抽样点的个数为 2^N 时才很快。否则,虽然 FFT 仍能使用,但速度并不理想,仅仅象 DFT。更进一步,有时系统可能是超定的 [10,11,13,15],因此 FFT 不能直接应用。还有傅立叶变换对于干涉成像的强度,相位有一些缺陷,如见 [8]。这样我们必须借助于其它方法。直接方法,如 [18,19,20] 中提出的 SVD 分解和本章我们要提出的方法就是一个很好的选择。

我们注意到傅立叶矩阵 \mathcal{F} 可以写成:

$$\mathcal{F} = \mathcal{F}_r + i\mathcal{F}_i,$$

其中 \mathcal{F}_r 和 \mathcal{F}_i 分别代表 \mathcal{F} 的实部和虚部。它们之间的差别就在于差一个角度位移 $\frac{\pi}{2}$ 。对于时间有限的信号,我们可以简单的通过 DFT 系统的实部来重构 (这更像余弦变换 [14, 16])。从这个意义上,输入信号 f 可以通过

$$f = \mathcal{F}_r^{-1} F. \tag{3.1.5}$$

重构。现在的问题是 \mathcal{F}_r 的条件数如何?它的逆是有界还是无界的?如果 \mathcal{F}_r 的条件数很大,而且由于频率信号 F 含有随机噪声,这就是一个离散的不适定系统,让我考虑 10×10

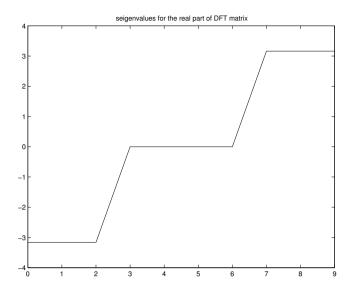


图 3.1: DFT 矩阵实部的特征值分布

的 DFT 矩阵: F_{10×10}:

$$\mathcal{F}_{10\times 10} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & w_N & w_N^2 & \cdots & w_N^9 \\ 1 & w_N^2 & w_N^4 & \cdots & w_N^{18} \\ 1 & w_N^6 & w_N^9 & \cdots & w_N^{27} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & w_N^9 & w_N^{18} & \cdots & w_N^{81} \end{bmatrix}.$$

其中 $w_N^{kn} = \exp(-i(2\pi kn/N)), N = 10$ 。从计算可知, \mathcal{F}_r 最大的和最小的奇异值分别为:

$$\sigma_{\text{max}}(\mathcal{F}_r) = 3.1623, \ \sigma_{\text{min}}(\mathcal{F}_r) = 7.1289 \times 10^{-19}.$$

因此 \mathcal{F}_r 的条件数非常大,达到 4.4359×10^{18} ,该矩阵高度病态。图 3.1 画出了 \mathcal{F}_r 的特征值分布。从图中,我们可以看出它的非零特征值分布在 x 轴两侧,中间部分的部分非常小,接近于零。

但是由于傅立叶矩阵的特殊结构, \mathcal{F}_r ,和 \mathcal{F}_i 有非常好的性质。根据这些性质,我们可以构造非常有效的数值算法,即截断奇异值分解算法(TSVD)来恢复信息。从后面我们的分析可以看出,计算的代价非常小。

§3.2 张量积

文献 [20,54] 列出了很多关于张量积的性质,我们这里对其只作简单的介绍。

定义3.2.1. 设 $A_{m_1 \times n_1}, A = (a_{ij}), B_{m_2 \times n_2}, B = (b_{ij}),$ 则矩阵 A 和 B 的张量积定义为:

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n_1}B \\ a_{21}B & a_{22}B & \cdots & a_{2n_1}B \\ \vdots & \vdots & \vdots & \vdots \\ a_{m_11}B & a_{m_12}B & \cdots & a_{m_1n_1}B \end{pmatrix}$$

定义3.2.2. 给定一个矩阵 $V \in \mathbb{C}^{n_x \times n_y}$, 我们可以按列叠加(或者说对列向量进行字典式排序)得到一个向量 $v \in \mathbb{C}^{n_x n_y}$, 即定义算子 $vec : \mathbb{C}^{n_x \times n_y} \to \mathbb{C}^{n_x n_y}$:

$$vec(V) = (V_{1,1}, \dots, V_{n_x,1}, V_{1,2}, \dots, V_{n_x,2}, \dots, V_{1,n_y}, \dots, V_{n_x,n_y}) , \qquad (3.2.1)$$

记 array 为 vec 的逆算子, 且满足:

$$array(vec(V)) = V$$
 , $vec(array(v)) = v$.

张量积在数字信号处理和数字图像处理中起着很重要的作用,我们下面给出张量积的一下性质,这些性质可以直接验证,我们这里略去不证。

性质3.2.3. $(A \otimes B) \otimes C = A \otimes (B \otimes C)$

性质3.2.4. $(A+B)\otimes (C+D)=A\otimes C+B\otimes C+A\otimes D+B\otimes D$

性质3.2.5. $(A \otimes B)^T = A^T \otimes B^T$

性质3.2.6. $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$

性质3.2.7. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$

性质3.2.8. $(A \otimes B)vec(X) = vec(BXA^T)$

下面我们再给出张量积的几个基本定理, 其证明都可以根据相关定义得到验证。

定理3.2.9. 对于矩阵 A,B, 我们有下面关系式成立:

- 1. 若 A, B 为对角(下或上三角)阵,则 $A \otimes B$ 亦为对角(下或上三角)阵
- 2. 若 A,B 为带状矩阵,则 $A \otimes B$ 亦为带状矩阵
- 3. 若 A.B 为对称矩阵,则 $A \otimes B$ 亦为对称矩阵
- 4. 若 A,B 为奇异矩阵,则 $A \otimes B$ 亦为奇异矩阵
- 5. 若 A,B 均正定,则 $A \otimes B$ 亦正定
- 6. 若 A.B 为正交矩阵,则 $A \otimes B$ 亦为正交矩阵
- 7. 若 A,B 为 Toeplitz 矩阵或循环矩阵,则 $A \otimes B$ 亦为 Toeplitz 矩阵或循环矩阵

定理3.2.10. 矩阵 A 和 B 的张量积的模、迹、秩、幂、谱和行列式满足以下关系式:

- 1. $||A \otimes B||_p = ||A||_p ||B||_p$, $p = 1, 2, \infty, F$
- 2. 若 A, B 均为方阵,则对迹有: $Tr(A \otimes B) = Tr(B \otimes A) = Tr(A) \cdot Tr(B)$

- 3. 若 A,B 的秩分别为 r(A),r(B) ,则 $r(A\otimes B)=r(A)\cdot r(B)$
- 4. 若 A, B 均为方阵,则对幂有 $(A \otimes B)^n = A^n \otimes B^n$
- 5. 若 A,B 均为方阵,则 $A\otimes B$ 的特征值为对应的 A 的特征值 λ_A 和 B 的特征值 λ_B 的 乘积 $\lambda_A\lambda_B$; $A\otimes B$ 的特征向量为对应的 A 的特征向量 v_A 和 B 的特征向量 v_B 的张量积 $v_A\otimes v_B$ 。

$$det(A \otimes B) = det(A)^m \cdot det(B)^n$$

7. 若 A, B 为均为方阵,则 $cond(A \otimes B) = cond(A) \cdot cond(B)$

§3.3 离散傅立叶矩阵 DFT 的 SVD 分解

令

$$\mathcal{F} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega_{N} & \omega_{N}^{2} & \cdots & \omega_{N}^{N-1} \\ 1 & \omega_{N}^{2} & \omega_{N}^{4} & \cdots & \omega_{N}^{2(N-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \omega_{N}^{N-1} & \omega_{N}^{2(N-1)} & \cdots & \omega_{N}^{(N-1)^{2}} \end{pmatrix} \in \mathbb{C}^{N \times N}, \tag{3.3.1}$$

其中 $\omega_N=e^{-i\frac{2\pi}{N}}$ 。显然, $\mathcal F$ 的不同元素仅为 $1,x,x^2,\cdots,x^{N-1}$,其中 $x=1,\omega_N,\cdots,\omega_N^{N-1}$,并且 $\mathcal F$ 满足:

$$\mathcal{F}^T = \mathcal{F} \quad \text{and} \quad \mathcal{F}^* \mathcal{F} = \mathcal{F} \mathcal{F}^* = NI_N.$$
 (3.3.2)

在得到我们的主要结论之前,我们需要以下几个引理:

引理3.3.1. 设 $\mathcal{F} \in \mathbb{C}^{N \times N}$ 是傅立叶矩阵,记 $\mathcal{K}_r := \mathcal{F}_r^T \mathcal{F}_r, \mathcal{K}_i := \mathcal{F}_i^T \mathcal{F}_i,$ 则

- (1) \mathcal{F} 是对称矩阵, 即 $\mathcal{F}^T = \mathcal{F}$;
- (2) \mathcal{F} 不同的元素仅为 1, x, x^2 , \dots , x^{N-1} , 其中 $x = \omega$, ω^2 , \dots , ω^{N-1} ;

(3) 设 N 偶数,则

$$\mathcal{K}_{T} = \begin{pmatrix}
N & 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 & 0 \\
0 & \frac{N}{2} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \frac{N}{2} \\
0 & 0 & \frac{N}{2} & \cdots & \cdots & \cdots & \cdots & \frac{N}{2} & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & \frac{N}{2} & 0 & \frac{N}{2} & \cdots & 0 & 0 \\
0 & 0 & 0 & \cdots & 0 & N & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & \cdots & \frac{N}{2} & 0 & \frac{N}{2} & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \frac{N}{2} & \cdots & \cdots & \cdots & \cdots & \frac{N}{2} & 0 \\
0 & \frac{N}{2} & 0 & \cdots & \cdots & \cdots & \cdots & 0 & \frac{N}{2}
\end{pmatrix},$$
(3.3.3)

并且,

$$\mathcal{K}_{i} = \begin{pmatrix}
0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 0 \\
0 & \frac{N}{2} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & -\frac{N}{2} \\
0 & 0 & \frac{N}{2} & \cdots & \cdots & \cdots & \cdots & \cdots & -\frac{N}{2} & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & \frac{N}{2} & 0 & -\frac{N}{2} & \cdots & 0 & 0 \\
0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & \cdots & -\frac{N}{2} & 0 & \frac{N}{2} & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & -\frac{N}{2} & \cdots & \cdots & \cdots & \cdots & \frac{N}{2} & 0 \\
0 & -\frac{N}{2} & 0 & \cdots & \cdots & \cdots & \cdots & 0 & \frac{N}{2}
\end{pmatrix} .$$
(3.3.4)

(4) 设 N 是奇数,则

$$\mathcal{K}_{r} = \begin{pmatrix}
N & 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 & 0 \\
0 & \frac{N}{2} & 0 & \cdots & \cdots & \cdots & \cdots & 0 & \frac{N}{2} \\
0 & 0 & \frac{N}{2} & \cdots & \cdots & \cdots & \cdots & \frac{N}{2} & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & \frac{N}{2} & \frac{N}{2} & \cdots & 0 & 0 \\
0 & 0 & 0 & \cdots & \frac{N}{2} & \frac{N}{2} & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \frac{N}{2} & \cdots & \cdots & \cdots & \frac{N}{2} & 0 \\
0 & \frac{N}{2} & 0 & \cdots & \cdots & \cdots & 0 & \frac{N}{2}
\end{pmatrix} ,$$
(3.3.5)

并且

$$\mathcal{K}_{i} = \begin{pmatrix}
0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 & 0 \\
0 & \frac{N}{2} & 0 & \cdots & \cdots & \cdots & \cdots & 0 & -\frac{N}{2} \\
0 & 0 & \frac{N}{2} & \cdots & \cdots & \cdots & -\frac{N}{2} & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & \frac{N}{2} & -\frac{N}{2} & \cdots & 0 & 0 \\
0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & \cdots & -\frac{N}{2} & \frac{N}{2} & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & -\frac{N}{2} & \cdots & \cdots & \cdots & \frac{N}{2} & 0 \\
0 & -\frac{N}{2} & 0 & \cdots & \cdots & \cdots & 0 & \frac{N}{2}
\end{pmatrix} .$$
(3.3.6)

(5) F 是正规矩阵,即 $F^*F = FF^*$,特别的 $F^*F = N \cdot I$,其中 I 是 N 阶的单位矩阵。证明. (1), (2) 和 (5) 非常明显,这里我们主要证明 (3) 和 (4)。

考虑一个单位圆周划分为 N 个等距节点: $\theta_l = \frac{2kl\pi}{N}$, $l = 0, 2, \dots, N-1$ 。注意到 $\cos \theta_l$ 是到 x 轴的投影, $\sin \theta_l$ 是到 y 轴的投影,因此下面的等式可以很容易的得到:

$$\sum_{l=0}^{N-1} \cos \frac{2kl\pi}{N} \cos \frac{2k'l\pi}{N} = \sum_{l=0}^{N-1} \sin \frac{2kl\pi}{N} \sin \frac{2k'l\pi}{N} = 0$$
 (3.3.7)

对于所有的整数 k, k', 且 $k \pm k' \notin N\mathbf{Z}$ 成立,

$$\sum_{l=0}^{N-1} \cos^2 \frac{2kl\pi}{N} = \sum_{l=0}^{N-1} \sin^2 \frac{2kl\pi}{N} = \frac{N}{2}$$
 (3.3.8)

对于所有的整数 $k \in \mathbf{Z}$,且 $2k \notin N\mathbf{Z}$ 成立,

$$\sum_{l=0}^{N-1} \cos 2k l \pi N \sin \frac{2k' l \pi}{N} = 0 \tag{3.3.9}$$

对于所有的整数 k, k' 成立。这些关系直接导出了 (3) 和 (4) 成立。 □ **引理3.3.2.** \mathcal{K}_r 和 \mathcal{K}_i 的特征值分解为:

$$\mathcal{K}_r = V^T \Lambda_1 V, \qquad \mathcal{K}_i = V^T \Lambda_2 V,$$

其中 $\Lambda_1 = diag(\underbrace{N, \cdots, N}_r, 0, \cdots, 0), \Lambda_2 = diag(0, \cdots, 0, \underbrace{N, \cdots, N}_{N-r})$ 。矩阵 V 是正交的,即 $V^TV = I$

并且当 N 为偶数时为以下形式:

$$V = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 & 0 \\ 0 & \frac{\sqrt{2}}{2} & 0 & \cdots & \cdots & \cdots & \cdots & 0 & \frac{\sqrt{2}}{2} \\ 0 & 0 & \frac{\sqrt{2}}{2} & \cdots & \cdots & \cdots & \cdots & \frac{\sqrt{2}}{2} & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \frac{\sqrt{2}}{2} & \cdots & \cdots & \cdots & \cdots & -\frac{\sqrt{2}}{2} & 0 \\ 0 & \frac{\sqrt{2}}{2} & 0 & \cdots & \cdots & \cdots & \cdots & 0 & -\frac{\sqrt{2}}{2} \end{pmatrix}$$

当 N 为奇数时为以下形式:

$$V = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 & 0 \\ 0 & \frac{\sqrt{2}}{2} & 0 & \cdots & \cdots & \cdots & \cdots & 0 & \frac{\sqrt{2}}{2} \\ 0 & 0 & \frac{\sqrt{2}}{2} & \cdots & \cdots & \cdots & \frac{\sqrt{2}}{2} & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \frac{\sqrt{2}}{2} & \cdots & \cdots & \cdots & \cdots & -\frac{\sqrt{2}}{2} & 0 \\ 0 & \frac{\sqrt{2}}{2} & 0 & \cdots & \cdots & \cdots & \cdots & 0 & -\frac{\sqrt{2}}{2} \end{pmatrix}$$

$$(3.3.11)$$

并且当 N 为偶数时 r = (N+2)/2, 当 N 为奇数时 r = (N+1)/2.

证明. 我们仅考虑偶数阶的傅立叶矩阵。记 det(·) 为给定矩阵的行列式, 我们得到:

$$\det(\mathcal{K}_r - \lambda I) = (-\lambda)^{\frac{N}{2} - 1} (N - \lambda)^{\frac{N}{2} + 1}, \tag{3.3.12}$$

因此 K_r 的不同特征值仅为 N 和 0。而且可以很容易的构造出正交的特征向量 (3.3.10) \Box 下面我们证明傅立叶矩阵 \mathcal{F} 实部和虚部的 SVD 分解。

定理3.3.3. 令 $\mathcal{F} \in \mathbb{C}^{N \times N}$ 为式 (3.3.1) 给出的傅立叶矩阵, \mathcal{F}_r 和 \mathcal{F}_i 分别是 \mathcal{F} 实部和虚部。对于所有的整数 k,令 $a_k = \cos 2k\pi/N$ 和 $b_k = \sin 2k\pi/N$ 。当 N 为偶数时,r = (N+2)/2,V

由式 (3.3.10) 给出, 并且定义

$$U = \sqrt{\frac{1}{N}} \begin{pmatrix} 1 & \sqrt{2} & \cdots & \sqrt{2} & 1 & 0 & \cdots & 0 \\ 1 & \sqrt{2}a_1 & \cdots & \sqrt{2}a_{r-2} & a_{r-1} & \sqrt{2}b_{r-2} & \cdots & \sqrt{2}b_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \sqrt{2}a_{N-1} & \cdots & \sqrt{2}a_{(N-1)(r-2)} & a_{(r-1)(N-1)} & \sqrt{2}b_{(N-1)(r-2)} & \cdots & \sqrt{2}b_{N-1} \end{pmatrix},$$
(3.3.13)

当 N 为奇数时, r = (N+1)/2, V 由式 (3.3.11) 给出, 并且定义

$$U = \sqrt{\frac{1}{N}} \begin{pmatrix} 1 & \sqrt{2} & \cdots & \sqrt{2} & 0 & \cdots & 0 \\ 1 & \sqrt{2}a_1 & \cdots & \sqrt{2}a_{r-1} & \sqrt{2}b_{r-1} & \cdots & \sqrt{2}b_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & \sqrt{2}a_{N-1} & \cdots & \sqrt{2}a_{(N-1)(r-1)} & \sqrt{2}b_{(N-1)(r-1)} & \cdots & \sqrt{2}b_{N-1} \end{pmatrix},$$
(3.3.14)

则 \mathcal{F}_r 和 \mathcal{F}_i 的奇异值分解如下:

$$\mathcal{F}_r = U\Sigma_1 V^T \quad , \quad \mathcal{F}_i = U\Sigma_2 V^T, \tag{3.3.15}$$

其中

$$\Sigma_1 = \operatorname{diag}(\underbrace{\sqrt{N}, \cdots, \sqrt{N}}_{r}, 0, \cdots, 0)$$
 , $\Sigma_2 = \operatorname{diag}(0, \cdots, 0, \underbrace{\sqrt{N}, \cdots, \sqrt{N}}_{r}).$

证明. 令 $\mathcal{F}_r = U\Sigma_1 V^T$,则 $\mathcal{F}_r^T \mathcal{F}_r = V\Sigma_1^2 V^T$ 是其特征值分解。从引理 3.3.2 中,我们可以看出

$$\Sigma_1 = diag(\underbrace{\sqrt{N}, \cdots, \sqrt{N}}_{\bullet}, 0, \cdots, 0),$$

并且很显然,V 可以分别取为 (3.3.10) 或 (3.3.11)。一旦 V 确定了,U 的前 r 列由于奇异值 非零也就从关系式 $\mathcal{F}_r V = U\Sigma_1$ 确定了,余下的列取为与前 r 列正交的向量就行了。注意 到 $\mathcal{F}_r^T \mathcal{F}_i = 0$,因此就得到了如 (3.3.13) 或 (3.3.14) 所示的 U。 \square

有意思的是,如果记 $\Sigma = \Sigma_1 + i\Sigma_2$,从定理 3.3.3 的结论中我们可以写出矩阵 \mathcal{F} 的乘积 形式:

$$\mathcal{F} = U\Sigma V^T \quad .$$

我们的目标是仅仅利用矩阵 \mathcal{F}_r 就重构出时间有限信号。但定理 3.3.3 却告诉我们 \mathcal{F}_r 是一个亏秩的实矩阵,它的秩等于 \mathcal{F}_r 严格正的奇异值的个数。这就给计算带来了很大的困难,即 \mathcal{F}_r 很小的扰动就能给解带来很大的不稳定。这就需要特殊的技巧来克服。对于秩亏问题,我们在第二章第四节有过比较详细叙述,由于 \mathcal{F}_r 的 SVD 分解的特殊性,很容易可以将截断 SVD 分解(TSVD)用来解这个问题。

§3.4 TSVD 在信号处理中的应用

§3.4.1 一维问题

在实际问题中,频率信息通常含有噪声,即实际上有的信息只是 $F = F_{true} + E$,其中 E是加法型噪声。如果我们想重构出信号 f,我们需要解如下优化问题:

$$\min \|E\|_2$$
 ,

即相当于求解最小二乘问题:

$$\min J(f) \stackrel{def}{=} \min \|\mathcal{F}_r f - F\|_2 \quad . \tag{3.4.1}$$

由定理 3.3.3,可得 $\mathcal{F}_r = U\Sigma_1 V^T$ 是矩阵 \mathcal{F}_r 的 SVD 分解,其中 $U = [u_1, u_2, \cdots, u_N]$ 和 $V = [v_1, v_2, \cdots, v_N]$ 按列排列。这样我们就可以将第二章第四节的截断 SVD 分解(TSVD)直接应用到这个问题,只需虑掉所有为零的奇异值,即有:

$$f = \mathcal{F}_r^{\dagger} F = \sum_{i=1}^r \frac{u_i^T F}{\sigma_i} v_i \quad .$$

我们把具体的算法描述如下:

算法3.4.1. 一维信号重构的最小二乘算法

步 1: 如果 N 是偶数,令 $r = \frac{N}{2} + 1$, $V = (v_1, v_2, \dots, v_r)$,构造 $U = \mathcal{F}_r(:, 1:r)$:

构造 $U = \mathcal{F}_r(:,1:r);$

 $U(:,2:r-1) = \sqrt{2}U(:,2:r-1);$ 转步 3;

步 2: 如果 N 是奇数,令 $r=\frac{N+1}{2}$, $V=(v_1,v_2,\cdots,v_r)$,

构造 $U = \mathcal{F}_r(:, 1:r)$;

 $U(:,2:r) = \sqrt{2}U(:,2:r);$ 转步 3;

步 3: $f_{LS} = \sum_{i=1}^{r} \frac{u_i^T F}{N} v_i$.

在算法 (3.4.1) 中,记号 $\mathcal{F}_r(:,1:r)$ 表示选择矩阵 \mathcal{F}_r 的 1 到 r 列形成的子矩阵,其它类似。

§3.4.2 二维问题

对于超定的二维问题, 计算过程比一维问题稍微复杂一点。然而, 利用张量积的性质, 我们仍然可以得到重构信号的简单方法。记

$$\mathcal{F}_x = A_x + B_x i, \qquad \mathcal{F}_y = A_y + B_y i,$$

另记 $N_x = dim(\mathcal{F}_x), \ N_y = dim(\mathcal{F}_y), \ r_x = rank(A_x), \ r_y = rank(A_y)$,因此有:

$$\mathcal{F}_x \otimes \mathcal{F}_y = (A_x + B_x i) \otimes (A_y + B_y i)$$
$$= (A_x \otimes A_y - B_x \otimes B_y) + (A_x \otimes B_y - B_x \otimes A_y)i.$$

二维的信号重构问题有如下形式:

$$\min ||\mathcal{K}\mathbf{f} - \mathbf{F}||_2, \tag{3.4.2}$$

其中 $\mathcal{K} = real(\mathcal{F}_x \otimes \mathcal{F}_y)$, $\mathbf{f} = vec(f)$, $\mathbf{F} = vec(F)$, 从定理 3.3.3, 可以得:

$$A_x = U_x \begin{pmatrix} \Sigma_{1x} & 0 \\ 0 & 0 \end{pmatrix} V_x^T \qquad A_y = U_y \begin{pmatrix} \Sigma_{1y} & 0 \\ 0 & 0 \end{pmatrix} V_y^T$$

$$B_x = U_x \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{2x} \end{pmatrix} V_x^T \qquad B_y = U_y \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{2y} \end{pmatrix} V_y^T$$

其中 $\Sigma_{1x} = \sqrt{N_x} I_{r_x}, \Sigma_{2x} = \sqrt{N_x} I_{N_x - r_x}, \Sigma_{1y} = \sqrt{N_y} I_{r_y}, \Sigma_{2y} = \sqrt{N_y} I_{N_y - r_y}$ 。因此

$$\mathcal{K} = A_x \otimes A_y - B_x \otimes B_y$$
$$= (U_x \otimes U_y) \Sigma (V_x \otimes V_y)^T. \tag{3.4.3}$$

其中

$$\Sigma = \begin{pmatrix} \Sigma_{1x} & 0 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} \Sigma_{1y} & 0 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{2x} \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{2y} \end{pmatrix} .$$

从前面的讨论中最小二乘解可以表示为:

$$\mathbf{f}_{LS} = \mathcal{K}^{\dagger} \mathbf{F} = (V_x \otimes V_y) \Sigma^{\dagger} (U_x \otimes U_y)^T \mathbf{F}, \tag{3.4.4}$$

因为 Σ 是对角矩阵, 则 Σ^{\dagger} 为:

$$\Sigma^{\dagger} = \begin{pmatrix} \Sigma_{1x}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} \Sigma_{1y}^{-1} & 0 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{2x}^{-1} \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{2y}^{-1} \end{pmatrix} \quad .$$

由张量积的定义, 我们展开式 (3.4.4)) 有:

$$(U_x \otimes U_y)^T \mathbf{F} = vec(U_y^T F U_x) = vec(\mathbf{F}'_n) \quad ,$$

$$\Sigma^{\dagger} vec(\mathbf{F}'_{n}) = vec \left(\begin{pmatrix} \Sigma_{1y}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \mathbf{F}'_{n} \begin{pmatrix} \Sigma_{1x}^{-1} & 0 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{2y}^{-1} \end{pmatrix} \mathbf{F}'_{n} \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{2x}^{-1} \end{pmatrix} \right)$$

$$= vec \left(\begin{pmatrix} \mathbf{F}''_{1n} & 0 \\ 0 & -\mathbf{F}''_{2n} \end{pmatrix} \right) = vec(\mathbf{F}''_{n}), \tag{3.4.5}$$

其中

$$\mathbf{F}_{1n}^{"} = U_{1y}^T F U_{1x} / \sqrt{N_x N_y}, \qquad \mathbf{F}_{2n}^{"} = U_{2y}^T F U_{2x} / \sqrt{N_x N_y}.$$

这里 U_{1x} 是 U_x 的前 r_x 列, U_{2x} 是 U_x 的后 $N_x - r_x$ 列,对于 U_{1y} , U_{2y} , V_{1x} , V_{2x} , V_{1y} , V_{2y} 有类似的定义。

同时我们注意到

$$(V_x \otimes V_y)vec(\mathbf{F}_n'') = vec \left(V_y \begin{pmatrix} \mathbf{F}_{1n}'' & 0 \\ 0 & -\mathbf{F}_{2n}'' \end{pmatrix} V_x^T \right)$$

$$= vec \left((V_{1y} \quad V_{2y}) \begin{pmatrix} \mathbf{F}_{1n}'' & 0 \\ 0 & -\mathbf{F}_{2n}'' \end{pmatrix} (V_{1x} \quad V_{2x})^T \right)$$

$$= vec(V_{1y}\mathbf{F}_{1n}''V_{1x}^T - V_{2y}\mathbf{F}_{2n}'V_{2x}^T) . \tag{3.4.6}$$

这样我们就得到了最小二乘问题的解:

$$f_{LS} = vec(V_{1y}\mathbf{F}_{1n}^{"}V_{1x}^{T} - V_{2y}\mathbf{F}_{2n}^{"}V_{2x}^{T}).$$
(3.4.7)

至此,我们就可以构造二维信号重构的一个非常有效的算法:

算法3.4.2. 二维信号重构的最小二乘算法

根据定理 3.3.3 构造 $U_{x}, V_{x}, U_{y}, V_{y}$;

$$\mathbf{F}_{1n}^{"} = U_{1n}^{T} F U_{1x} / \sqrt{N_x N_y};$$

$$\mathbf{F}_{2n}^{"} = U_{2n}^T F U_{2n} / \sqrt{N_x N_y};$$

$$\mathbf{f}_{LS} = vec(V_{1y}\mathbf{F}_{1n}^{"}V_{1x}^{T} - V_{2y}\mathbf{F}_{2n}^{"}V_{2x}^{T}).$$

§3.5 数值算例

在这一节中,我们给出七个测试算例。这些算例包括一维和二维的信号重构及图像恢复。

例3.5.1. 一维双脉冲信号重构

这是文献[9]中考虑的信号重构问题。输入信号由:

$$f(x) = 1000\left[\text{rect}\left(16\left(x - \frac{3.5}{32}\right)\right) + \text{rect}\left(16\left(x + \frac{3.5}{32}\right)\right)\right],\tag{3.5.1}$$

给出。这个信号由两个 $\frac{1}{16}$ 单位长度的脉冲组成,对称分布在原点的两侧,并相距 $\frac{7}{32}$ 单位长度。脉冲为 1000 单位高。在数值上,我们在区间 [-4,4] 定义信号 f(t) 。函数 f 可由一个抽样点为 N=500 的离散向量 \vec{f} 来近似。我们记 \vec{f} 是函数 f 的离散化,其它记号类似。

在我们的测试中,频率信息通过如下方式产生: 首先,令时域信号 f 通过傅立叶系统 \mathcal{F} ,这样我们得到频率信息 F: 然后给 F 添加 Gaussian 噪声:

$$\vec{F}_n = \vec{F} + \delta \cdot \text{randn}(\text{size}(\vec{F})),$$

其中 randn(·) 代表 Gaussian 噪声函数,size(·) 表示 \vec{F}_n 和 \vec{F} 有相同的维数。 δ 是噪声水平,在数值试验中,我们选择 $\delta=0.5$ 。以这种方式, \vec{F} 虽然被噪声扰乱但不丢失其它信息。为了模拟实际应用,我们假定部分频率信息缺失。在计算中,这些部分用零来代替,如 令 $\vec{F}_n(1)=0$, $\vec{F}_n(N/5)=0$, $\vec{F}_n(N/4+1:N)=0$ 。这样,频率信息被严重破坏了。图 3.2 画出了真实的一维信号和它的频率信息。图 3.3 的左边画出带噪声,残缺的频率信息;图 3.3 的右边,原信号用实线画出,重构的信号用点线画出。我们发现重构更多的依赖于频率的品质。

同时从图 3.4 我们发现,噪声不太影响重构的效果,我们发现如果原信号只被噪声影响,重构的信号近乎完美。这个现象的原因是噪声被截断 SVD 分解成功的压制住了,即小奇异值不起多大作用。

例3.5.2. 一维四脉冲信号重构

这个例子与例 3.5.1 类似。输入信号为:

$$f(x) = 1000 \left[\operatorname{rect}(16((x+2) - \frac{3.5}{32})) + \operatorname{rect}(16((x+2) + \frac{3.5}{32})) + \operatorname{rect}(16((x-2) - \frac{3.5}{32})) + \operatorname{rect}(16((x-2) + \frac{3.5}{32})) \right],$$
(3.5.2)

很显然,它和例 3.5.1 差别是这里信号有个为 2 的平移。

为了描述实际应用,我们使用例 3.5.1 中同样的模拟方式。首先,我们加误差水平为 $\delta=0.5$ 的 Gaussian 白噪声到 \vec{F} ,然后假定部分频率信息缺失。在计算中,这些缺失部分用零来代替,即:令 $\vec{F}_n(N/5)=0$, $\vec{F}_n(1:N/4+1)=0$ 。这样,频率信息被严重破坏了。图 3.5 画出了真实信号和它的频率谱。图 3.6 画出了带噪声,残缺的频谱信号;图 3.6 的右边,原信号用实线画出,重构的信号用点线画出。我们发现重构也更加依赖于频率的品质。同时从图 3.7 我们发现,噪声也不太影响重构的效果,

例3.5.3. 一维三脉冲 "on-off" 信号重构

这个在输入信号是工程中经常遇到,可以描述为:

$$f(t) = rect(t+2) + rect(t) + rect(t-2)$$

$$(3.5.3)$$

为了描述实际应用,我们使用例 3.5.1 中同样的模拟方式。图 3.8 画出了真实信号和它的频率谱。图 3.9 画出了带噪声,残缺的频谱信号;图 3.9 的右边,原信号用实线画出,重构的信号用点线画出。我们还是发现重构更加依赖于频率的品质。同时从图 3.10 我们发现,噪声也不太影响重构的效果,

例3.5.4. 二维信号重构

在下面测试中,我们考虑二维信号重构问题。输入信号为:

$$f(x,y) = \sqrt{4 - x^2/9 - y^2/4},$$

where $x, y \in [-2, 2]$ 。数值上,函数 f 由 $M_x = N_y = 2^8$ 个抽样点的离散信号 \vec{f} 来近似。

频谱信息如同例 3.5.1 中的模拟方式产生。首先,我们加误差水平为 $\delta=0.5$ 的 Gaussian 白噪声到 \vec{F} ,然后假定部分频率信息缺失。在计算中,这些缺失部分用零来代替,即:令 $\vec{F}_n(1,2)=0$, $\vec{F}_n(N_x/4+1:N_x,N_y/4+1:N_y)=0$ 。这样,频率信息被严重破坏了。为了形象的显示频谱信息,我们将零频率分量平移到谱的中心。图 3.11 画出了真实的二维信号和它的频率谱。图 3.12 画出了带噪声,残缺的频谱信号和重构信号。

例3.5.5. 图像恢复

我们现在考虑图像恢复问题。我们取信号处理文献中经常使用的测试图像"Lena"。频谱信息如同例 3.5.4 中的模拟方式产生。首先,我们加误差水平为 $\delta = 0.5$ 的 Gaussian 白噪声到 \vec{F} ,然后假定部分频率信息缺失。在计算中,这些缺失部分用零来代替,即:令

$$\vec{F}_n(1,2) = 0, \ \vec{F}_n(flr(N_x/4) + 1: flr(N_x), flr(N_y/4) + 1: N_y) = 0,$$

其中 $flr(\cdot)$ 表示取比给定数小的最大整数。这样,频率信息被严重破坏了。图 3.13 画出了真实的二维信号和它的频率谱。图 3.14 画出了带噪声,残缺的频谱信号和重构信号。

例3.5.6. 医学图像恢复

这个测试例子考虑人体肺部的医学图像。所有的模拟过程如同例 3.5.5 除了损失的信息由下式贴换:

$$\vec{F}_n(flr(N_x/16) + 1 : flr(N_x/4), flr(N_y/8) + 1 : flr(N_y/2)) = 0,$$

 $\vec{F}_n(flr(N_x/4) + 1 : flr(N_x/2), flr(N_y/16) + 1 : flr(N_y/8)) = 0.$

3.15 画出了真实的二维信号和它的频率谱。图 3.16 画出了带噪声, 残缺的频谱信号和重构信号。

§3.6 讨论和结论

在这一章中我们提出了通过傅立叶系统恢复信号的截断 SVD 分解方法。这一方法基于对傅立叶矩阵 \mathcal{F} 实部和虚部的特殊结构的研究而得到的。这一方法当系统是超定的时候也适用,因为此时最小二乘解是有意义的。该方法对 [6, 8, 12] 中的雷达问题(radar problems)和医学中的图像处理问题,如磁共振成像(magnetic resonance imaging (MRI)),

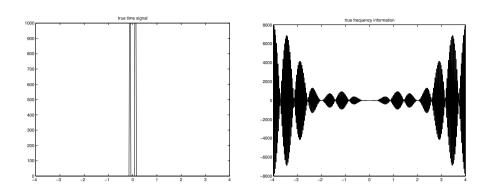


图 3.2: 测试 1) 左: 输入信号; 右: 精确频率信息

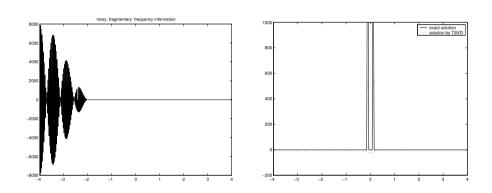


图 3.3: 测试 1) 左: 带噪声, 残缺的频谱信号; 右: 重构信号

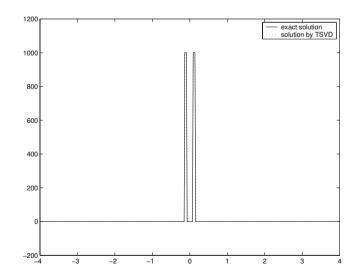


图 3.4: 测试 1) 仅带噪声的重构信号

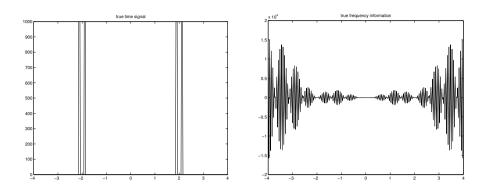


图 3.5: 测试 2) 左: 输入信号; 右: 精确频率信息

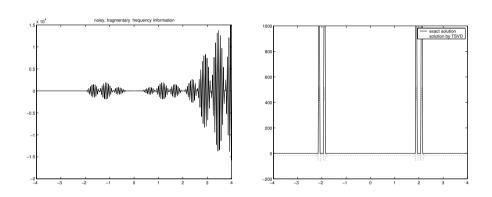


图 3.6: 测试 2) 左: 带噪声, 残缺的频谱信号; 右: 重构信号

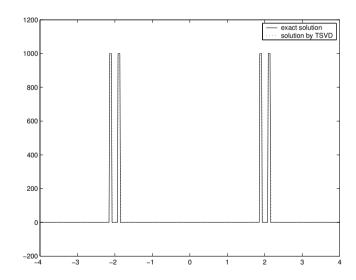


图 3.7: 测试 2) 仅带噪声的重构信号

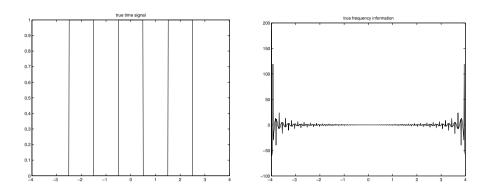


图 3.8: 测试 3) 左: 输入信号; 右: 精确频率信息

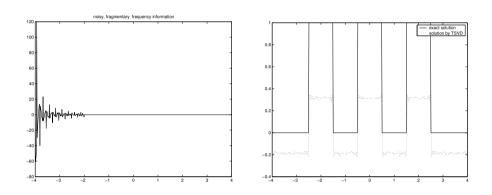


图 3.9: 测试 3) 左: 带噪声, 残缺的频谱信号; 右: 重构信号

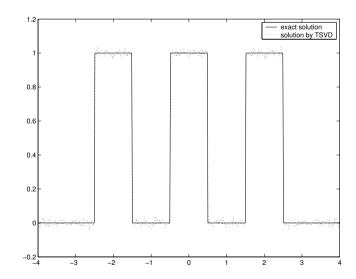


图 3.10: 测试 3) 仅带噪声的重构信号

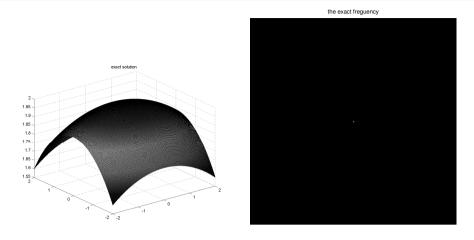


图 3.11: 测试 4) 左: 输入信号; 右: 精确频率信息

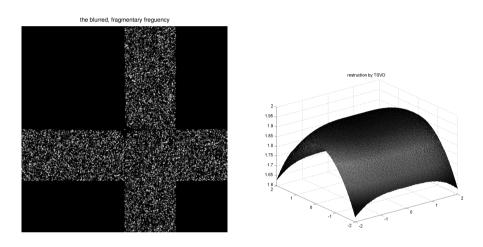


图 3.12: 测试 4) 左: 带噪声, 残缺的频谱信号; 右: 重构信号



图 3.13: 测试 5) 左: 输入信号; 右: 精确频率信息

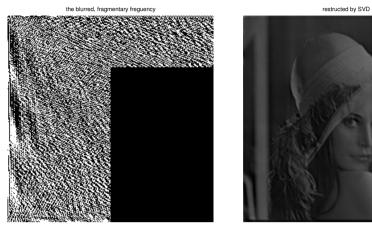




图 3.14: 测试 5) 左: 带噪声, 残缺的频谱信号; 右: 重构图像



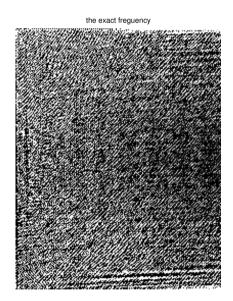
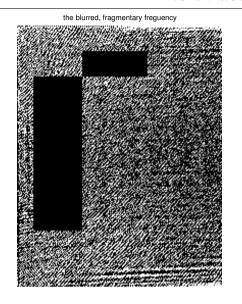


图 3.15: 测试 6) 左: 输入信号; 右: 精确频率信息



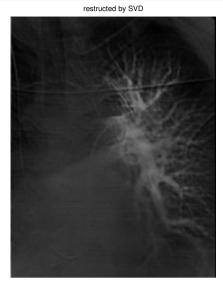


图 3.16: 测试 6) 左: 带噪声, 残缺的频谱信号; 右: 重构图像

计算 X 线断层摄影术 (computerized tomography), 磁共振分光镜成像 (magnetic resonance spectroscopic imaging[10]) 等等也同样很有帮助。

我们的算法对不等距节点 F_k , $k=1,2,\cdots$ 同样也适用。在这种情况下,为了恢复 f,最简单的办法是先将右端项 F 插值形成等距节点 \bar{F} ,然后用我们的算法针对 \bar{F} 求解,最后从 \bar{F} 中提取出对应于原信号 F 的点即是要的恢复信号。

第四章 图像恢复的信赖域方法

数字图像的恢复(Image Restoration)或者重构(Image Reconstruction)是图像信息处理中的重要内容。图像恢复指的是去除或减轻在获取数字图像处理过程中发生的图像质量下降。造成该现象的原因可能是由光学系统(如聚焦不准(Out of Foucus)),运动(Motion)等造成的图像模糊,也可能是由电路和广度学因素的噪声,对于遥感图像质量的下降,还可能是由大气湍流等随机因素造成的。这类问题通常表征为离散或连续的卷积型方程,具有特殊的结构,但同时具有典型的不适定性,图像恢复的目的是对模糊的图像进行处理,使它趋向于复原或没有噪声影响的理想图像。在这一章我们将以遥感图像为例集中讨论数字图像恢复的一些技巧。

§4.1 引言

在遥感和天文图像中,图像经常由于大气湍流和噪声的影响而失真,甚至被破坏。描述这一过程的向前模型 (forward model) 通常有如下形式:

$$(Kf)(x,y) := \int_{a}^{b} \int_{c}^{d} k(x-\xi,y-\eta)f(\xi,\eta)d\xi d\eta = h(x,y). \tag{4.1.1}$$

其中 h 是观测到的图像,k 是点扩散函数(PSF, point spread function,见 [20, 23, 24, 25]),f 为要恢复的真实的图像。(4.1.1) 是一个卷积过程,K 表示卷积算子,由算子理论,K 将空间 D(f) 中的函数 f 映射到空间 D(h) 中的函数 g。图像恢复的目的是要找一逆算子 K^{-1} 使得:

$$K^{-1}h \longrightarrow f$$
 . (4.1.2)

本质上讲,图像恢复是一个不适定问题。这类问题关注逆算子的存在性和唯一性,特别的,关注它的稳定性(如见 [20, 26])。存在性和唯一性通常能容易的解决,而稳定性是需要更多考虑。因为在遥感中,观测图像 h 的一个微小的扰动就能对要得到的真实图像 f 产生显著的扰动。例如,扰动 δ_{out} 可以任意小,但通过变换:

$$K^{-1}(h + \delta_{out}) = f + \delta_{in},$$
 (4.1.3)

 $\delta_{in} >> \delta_{out}$,即 δ_{in} 不能任意小,因此不能被忽略。

在数字图像恢复中,我们经常建立其离散系统来进行处理。对于模型 (4.1.1),它对应的离散系统为线性方程组:

$$\mathcal{K}\mathbf{f} = \mathbf{h},\tag{4.1.4}$$

其中 f 和 h 是从采样对象和图像建立的字典序向量。而矩阵 κ 由采样点扩散函数产生。由于误差总是存在于试验和采样中,我们不能忽略上面模型中的噪声,因此线性方程组可以描述为:

$$\mathcal{K}\mathbf{f} = \mathbf{h} + \mathbf{n},\tag{4.1.5}$$

其中 \mathbf{n} 字典序排列的向量。简单的说矩阵 \mathcal{K} 是 $N^2 \times N^2$ 阶的,通常是一个张量; \mathbf{f} , \mathbf{h} , \mathbf{n} 是 N^2 阶的向量。由于问题 (4.1.1) 的不适定性,问题 (4.1.5) 也是不适定的。因此很难用解 线性方程组的直接方法,如 LU 分解来求解。

我们也可以将 (4.1.5) 描述为最小二乘问题:

$$J_1[\mathbf{f}] := \|\mathcal{K}\mathbf{f} - \mathbf{h}\|^2 \quad . \tag{4.1.6}$$

但正如我们在第二章中提到的,由于系统 (4.1.6) 的不适定性,通常的最小二乘解不能满足要求。

Tikhonov 正则化方法是另外一个流行的可能最小二乘法不稳定性的方法。它求解问题:

$$J_2[\mathbf{f}] := \|\mathcal{K}\mathbf{f} - \mathbf{h}\|^2 + \alpha \|\mathbf{f}\|_L^2, \tag{4.1.7}$$

关于这个方法,我们在第二章第五节中有过详细的叙述, α 称为正则化参数,在正则化方法中起关键作用,但它的选取比较困难,必须考虑噪声和正则化误差之间的平衡。 α 取得太小,问题仍是不稳定;当 α 取得太大,则得到的稳定解不是我们所需要的。

信赖域方法是克服不适定问题 (4.1.6) 的另一个方法。在 [27, 28] 中,利用截断共轭梯度法解信赖域子问题的方法应用到这类问题,取得良好的计算效果。我们在第二章第六节中对该方法也有过详细的叙述。在这一章中,我们将用截断的 Lanczos 方法解信赖域子问题。对于数字图像恢复问题,该方法是快速稳定的。

§4.2 Toeplitz 型与 Circulant 型矩阵

由于表征图像恢复的方程通常是卷积型方程,而这类方程离散化之后矩阵是 Toeplitz 型或 Circulant 型的,仔细研究这些特殊矩阵对实际计算效果有相对大的帮助,这里我们给出关于这些矩阵的定义和一些有用的性质(可参见[20, 21, 54])。

定义4.2.1. 矩阵 $T = [t_i j]$ 称为 Toeplitz 矩阵, 如果满足:

$$t_{ij} = t_{j-i}$$
 , $i, j = 1, \cdots, n$.

具体的, 形如:

$$T = \begin{pmatrix} t_0 & t_{-1} & \cdots & t_{2-n} & t_{1-n} \\ t_1 & t_0 & t_{-1} & \ddots & t_{2-n} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ t_{n-2} & \ddots & t_1 & t_0 & t_{-1} \\ t_{n-1} & t_{n-2} & \cdots & t_1 & t_0 \end{pmatrix}$$

对于任何向量 $f \in \mathbb{R}^n$, 矩阵 - 向量积 Tf 具有如下离散卷积形式:

$$[Tf]_k = [t \star f]_k = \sum_{j=0}^{n-1} t_{k-j} f_j$$
 , $i = 0, \dots, n-1$

其中 $t = (t_{1-n}, \dots, t_{-1}, t_0, t_1, \dots, t_{n-1})^T \in \mathbb{C}^{2n-1}$ 。我们记 T = toeplitz(t)。

定义4.2.2. 矩阵 $C = [c_i j]$ 称为 Circulant 矩阵,如果该矩阵是 Toeplitz 型的并且其行(和列) 首尾末交替循环出现,即具有如下形式:

$$T = \begin{pmatrix} c_0 & c_{n-1} & \cdots & c_2 & c_1 \\ c_1 & c_0 & c_{n-1} & \ddots & c_2 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ c_{n-2} & \ddots & c_1 & c_0 & c_{n-1} \\ c_{n-1} & c_{n-2} & \cdots & c_1 & c_0 \end{pmatrix} .$$

容易看出 $C = toeplitz(c^{ext})$,其中 $c^{ext} = (c_1, \cdots, c_{n-1}, c_0, c_1, \cdots, c_{n-1})^T$ 为对 $c^{ext} = (c_0, c_1, \cdots, c_{n-1})^T$ 以 n 为周期的长度为 2n-1 的延拓。我们记 C = circulant(c)。事实上,c 为矩阵 C 的第一列。

如果 C = circulant(c), 则:

$$C = F_n^* diag(F_n c) F_n$$
,

其中 F_n 为 Fourier 矩阵,由此可以看出 F_nc 的元素为 C 的特征值, F_n^* 的列为相应的特征向量。这意味着形如 y = Cx 的积能以 "FFT 速度" 求解:

$$y = F_n^*((F_n c). * (F_n x)) \quad , \tag{4.2.1}$$

换句话说,三次 DFT 运算和一次向量乘法就能计算一个循环阵与一个向量的乘积。

令 $T \in \mathbb{R}^n$,的 Toeplitz 矩阵, $v \in \mathbb{R}^n$,并记 S = toeplitz(s),其中

$$s = (t_{-1}, t_{-2}, \dots, t_{1-n}, 0, t_{n-1}, \dots, t_2, t_1)^T$$
,

则:

$$C^{ext} \begin{pmatrix} v \\ 0_{n \times 1} \end{pmatrix} = \begin{pmatrix} Tv \\ Sv \end{pmatrix}, 其中, C^{ext} = \begin{pmatrix} T & S \\ S & T \end{pmatrix}$$
 ,

此处 $2n \times 2n$ 阶的块状矩阵 C^{ext} 称为矩阵 T 的 circulant 延拓,且可以表示为: $C^{ext} = circulant(c^{ext})$,其中:

$$c^{ext} = (t_0, t_1, \dots, t_{n-1}, 0, t_{1-n}, \dots, t_{-1})^T \in \mathbb{C}^{2n}$$

通过将一个 Toeplitz 矩阵嵌入到了一个 circulant 矩阵,就可以将计算乘积 y = Tx 也能以 "FFT 速度"进行。具体方式是首先将 x 延拓为 $x^{ext} = [x^T \quad 0_{1\times n}]^T$,然后用 (4.2.1) 求解 $y^{ext} = C^{ext}x^{ext}$,取 y^{ext} 的前 n 个元素就是要求的 y。

接下来我们给出二维情形下的 Toeplitz 矩阵和 circulant 矩阵的一些性质。

定义4.2.3. 称一 $n_x n_y \times n_x n_y$ 维的矩阵 T 为带 Toeplitz 块的 Toeplitz 块矩阵 (BTTB),如果 T 具有下述块状结构:

$$T = \begin{pmatrix} T_0 & T_{-1} & \cdots & T_{2-n_y} & T_{1-n_y} \\ T_1 & T_0 & T_{-1} & \ddots & T_{2-n_y} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ T_{n_y-2} & \ddots & T_1 & T_0 & T_{-1} \\ T_{n_y-1} & T_{n_y-2} & \cdots & T_1 & T_0 \end{pmatrix}$$

其中块 T_i 为 $n_x \times n_y$ 维的 Toeplitz 矩阵。

定义4.2.4. 给定一个矩阵 $V \in \mathbb{C}^{n_x \times n_y}$, 我们可以按列叠加(或者说对列向量进行字典式排序)得到一个向量 $v \in \mathbb{C}^{n_x n_y}$, 即定义算子 $vec: \mathbb{C}^{n_x \times n_y} \to \mathbb{C}^{n_x n_y}$:

$$vec(V) = (V_{1,1}, \dots, V_{n_x,1}, V_{1,2}, \dots, V_{n_x,2}, \dots, V_{1,n_y}, \dots, V_{n_x,n_y})$$
 (4.2.2)

记 array 为 vec 的逆算子, 且满足:

$$array(vec(V)) = V \quad , \quad vec(array(v)) = v$$

定义4.2.5. 称一 $n_x n_y \times n_x n_y$ 维的矩阵 C 为带 circulant 块的 circulant 块矩阵 (BCCB), 如果 C 满足以下条件:

- 1. C 为 BTTB 矩阵
- 2. 每个块 $(C_i)_{n_m \times n_m}$ 其元素按行或按列均为 circulant 型的。
- 3. 每个块 $(C_i)_{n_x \times n_x}$ 均为 circulant 型矩阵。

换言之, C 具有如下形式:

$$C = \begin{pmatrix} C_0 & C_{n_y-1} & \cdots & C_2 & C_1 \\ C_1 & C_0 & C_{n_y-1} & \ddots & C_2 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ C_{n_y-2} & \ddots & C_1 & C_0 & C_{n_y-1} \\ C_{n_y-1} & C_{n_y-2} & \cdots & C_1 & C_0 \end{pmatrix} .$$

其中 C_i 为 $n_x \times n_x$ 维的 circulant 型矩阵。

类似一维的情况,BCCB 矩阵 - 向量的积运算也可以通过 DFT 进行计算。而 BTTB 矩阵 - 向量的积运算可以先将 BTTB 矩阵嵌入到 BCCB 矩阵,然后利用 BCCB 矩阵的性质进行计算。我们在这里就不在叙述。

§4.3 解信赖域子问题的截断 Lanczos 方法

在第二章第六节中,我们介绍了用截断共轭梯度法解信赖域子问题,下面我们介绍解信赖域子问题的截断 Lanczos 方法。首先我们回顾一下共轭梯度法求解极小化问题:

$$\min J_1[\mathbf{f}] = \frac{1}{2} \| \mathcal{K}\mathbf{f} - \mathbf{h} \|^2 = \frac{1}{2} \mathbf{f}^T \mathcal{K}^T \mathcal{K}\mathbf{f} - \mathbf{h}^T \mathcal{K}\mathbf{f} + \frac{1}{2} \mathbf{h}^T \mathbf{h}, \tag{4.3.1}$$

它是一种 Krylov 子空间方法, 其一般的迭代格式如下:

算法4.3.1. 共轭梯度法

步 0: 给定 $\mathbf{f_0}$, 设 $g_0 = grad(J_1[\mathbf{f_0}]), p_0 = -g_0$. 令 $j = 0, \epsilon > 0$ 执行迭代:

步 1: 计算 $a_i = (g_i, g_i)/(p_i, \mathcal{K}^T \mathcal{K} p_i)$,

步 2. 设置 $\mathbf{f}_{i+1} = \mathbf{f}_i + a_i p_i$ 并计算 $g_{i+1} = g_i + a_i \mathcal{K}^T \mathcal{K} p_i$, 如果 $||g_{i+1}|| < \epsilon$, 停机。

步 3: 计算 $\beta_i = (g_{i+1}, g_{i+1})/(g_i, g_i)$,

上述算法生成 Krylov 子空间 $\mathbb{K}(\mathcal{K}^T\mathcal{K}, g_0, k)$ 的一个基底:

$$\mathbb{K}_k = \text{span}\{p_0, p_1, \cdots, p_k\} = \text{span}\{g_0, g_1, \cdots, g_k\},\$$

注意到 Lanczos 方法 (算法 4.3.2) 也产生 Krylov 子空间 \mathbb{K}_k 的另外一组正交基: $\{q_0,q_1,...,q_k\}$

算法4.3.2. Lanczos 算法

步 0: 给定 $\mathbf{f_0}$, $Set\ g_0=grad(J_1[\mathbf{f_0}])$, 令 $y_0=g_0,\ q_{-1}=0$, 对于 $j=0,1,\cdots$, 执行迭代:

步 1: 计算 $\gamma_j = (y_j, y_j)$,

步 2: 计算 $q_j = y_j/\gamma_j$,

步 3: 计算 $\delta_j = (q_j, \mathcal{K}^T \mathcal{K} q_j)$,

如果我们定义 $Q_k = (q_0, ..., q_k)$, 并令 $q_{-1} = 0$,则 Lanczos 算法可以写成矩阵形式:

$$\mathcal{K}^T \mathcal{K} Q_k - Q_k T_k = \gamma_{k+1} q_{k+1} e_{k+1}^T \tag{4.3.2}$$

$$Q_k^T Q_k = I_{k+1}, (4.3.3)$$

其中 T_k 为三对角矩阵:

$$T_{k} = \begin{pmatrix} \delta_{0} & \gamma_{1} & & & \\ \gamma_{1} & \delta_{1} & . & & & \\ & . & . & . & & \\ & & . & \delta_{k-1} & \gamma_{k} & \\ & & & \gamma_{k} & \delta_{k} \end{pmatrix}. \tag{4.3.4}$$

从算法中我们还能得到以下有用性质: 4.3.2:

$$Q_k^T \mathcal{K}^T \mathcal{K} Q_k = T_k \tag{4.3.5}$$

$$Q_k^T g_0 = \gamma_0 e_1 \tag{4.3.6}$$

$$g_0 = y_0 = \gamma_0 q_0 \tag{4.3.7}$$

以上两个算法在某种程度上是等价的,因为 T_k 的元素可以用算法 4.3.1 产生的参数来表达:

$$\gamma_k = \sqrt{\beta_{k-1}} / |\alpha_{k-1}|, \quad \delta_k = \frac{1}{\alpha_k} + \frac{\beta_{k-1}}{\alpha_{k-1}}.$$
(4.3.8)

因此完全可以用 Lanczos 方法来求解信赖域子问题。

当用截断共轭梯度法求解信赖域子问题时,如果试探步越过信赖域边界,算法就简单的在边界处截取步长。即截断共轭梯度法并不关心信赖域边界直到试探步走到边界并要跨过它的时候。这种做法很简单,也很实用。不过我们可以进一步的探索这个方法的性质。注意到共轭梯度法和 Lanczos 方法之间的等价性,很容易将原系统简化成一个三对角矩阵,而求解一个三对角矩阵系统的代价远远要小于原问题。这样我们将 Lanczos 方法整合到截断共轭梯度法中: 当试探步停留在信赖域中的时候,仍采用截断共轭梯度法; 当试探步碰到边界时,我们迭代的求解从原问题得到的三对角系统直到一个满意的解满足中止准则。

令

$$s \in \mathcal{S} = \left\{ s \in \mathbb{R}^{N^2} | s = Q_k u \right\} ,$$

并且寻找

$$s_k = Q_k u_k, \tag{4.3.9}$$

其中 s_k 为子问题 (2.6.4) 的解。则由 (4.3.5)-(4.3.7) 计算可得 u_k 为下述问题

$$\min_{u \in \mathbb{R}^{k+1}} \quad J_3[u] := (u, \gamma_0 e_1) + \frac{1}{2}(u, T_k u), \tag{4.3.10}$$

$$s.t. ||u|| \le \Delta. (4.3.11)$$

的解。设 u_k 为问题 (4.3.10)-(4.3.11) 的可行解,则 KKT 条件可以计算为:

$$T_k(\lambda_k) := (T_k + \lambda_k I)u_k = -\gamma_0 e_1,$$

$$\lambda_k(\|u_k\| - \Delta) = 0,$$

$$\|u_k\| \le \Delta, \ \lambda_k \ge 0.$$

其中, λ_k 为相应的 Lagrangian 乘子。

记 g 为函数 $J_1[\mathbf{f}]$ 的梯度。我们有以下逼近定理:

定理4.3.3.

$$(\mathcal{K}^T \mathcal{K} + \lambda_k I) s_k + g = \gamma_{k+1} (e_{k+1}, u_k) q_{k+1}$$
 (4.3.12)

$$\|(\mathcal{K}^T \mathcal{K} + \lambda_k I) s_k + g\| = \gamma_{k+1} \|(e_{k+1}, u_k)\|$$
(4.3.13)

证明. 注意到

$$\mathcal{K}^{T}\mathcal{K}s_{k} = \mathcal{K}^{T}\mathcal{K}Q_{k}u_{k}
= Q_{k}T_{k}u_{k} + \gamma_{k+1}(e_{k+1}, u_{k})q_{k+1} \quad \mathcal{M}(4.3.2)
= -Q_{k}(\lambda_{k}u_{k} + \gamma_{0}e_{1}) + \gamma_{k+1}(e_{k+1}, u_{k})q_{k+1}
= -\lambda_{k}Q_{k}u_{k} - \gamma_{0}Q_{k}e_{1} + \gamma_{k+1}(e_{k+1}, u_{k})q_{k+1}
= -\lambda_{k}s_{k} - \gamma_{0}q_{0} + \gamma_{k+1}(e_{k+1}, u_{k})q_{k+1}
= -\lambda_{k}s_{k} - g + \gamma_{k+1}(e_{k+1}, u_{k})q_{k+1} \quad \mathcal{M}(4.3.7)$$

于是由 (4.3.12) 立得,(4.3.13) 由向量 q_{k+1} 的正交性得到。 \Box

从上述定理我们看到,只要知道了 γ_{k+1} 和 u_k 的最后一个分量,则误差量 $\gamma_{k+1} \| (e_{k+1}, u_k) \|$ 就完全确定了。因此,完全没有必要知道或求解 s_k 和 Q_k 。

算法4.3.4. 求解信赖域子问题的截断 Lanczos 方法

- 步 g: ϕ $g_0 = grad(J_1[\mathbf{f}])$, $\gamma_0 = \sqrt{(g_0, g_0)}$, 并置 $p_0 = -g_0$; 给定 ϵ (误差允许量); ϕ 标记符 INTERIOR 为 true, k := 0.
- 步 1: 计算 $\alpha_k = (g_k, g_k)/(p_k, Hess(J_1[\mathbf{f}])p_k)$, 并由 (4.3.4) 从 T_{k-1} 得到 T_k 。
- 步 2. 如果标记符 INTERIOR 为真,但 $\alpha_k < 0$ 或 $||s_k + \alpha_k p_k|| \ge \Delta$,重置 INTERIOR 为 false。
- 步 3. 如果标记符 INTERIOR 为真,则令 $s_{k+1} = s_k + \alpha_k p_k$; 否则,求解信赖域子问题 (4.3.10)-(4.3.11) 求得 u_k 。
- 步 5: 如果标记符 INTERIOR 为 true 并且 $||g_{k+1}|| \le \epsilon$,停机;如果标记符 INTERIOR 为 false 并且 $\gamma_{k+1}|(e_{k+1},u_k)| \le \epsilon$,转步 7

步 6. 计算 $\beta_k = (g_{k+1}, g_{k+1})/(g_k, g_k)$ 并令 $p_{k+1} = -g_{k+1} + \beta_k p_k$, k := k+1,转步 1。 步 7. 计算 $s_k = Q_k u_k$,停机。

在上述算法中,我们需要从 u_k 重构 s_k ,因此,Lanczos 向量可以存储起来或者重新产生。当然在计算过程中可以存储 α_i 和 β_i ,然后利用这些参数重新执行迭代,恢复 Lanczos 向量,这样可以节省存储空间。另外我们需要计算三对角系统 (4.3.10)-(4.3.11) 的解,这时由于问题很简单,用第二章第六节中提到的牛顿法来解就能快速的得到解。

使用预条件技巧能经常能有效的提高共轭梯度法的速度。在我们的论文 [50] 中也提出了一种利用预条件的 Lanczos 算法,这里不再赘述。值得注意的是,预条件矩阵的选取相当有技巧。

下面我们设计了一个对于 Gaussian 点扩散函数的计算 y = Kx 的算法, 其中 K 是 $mn \times mn$ 阶矩阵, 假设 A,B 是三对角矩阵, 因此 K 的不同元素仅为: $C = A(1:2,1) \otimes B(1:2,1)$ 。 我们用 Matlab 语言描述这一算法:

算法4.3.5. Gaussian 点扩散函数的块状矩阵 - 向量乘法

```
\begin{split} &function \ y = MatVecTri(m \ , \ n \ , \ C \ , \ x) \\ &y = blockmulti(m \ , \ n \ , \ C(1:2) \ , \ x \ ); \\ &x = blockmulti(m \ , \ n \ , \ C(3:4) \ , \ x \ ); \\ &y(1:n*(m-1)) = y(1:n*(m-1)) + x(n+1:m*n); \\ &y(n+1:m*n) = y(n+1:m*n) + x(1:n*(m-1)); \\ &function \ b = blockmulti(m \ , \ n \ , \ D \ , \ x) \\ &b = zeros(m*n,1); \\ &tmp = D(2)*x; \\ &b(1:m*n-1) = tmp(2:m*n); \\ &b(n:n:m*n) = 0; \\ &b = D(1)*x + b; \\ &tmp(2:m*n) = tmp(1:m*n-1); \\ &tmp(1:n:m*n) = 0; \end{split}
```

b = b + tmp:

在这一算法中,矩阵 - 向量的乘法的代价仅为 4mn 次乘法。而如果用 FFT 计算,工作量为 $O(mn\log_2(mn))$ 。当 A, B 的带宽取为 5 时,这个技巧仍很适用。

§4.4 数值算例

在这一节中,我们给出遥感图像的重构。模糊化的过程由 Gaussian 点扩散函数来描述:

$$k(x, y, \xi, \eta) = \frac{1}{2\pi\rho\bar{\rho}} \exp(-\frac{1}{2} (\frac{x-\xi}{\rho})^2 - \frac{1}{2} (\frac{y-\eta}{\bar{\rho}})^2) \quad , \tag{4.4.1}$$

显然, (4.4.1) 可以写成:

$$k(x, y, \xi, \eta) = k_{1,\rho}(x, \xi) k_{2,\bar{\rho}}(y, \eta),$$
 (4.4.2)

其中 $k_{i,\zeta}(x,\xi) = \frac{1}{\sqrt{2\pi\zeta}} \exp(-\frac{1}{2}(\frac{x-\xi}{\zeta})^2)$, ρ 和 $\bar{\rho}$ 是分别描述 x 和 y 方向的模糊化的常量。在我们的测试中,取 $\rho = \bar{\rho} = 0.7$ 。

k 的离散化是通过 k_1 和 k_2 的离散化来实现的。显然 k_1 和 k_2 是 Toeplitz 型的。 k_1 和 k_2 的张量积构成 k。我们选择 k_1 和 k_2 为带状矩阵,带宽为 5。因此 k 的离散化为一个块状 Toeplitz 矩阵。注意到 k_1 , k_2 的窄带宽,矩阵 - 向量乘法的代价极为低廉。

为了描述模糊化过程,我们假设噪声满足零均值,标准差为 σ 的 Gaussian 分布。即我们在数据中加入 Gaussian 白噪声:

$$\mathbf{f_n} = \mathbf{f} + \delta \cdot \text{rand}(mn, 1),$$

其中δ是噪声水平。

我们将我们的算法应用到三幅不同的遥感图像中,第一幅是加利福利亚一个飞机场的 256×256 图像,第二幅是 Madrid 体育场的 256×256 图像,第三幅是北京市的一副 256×256 的鸟瞰图。图 4.1 列出了真实图像。为了测试算法的有效性,我们给真实图像添加不同噪声水平。首先,我们给这些图像添加噪声水平为 $\delta=0.001$ 的白噪声,图 4.2 列出了模糊化的图像。而图 4.6 列出了由信赖域算法恢复的图像,其子问题由不带预条件的 Lanczos 方法来解。在图 4.10 列出了由带预条件的 Lanczos 方法恢复的图像,其预条件取为 $\mathcal{K}^T\mathcal{K}$ 的主对角。同样的预条件技巧使用于其它不同噪声水平的测试。接着我们添加噪声水平为 $\delta=0.005$ 的白噪声,图 4.3 列出了模糊化的图像,图 4.7 列出了不带预条件的信赖域算法恢复的图像,图 4.11 列出了带预条件的信赖域算法的恢复的图像。接着我们添加更高的噪声水平 $\delta=0.01,0.02$,图 4.4,4.5 分别列出了模糊化的图像,图 4.8,4.9 分别列出了不带预条件的信赖域算法恢复的图像,图 4.8,4.9 分别列出了不带预条件的信赖域算法恢复的图像,图 4.12,4.13 分别列出了带预条件的信赖域算法的恢复的图像。

在所有测试中,选取 $\eta_1 = 0.9$, $\eta_2 = 0.9$, $\gamma_1 = 0.5$, $\gamma_2 = 1.5$, $\gamma_3 = 2$, 初始的信赖域半径为 $\Delta = 0.1 \|\mathbf{h_n}\|_2$ 。这里 $\mathbf{h_n} := \mathbf{h} + \mathbf{n}$ 。记总内迭代步数为 *initer*,总外迭代不是为 *outiter*,真实图像与恢复图像的绝对误差为 *error*:

 $error = \|image_{true} - image_{restoration}\|_{l_2}.$

我们在表 4.1-4.4 中列出迭代结果。从表 4.1 和表 4.2,我们看到当误差水平 δ 不是很大时,由不带预条件的信赖域方法得到的绝对误差 error 比用带预条件技巧得到的误差小,但使用预条件需要的迭代步数少。然而,当误差水平 δ 比较大时,情况完全不同,我们可以从表 4.3 和表 4.4 中看出。我们注意到带预条件技巧的算法要好于不带预条件的算法,不论是从迭代步数还是解的绝对误差来看。

表 **4.1**: 噪声水平 $\delta = 0.001$ 下的迭代结果

images	initer		outiter		error	
	LTRS	PLTRS	LTRS	PLTRS	LTRS	PLTRS
飞机场	41	35	3	5	4379.5589	4873.0837
体育场	27	27	3	5	3117.8386	3335.4046
鸟瞰图	43	37	3	5	4640.2131	5196.2899

表 **4.2**: 噪声水平 $\delta = 0.005$ 下的迭代结果

<u> </u>							
images	initer		out iter		error		
	LTRS	PLTRS	LTRS	PLTRS	LTRS	PLTRS	
飞机场	44	36	3	5	4627.4995	4937.3513	
体育场	39	27	3	5	3000.2422	3392.6917	
鸟瞰图	45	38	3	5	4854.0366	5251.5561	

表 **4.3**: 噪声水平 $\delta = 0.01$ 下的迭代结果

R 10 R/ WY 0 001 T FIZE VALVE							
images	initer		out iter		error		
	LTRS	PLTRS	LTRS	PLTRS	LTRS	PLTRS	
飞机场	45	37	3	5	5972.9118	5443.6936	
体育场	47	37	4	5	5884.7627	5709.6846	
鸟瞰图	44	27	3	5	4774.5657	3607.3968	

表 4.4: 噪声水平 $\delta=0.02$ 下的迭代结果

images	initer		out iter		error	
	LTRS	PLTRS	LTRS	PLTRS	LTRS	PLTRS
飞机场	69	48	5	6	11364.7184	8572.6459
体育场	67	44	5	6	10652.6523	7542.3392
鸟瞰图	67	53	5	6	10838.7183	8687.0703

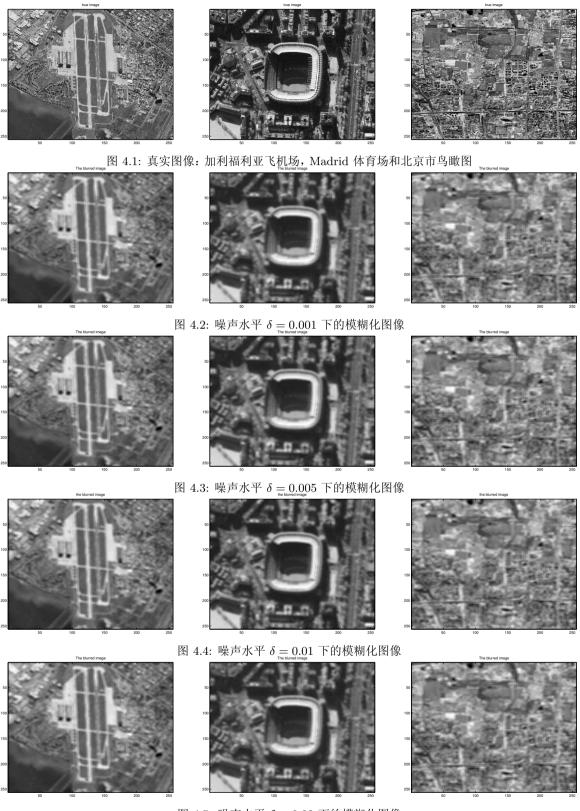


图 4.5: 噪声水平 $\delta = 0.02$ 下的模糊化图像

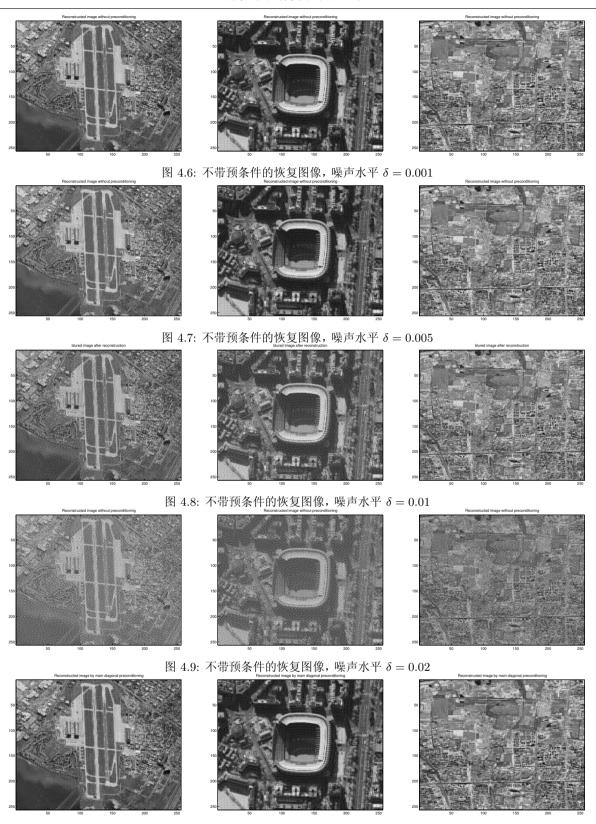
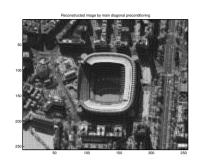


图 4.10: 带预条件的恢复图像, 噪声水平 $\delta=0.001$





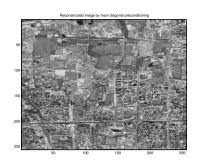
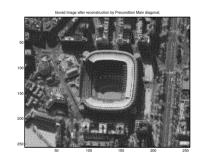


图 4.11: 带预条件的恢复图像, 噪声水平 $\delta=0.005$





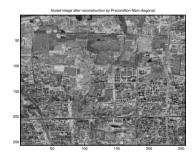
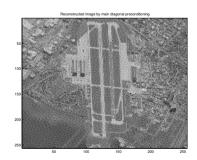
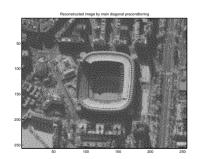


图 4.12: 带预条件的恢复图像, 噪声水平 $\delta=0.01$





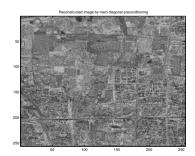


图 4.13: 带预条件的恢复图像, 噪声水平 $\delta=0.02$

第二部分

非线性最小二乘法及其应用

第五章 非线性最小二乘问题

在第一部分中,我们讨论了线性最小二乘问题。但在很多实际应用中,线性函数往往 不能真正描述系统的性质,必须考虑非线性模型。这时目标函数可能是一个显式的非线 性函数,也可能是一个无法显式积出的复杂积分,还可能是一个含参数的动力系统或偏 微分方程组的解。因此在这一部分中,我们将集中讨论非线性最小二乘问题及其解法。

§5.1 无约束非线性最小二乘问题

为了叙述的方便,我们重新提一下非线性最小二乘问题:

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2} ||r(x)||_2^2 \quad , \tag{5.1.1}$$

其中 $r(x) = (r_1(x), \dots, r_m(x))^T$, $r_j(x)$ 是从 \mathbb{R}^n 到 \mathbb{R} 的光滑函数。

引入 f(x) 的 Jacobian 矩阵:

$$J(x) = \begin{pmatrix} \frac{\partial r_1}{\partial x_1} & \cdots & \frac{\partial r_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial r_m}{\partial x_1} & \cdots & \frac{\partial r_m}{\partial x_n} \end{pmatrix}, \tag{5.1.2}$$

此时, J(x) 一般是一个与 x 有关的矩阵。则我们可以将 f(x) 的梯度和 Hessian 阵表达为:

$$\nabla f(x) = \sum_{j=1}^{m} r_j(x) \nabla r_j(x) = J(x)^T r(x) , \qquad (5.1.3)$$

$$\nabla^{2} f(x) = \sum_{j=1}^{m} \nabla r_{j}(x) \nabla r_{j}(x)^{T} + \sum_{j=1}^{m} r_{j}(x) \nabla^{2} r_{j}(x)$$
$$= J(x)^{T} J(x) + S(x) , \qquad (5.1.4)$$

其中:

$$S(x) = \sum_{i=1}^{m} r_i(x) \nabla^2 r_i(x) \quad . \tag{5.1.5}$$

在几何上, r(x) 可以看成是 \mathbb{R}^m 空间中的 n 维曲面, 因此问题 (5.1.1) 可以写为:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} ||r(x) - 0||_2^2 \quad ,$$

即在由 r(x) 确定的曲面上寻找一个离原点最近的点。如果记 $\mathcal{R}(J)$ 上的正交投影为 P_R ,则 r(x) 的切平面上距原点最近的点是 $(I_m - P_R)r(x)$ 。

在 Bates 和 Watts [48] 的讨论中,2 阶导数可以看成m维的加速向量(acceleration vectors)。这些向量可以分成两部分,一部分落在切平面上,另一部分正交于切平面。我们可以很方

便的将 m 个矩阵 $\nabla^2 r_j(x)$ 排列成一个 $m \times (n \times n)$ 的矩阵 G_{acc} , 于是切平面和正交于切平面的分量分别为 $G^T = P_R G_{acc}$ 和 $G^N = (I_m - P_R) G_{acc}$ 。如果定义:

$$K_h^T = \frac{\|h'G^Th\|}{\|Jh\|_2^2} \quad , \quad K_h^N = \frac{\|h'G^Nh\|}{\|Jh\|_2^2}$$
 (5.1.6)

分别为方向 h 上的参数效应曲率 (Parameter-effect curvature)和内在曲率 (Intrinsic curvature)。这两个曲率可以用来衡量非线性最小二乘的非线性程度,通常所说的参数变换可以影响参数效应曲率,即可以使参数在切平面上的投影分布相对均匀,但并能不改变函数的内在非线性,即内在曲率。

§5.2 Gauss-Newton 法

有了上节中的铺垫,可以很快的得到非线性最小二乘目标函数 f(x) 的二次模型为:

$$m_k(x_k + s) = f(x_k) + \nabla f(x_k)^T s + \frac{1}{2} s^T \nabla^2 f(x_k) s$$

= $\frac{1}{2} ||r(x_k)||_2^2 + (J(x_k)^T r(x_k))^T s + \frac{1}{2} s^T (J(x_k)^T J(x_k) + S(x_k)) s$, (5.2.1)

从而,解问题 (5.1.1) 的牛顿法为:

$$x_{k+1} = x_k + s_k$$

 $s_k = -(J(x_k)^T J(x_k) + S(x_k))^{-1} J(x_k)^T r(x_k)$.

在一定的假设下,牛顿法具有局部的二阶收敛速度。但是牛顿法的主要问题是 Hessian 阵 $\nabla^2 f(x)$ 中的二阶信息项 S(x) 通常难以计算或者花费的工作量很大。值得注意的是在计算梯度 $\nabla f(x)$ 时就已经得到了 J(x),即实际上已经得到了 Hessian 阵的一阶信息项 $J(x)^T J(x)$ 。特别当模型在解的附近时(此时模型经常接近线性,从而 $\nabla^2 r_j$ 比较小),或者当余量较小(即 r_j 比较小)时,由式 (5.1.5) 可以看出,S(x) 常可忽略不计。鉴于此,我们或者忽略 S(x),或者用一阶导数信息逼近 S(x)。这是求解非线性最小二乘问题的关键所在 [53]。

最简单的求解非线性最小二乘问题的算法之一是 Gauss-Newton 法,它直接忽略 $\nabla^2 f(x)$ 中的二阶信息项 S(x),求解方程:

$$(J(x_k)^T J(x_k)) s_k^{GN} = -J(x_k)^T r(x_k)$$
(5.2.2)

得到试探步 s_k^{GN} 。它的一个优点在于不管 $J(x_k)$ 是否满秩,由关系式:

$$(s_k^{GN})^T \nabla f_k = (s_k^{GN})^T J_k^T r_k = -(s_k^{GN})^T J_k^T J_k s_k^{GN} = -\|J_k s_k^{GN}\|_2^2 \le 0$$

知 s_k^{GN} 是 f(x) 一个下降方向。此外,观察式 (5.2.2),可以发现它实际上是线性最小问题;

$$\min_{s} \frac{1}{2} \|J_k s + r_k\|_2^2 \tag{5.2.3}$$

的正规化方程组,因此我们可以将求解线性最小二乘问题的算法应用到求解子问题 (5.2.3)。 如果使用 QR 或 SVD 算法,则没有必要显式的形成 Hessian 阵的近似 $J_t^T J_k$ 。

子问题 (5.2.3) 也说明 Gauss-Newton 法的另外一个动机: 我们并不寻求函数 f(x) 的二次模型,而是形成向量函数 r(x+s) 的线性模型 $r(x+s)\approx r(x)+J(x)s$ 。然后将这个线性模型代入 $f(x)=\frac{1}{2}||r(x)||_2^2$,极小化得到 Gauss-Newton 步 s^{GN} 。

在实际算法中,我们采用的 Gauss-Newton 往往加上线搜索策略:

$$x_{k+1} = x_k + \alpha_k s_k^{GN} \quad ,$$

其中, α_k 为一维搜索因子, 满足 Wolfe 搜索条件:

$$f(x_k + \alpha_k s_k^{GN}) \leq f(x_k) + c_1 \alpha_k \nabla f_k^T s_k^{GN}$$

$$(5.2.4)$$

$$\nabla f(x_k + \alpha_k s_k^{GN})^T s_k \ge c_2 \nabla f_k^T s_k^{GN} \tag{5.2.5}$$

其中 $0 < c_1 < c_2 < 1$ 。

对于带 Wolfe-Powell 搜索的 Gauss-Newton 法,我们有下面的收敛性定理: **定理5.2.1**.([49])设每个余量函数 r_i 在初始点 x_0 的水平集:

$$\mathcal{L} = \{x | f(x) \le f(x_0)\} \tag{5.2.6}$$

的邻域 N 内 Lipshitz 连续可微, 并且 Jacobian 矩阵 J(x) 满足条件

$$||J(x)z||_2 \ge \gamma ||z||_2 \quad , \quad \gamma > 0$$

则带 Wolfe 搜索的 Gauss-Newton 法有:

$$\lim_{k \to \infty} J_k^T r_k = 0 \quad .$$

但 Gauss-Newton 法在解 x^* 附件的收敛速度取决于线性项 J^TJ 在 Hessian 阵中所占的比重。对于零残量问题(即 $r(x^*)=0$),Gauss-Newton 法有局部二阶收敛速度;对于小残量问题(即残量 r(x) 较小,或 r(x) 接近线性),该方法也有较快的局部收敛速度。但对于大残量问题,收敛速度就很慢了。当 J(x) 奇异的时候,虽然此时由于它与线性最小二乘问题 (5.2.3) 的等价性,仍能找到一个解,但不能得到类似定理 5.2.1 的结果。

§5.3 Levenberg-Marquardt 方法

从上节可以看出,Gauss-Newton 法实际上是带线搜索的近似牛顿法,只不过用 J^TJ 来代替 Hessian 阵,同样我们也可以在信赖域的框架内求解。但信赖域的使用可以避免 Gauss-Newton 法中 Jacobian 矩阵 J(x) 亏秩或近似亏秩的困难。

求解非线性最小二乘问题

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2} ||r(x)||_2^2 \tag{5.3.1}$$

的 Levenberg-Marquardt 方法,可以在求解非线性优化问题的信赖域算法的框架内来进行描述和分析。实际上, Levenberg-Marquardt 方法通常被认为是解无约束问题的信赖域算法的先驱。对于球形信赖域,信赖域算法每次迭代求解子问题:

$$\min m_k(s) = \frac{1}{2} ||r(x_k) + J(x_k)s||_2^2$$
s.t. $||s|| \le \Delta_k$. (5.3.2)

或等价于求解子问题

$$\min \phi_k(s) = s^T J(x_k)^T r(x_k) + \frac{1}{2} s^T (J(x_k)^T J(x_k)) s$$
s.t. $||s|| < \Delta_k$. (5.3.3)

其中 Δ_k 是信赖域半径。定义比值:

$$r_k = \frac{Ared_k}{Pred_k} = \frac{f(x_k) - f(x_k + s_k)}{\phi_k(0) - \phi_k(s_k)}$$
 (5.3.4)

来控制是否接受信赖域试探步以及是否调整信赖域半径。下面我们来具体描述求解非线性最小二乘问题的信赖域算法:

算法5.3.1. 解非线性最小二乘问题的信赖域算法

- 步 0: 初始化。给定初始点 x_0 和信赖域半径 Δ_0 。选取信赖域参数 $\eta_1, \eta_2, \gamma_1, \gamma_2, \gamma_3$ 使 得 $0 < \eta_1 \le \eta_2 < 1$ 和 $0 < \gamma_1 \le 1 \le \gamma_2 \le \gamma_3$ 。计算 $f(\mathbf{x_0})$ 并令 k := 0。
- 步 1: **求步长**。若满足停机条件,则停机;否则求解子问题 (5.3.2) 或 (5.3.3) 得到步长 s_k ,使得"目标函数模型充分下降"。
- 步 2: 是否接受试探步。计算 $f(\mathbf{x}_k + s_k)$,并由 (5.3.4) 来计算 r_k ,令

步 3: 更新信赖域半径。令

$$\Delta_{k+1} \in \begin{cases} [\gamma_2 ||s_k||, \gamma_3 \Delta_k], & \ddot{\mathcal{Z}} \quad r_k \ge \eta_2, \\ [\gamma_1 \Delta_k, \gamma_2 \Delta_k], & \ddot{\mathcal{Z}} \text{ 否则}. \end{cases}$$

步 4: 更新模型函数 $m_{k+1}(s)$ 或 $\phi_{k+1}(s)$, 并令 k := k+1; 转步 1。

同样对于子问题 (5.3.2) 我们也有以下定理:

定理5.3.2. s_k 是子问题 (5.3.2) 的解当且仅当存在 $\lambda_k \geq 0$ 使得:

$$(J(x_k)^T J(x_k) + \lambda_k I) s_k = -J(x_k)^T r(x_k)$$

$$\|s_k\|_2 \le \Delta_k$$

$$\lambda_k (\Delta_k - \|s_k\|_2) = 0$$
(5.3.5)

并且 $J(x_k)^T J(x_k) + \lambda_k I$ 是半正定矩阵。

从定理可以看出子问题 (5.3.2) 的解由方程组:

$$(J(x_k)^T J(x_k) + \lambda_k I)s = -J(x_k)^T r(x_k)$$
(5.3.6)

来表征,而该方程恰好是线性最小二乘问题:

$$\min_{s} \frac{1}{2} \left\| \begin{pmatrix} J_k \\ \sqrt{\lambda}I \end{pmatrix} s + \begin{pmatrix} r_k \\ 0 \end{pmatrix} \right\|_{2}^{2}$$
(5.3.7)

的正规方程组。这样就像前一节的 Gauss-Newton 法,不用显式的形成矩阵 $J_k^T J_k$ 来解子问题。假设我们已经用 Householder 变换单独计算了矩阵 J_k 的 QR 分解:

$$J_k = Q \begin{pmatrix} R \\ 0 \end{pmatrix} \quad ,$$

则有:

$$\begin{pmatrix} R \\ 0 \\ \sqrt{\lambda}I \end{pmatrix} = \begin{pmatrix} Q^T \\ I \end{pmatrix} \begin{pmatrix} J_k \\ \sqrt{\lambda}I \end{pmatrix} .$$

上式左边的矩阵上半部分是一个上三角矩阵,下半部分为一个只有n个非零元素的对角矩阵 $\sqrt{\lambda}I$,因此我们可以利用 n(n+1)/2 次 Givens 变换将整个矩阵变为一个上三角矩阵:

$$\bar{Q}_{\lambda}^{T} \begin{pmatrix} R \\ 0 \\ \sqrt{\lambda}I \end{pmatrix} = \begin{pmatrix} R_{\lambda} \\ 0 \\ 0 \end{pmatrix} \quad ,$$

其中 \bar{Q}_{λ}^{T} 是一系列 Givens 旋转矩阵的乘积。因此,令

$$Q_{\lambda} = \begin{pmatrix} Q^T & \\ & \\ & I \end{pmatrix} \bar{Q}_{\lambda}^T$$

这样我们就得到了式 (5.3.7) 中系数矩阵的 QR 分解:

$$\begin{pmatrix} R_{\lambda} \\ 0 \end{pmatrix} = Q_{\lambda} \begin{pmatrix} J_k \\ \sqrt{\lambda}I \end{pmatrix} \quad ,$$

于是有分解 $R_{\lambda}^{T}R_{\lambda} = J_{k}^{T}J_{k} + \lambda I$, 可以很方便的求解出方程 (5.3.6)。

在最小二乘问题中,变量的尺度经常不一样,如有些变量的量级可能达到 10^4 ,而有些变量的量级仅为 10^{-6} ,如果不考虑这些因素,算法可能遇到数值困难,或者得到的解性态比较差。一个办法是使用调比矩阵 D_k ,考虑求解椭球形信赖域子问题:

$$\min \frac{1}{2} ||r(x_k) + J(x_k)s||_2^2$$
s.t. $||D_k s|| \le \Delta_k$. (5.3.8)

来克服。其中调比矩阵在迭代过程中 D_k 自动调整大小,如使得它们与 $J_k^T J_k$ 的对角元的量级匹配。在本章第五节中,我们将考虑另外一个克服技巧。

Levenberg-Marquardt 方法的局部收敛性质和 Gauss-Newton 方法的类似。对于小残量问题,在解的附近,模型 (5.3.2) 能很好的逼近函数 f(x),并且随着迭代信赖域将最终变得松弛,试探步也逐渐取为 Gauss-Newton 步。

§5.4 拟牛顿方法

对于大残量问题 (即 $r(x^*)$ 很大或 r(x) 的非线性程度很高), Gauss-Newton 法和 Levenberg-Marquardt 方法可能收敛得很慢,这主要是因为 Hesse 矩阵 $\nabla^2 f(x)$ 中得二阶信息项 S(x) 不能忽略。但正如前面所指出一样,S(x) 通常难以计算或花费的计算量很大。这就启示我们构造 $S(x_k)$ 的割线近似。

设 S_k 是 $S(x_k)$ 的割线近似,则迭代 (5.2.2) 成为:

$$(J(x_k)^T J(x_k) + S_k)s = -J(x_k)^T r(x_k) .$$

我们要求 S_k 满足某种拟牛顿条件。由于:

$$S(x_{k+1}) = \sum_{i=1}^{m} r_i(x_{k+1}) \nabla^2 r_i(x_{k+1}) ,$$

故我们用:

$$S_{k+1} = \sum_{i=1}^{m} r_i(x_{k+1})(H_i)_{k+1}$$

去近似 $S(x_{k+1})$ 。这里 $(H_i)_{k+1}$ 是 $\nabla^2 r_i(x_{k+1})$ 的拟牛顿近似,故有:

$$(H_i)_{k+1}(x_{k+1} - x_k) = \nabla r_i(x_{k+1}) - \nabla r_i(x_k)$$
,

于是:

$$S_{k+1}(x_{k+1} - x_k) = \sum_{i=1}^{m} r_i(x_{k+1})(H_i)_{k+1}(x_{k+1} - x_k)$$

$$= \sum_{i=1}^{m} r_i(x_{k+1})(\nabla r_i(x_{k+1}) - \nabla r_i(x_k))$$

$$= (J(x_{k+1}) - J(x_k))^T r(x_{k+1}) \cong y_k , \qquad (5.4.1)$$

这就是 Sk 要满足的拟牛顿条件。

下面的定理给出了 S_k 校正公式 ([53]):

定理5.4.1. 令 $v_k^T \Delta x_k \ge 0$, $T \in \mathbb{R}^{n \times n}$ 是对称正定的加权矩阵, 满足:

$$TT^Ts_k = v_k$$

其中:

$$v_k \cong \nabla f(x_{k+1}) - \nabla f(x_k) = J(x_{k+1})^T r(x_{k+1}) - J(x_k)^T r(x_k)$$

那么,校正:

$$\begin{split} S_{k+1} = & S_k & + \frac{(y_k - S_k \Delta x_k) v_k^T + v_k^T (y_k - S_k \Delta x_k)}{\Delta x_k^T v_k} \\ & - & \frac{\Delta x_k^T (y_k - S_k \Delta x_k) v_k v_k^T}{(\Delta x_k^T v_k)^2} \end{split}$$

是极小化问题:

的唯一解。

Dennis, Gay 和 Welsch [35] 最早提出了解非线性最小二乘问题的拟牛顿算法,他们解信赖域子问题:

$$\min \phi_k(s) = s^T J(x_k)^T r(x_k) + \frac{1}{2} s^T (J(x_k)^T J(x_k) + S_k) s$$
s.t. $||s|| \le \Delta_k$. (5.4.3)

即用 $J(x_k)^T J(x_k) + S_k$ 来近似 Hessian 阵。值得注意的是, 在

$$S(x) = \sum_{i=1}^{m} r_i(x) \nabla^2 r_i(x)$$

中,分量 $r_i(x)$ 有时比二阶导数分量 $\nabla^2 r_i(x)$ 变化得更快,这在零残量和小残量问题中尤为明显。但对于 S_k ,基本的校正策略并不保证在迭代趋于零残量解或小残量解时该矩阵也趋于零。这就有可能扰乱超线性收敛性。因此,可选择调比因子:

$$\tau_k = \min\{1, \frac{s_k^T y_k}{s_k^T S_k s_k}\}$$

然后用 $\tau_k S_k$ 而不是 S_k 对式 (5.4.2) 的右边进行校正。当这个策略采用时,如果 $r(x_{k+1}) = 0$,则 $y_k = 0$,因而 $\tau_k = 0$, $S_{k+1} = 0$ 。数值试验表明这对拟牛顿算法的表现有显著的影响。

Dennis , Gay 和 Welsch 提出的算法另一个特点是根据迭代的表现选择模型,当 Gauss-Newton 模型能产生一个很好的试探步时,则还是从 Hessian 近似中忽略掉 S_k 。

§5.5 M 范数意义下的信赖域方法

在前面的讨论中, Levenberg-Marquardt 方法和拟牛顿法都是在信赖域框架下来解的。信赖域算法的一个核心是如何有效的求解信赖域子问题。在不同的范数意义下,解信赖域子问题的难易程度可能不一样,这在文献 [47] 中有比较详细的叙述。由于最小二乘问题的特殊结构,使得寻找一个特定的范数成为可能。但究竟使用何种范数,对信赖域算法本身的理论性质没有影响。

首先考虑一般的信赖域子问题:

min
$$g^T s + \frac{1}{2} s^T B s$$

s.t. $\|s\|_M \le \Delta_k$. (5.5.1)

其中 B 对称且非奇异。那么写出 B 的谱分解为:

$$B = U\Lambda U^T \quad , \tag{5.5.2}$$

其中对角矩阵 Λ 的对角元素即为 B 的特征值,而正交矩阵 U 是 B 的特征向量排成的矩阵。记 B 的绝对值为:

$$|B| = U|\Lambda|U^T \quad ,$$

其中绝对值 |Λ| 是对角矩阵 Λ 相应的元素取绝对值得到的。

很自然的一个 M 范数可取 M = |B|,即 $||s||_M^2 = \langle s, |B|s \rangle$ 。这个范数在一定程度上也调比原问题 — 那些使模型变化最快的方向,还有那些使模型可能最远离真实函数的方向比曲率小的方向更多的受到限制。另外一个有趣的性质是,只需要进行一次谱分解 (5.5.2) 就足够求解子问题 (5.5.1)。令

$$s_D = U^T s$$
 . $a_D = U^T a$

利用 U 的正交性, 子问题 (5.5.1) 的解可以表示为 $s = Us_D$, 其中 s_D 是对角信赖域子问题:

min
$$g_D^T s_D + \frac{1}{2} s_D^T \Lambda s_D$$

s.t. $\langle s_D, |\Lambda| s_D \rangle \leq \Delta_k^2$ (5.5.3)

的解。可以看出这种对角的信赖域子问题是非常容易求解的,其解法我们将在后面详细 讨论。

§5.5.1 Levenberg-Marquardt 方法的 M 范数

求解 Levenberg-Marquardt 方法的信赖域子问题为:

min
$$s^T J(x_k)^T r(x_k) + \frac{1}{2} s^T (J(x_k)^T J(x_k)) s$$

s.t. $||s||_M \le \Delta_k$. (5.5.4)

我们不想显式的形成矩阵 $J(x_k)^T J(x_k)$ 再求谱分解。然而我们可以求矩阵 $J(x_k)$ 的奇异值分解:

$$J(x_k) = U\Sigma V^T$$
 , $\Sigma = diag(\sigma_1, \dots, \sigma_n)$,

则有:

$$J(x_k)^T J(x_k) = V \Sigma^2 V^T \quad .$$

显然 M 范数矩阵就可以取为 $M = V\Sigma^2V^T$ 。但有时 $J(x_k)$ 可能是奇异的,或者存在小奇异值簇,这时 M 范数矩阵的取法就应该有所调整,即取 $M = V\hat{\Sigma}^2V^T$,其中

$$\hat{\Sigma} = \begin{cases} \sigma_i, & \text{如果 } \sigma_i \ge \delta, \\ \delta, & \text{否则.} \end{cases}$$

这里 $\delta > 0$ 是一个小正数。定义:

$$s_s = \hat{\Sigma} V^T s$$
 , $g_s = \hat{\Sigma}^{-1} V^T J(x_k)^T r(x_k)$,

则解子问题 (5.5.4) 相当于求解:

$$\min \quad g_s^T s_s + \frac{1}{2} s_s^T D s_s$$
s.t.
$$||s_s||_2 \le \Delta_k$$
(5.5.5)

其中 $D \equiv \hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}$ 是一个对角矩阵。而解 s 可以通过 $s = V\hat{\Sigma}^{-1}s_s$ 来恢复。

该算法对中小规模的问题或者 $m \gg n$ 时是相对有效的。

§5.5.2 拟牛顿法的 M 范数

求解非线性最小二乘法的拟牛顿法的信赖域子问题为:

$$\min \quad g^T s + \frac{1}{2} s^T B_k s$$
s.t. $||s|| \le \Delta_k$. (5.5.6)

其中 $g = J(x_k)^T r(x_k)$, $B_k = J(x_k)^T J(x_k) + S_k$, 这时我们仍然可以求矩阵 B_k 的谱分解来把问题化成对角系统。但当问题规模比较大时,谱分解的代价很大,因此要考虑其它途径。

当拟牛顿矩阵 S_k 正定时,由于 $J(x_k)^T J(x_k)$ 总是半正定的, B_k 也是半正定的,这时可以通过 LDL^T 分解来达到目标。但有些拟牛顿校正得到的矩阵 S_k 可能是不正定的,这时 B_k 也不正定。而一个不定矩阵的 LDL^T 分解可能不存在,方法会失败。幸好我们有 B_k 的对称不定分解:

$$B_k = PLBL^T P^T \quad , \tag{5.5.7}$$

其中 P 为置换矩阵,L 为单位下三角矩阵,B 为最大块为 2 维的块对角矩阵。再计算矩阵 B 的谱分解:

$$B = Q\Theta Q^T$$
 , $\Theta = diag(\theta_1, \dots, \theta_n)$,

由于 B 的特殊性,可以看出 Q 的每列至多只有两个非零元素。然后取

$$|B| = Q|\Theta|Q^T \quad .$$

其中 $|\Theta|$ 为 Θ 的绝对值。这时很自然 M 范数矩阵就可以取为 $M = PL|B|L^TP^T$ 。同样为了克服奇异性,我们需要进行一些修正,取 $M = PLCL^TP^T$,其中

$$C = Q\Gamma Q^T$$
 ,
$$\Gamma = \begin{cases} \theta_i, & \text{如果} & \theta_i \ge \delta, \\ -\theta_i, & \text{如果} & \theta_i \le -\delta, \\ \delta, & \text{否则} \end{cases}$$

这样定义:

$$s_s = \Gamma^{\frac{1}{2}} Q^T L^T P^T s$$
 , $g_s = \Gamma^{-\frac{1}{2}} Q^T L^{-1} P^T g$,

则解子问题 (5.5.6) 相当于求解:

$$\min \quad g_s^T s_s + \frac{1}{2} s_s^T D s_s$$
s.t.
$$||s_s||_2 \le \Delta_k$$
(5.5.8)

其中 $D \equiv \Gamma^{-\frac{1}{2}}\Theta\Gamma^{-\frac{1}{2}}$ 是一个对角矩阵。而解 s 可以通过 $s = PL^{-T}Q\Gamma^{-\frac{1}{2}}s$ 。来恢复。

§5.5.3 解对角信赖域子问题

从前几节的分析可以看出,当问题对角化后,问题的关键是解对角信赖域子问题 (5.5.5) 或 (5.5.8)。由于对角系统非常简单,矩阵 D 的特征值就是它的对角线元素,因此用牛顿法来解就相当有效。下面我们给出具体算法:

算法5.5.1. 解对角信赖域子问题

$$\diamond \epsilon \in (0,1)$$
.

步 1: 如果 D 严格正定,令 $\lambda = 0$ 且 $s_s = -D^{-1}g_s$ 。

步 1a: 如果 $||s_s||_2 \leq \Delta$, 迭代停止。

步 2. 否则, 计算 D 的最左边的特征值 θ , 令 $\lambda = -\theta$, 并且定义 g_n^n 使得:

$$(g_s^n)_i = \begin{cases} (g_s)_i, & \text{如果} (D)_{ii} + \lambda = 0, \\ 0, & \text{否则}. \end{cases}$$

步 2a: 如果 $g_s^n = 0$,令 $s_s = -(D + \lambda I)^{\dagger} g_s$ 。

如果 $\|s_s\|_2 \leq \Delta$, 计算 θ 所对应的一个特征向量 u, 并寻找方程 $\|s_s + \alpha u\|_2 = \Delta$ 的一个根 α , 使得模型 $q_s(s_s + \alpha u)$ 最小,令 $s_s = s_s + \alpha u$,迭代中止。

步 2b: 否则,令 $\lambda = \lambda + \|q_s^n\|_2/\Delta$,并置 $s_s = -(D + \lambda I)^{\dagger}q_s$

步 3. 如果 $|||s_s||_2 - \Delta| \le \epsilon \Delta$, 迭代中止。

步 4: 令:

$$\lambda = \lambda + \left(\frac{\|s_s\|_2 - \Delta}{\Delta}\right) \left(\frac{\|s_s\|_2^2}{\langle s_s, (D + \lambda I)^{\dagger} s_s \rangle}\right)$$

步 5: $\diamondsuit s_s = -(D + \lambda I)^{\dagger} g_s$, 转步 3。

§5.6 带约束的最小二乘问题及其解法

在很多实际问题中,参数只有在一定的取值范围内才有意义,或者参数还要满足一些的附加条件才能描述真实的系统。这时建立的模型更为复杂,往往具有如下形式:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \sum_{i=1}^m r_i(x)^2$$
s.t. $c_i(x) = 0$, $i = 1, \dots, q_e$,
$$c_i(x) \ge 0$$
 , $i = q_e + 1, \dots, q$

$$(5.6.1)$$

这就是我们所说的带约束最小二乘的问题。对于这类问题,可以用非线性优化领域内的传统算法来解,如罚函数法,可行方向法,逐步二次规划(SQP)等等,这些方法在 [53, 49, 47] 中都有详细的叙述。此外有很多数值表现优异的程序,软件包实现了这些方法,如 VFO2AD、NLPQL、NPSOL、Knitro、Lancelot、Galahad 等等。在这一节里,我们不打算详细的展开,只简要的概述一下解带约束最小二乘的问题的特殊算法。

由于大多数最小二乘问题是病态的,Schittkowski[34] 并不建议直接应用前面所提到的算法来求解问题 (5.6.1),而是引入 m 个附加变量 $z = (z_1, \dots, z_m)^T$ 和 m 个等式约束:

$$r_i(x) - z_i = 0$$
 , $i = 1, \dots, m$,

这样问题 (5.6.1) 转换成:

$$\min_{(x,z)\in\mathbb{R}^{n+m}} \quad \frac{1}{2}z^{T}z$$
s.t. $r_{i}(x) - z_{i} = 0$, $i = 1, \dots, m$,
$$c_{i}(x) = 0$$
 , $i = 1, \dots, q_{e}$,
$$c_{i}(x) \geq 0$$
 , $i = q_{e} + 1, \dots, q$

$$(5.6.2)$$

然后再用 SQP 来解这个变换后的问题。Schittkowski 用大量的算例表明该方法在实际应用中非常有效。

Lindstrőm [36] 将 SQP 方法和 Gauss-Newton 方法结合起来,每次迭代解:

$$\min_{s \in \mathbb{R}^n} \quad \frac{1}{2} ||J(x_k)s + r(x_k)|| 2$$
s.t.
$$\hat{A}(x_k)s + \hat{c}(x_k) = 0$$
(5.6.3)

其中 $\hat{A}(x_k)$ 是当前积极集所对应的约束 $\hat{c}(x_k)$ 的 Jacobian 矩阵。这是一个带约束的线性最小二乘问题,其特殊的结构可使它比一般的二次规划问题更容易求解。此外 Hanson 和 Krogh [37] 基于张量模型提出了一种解带线性约束的最小二乘问题的算法(实际上他们的程序还能解一些非线性约束的问题)。

Port 数学库 [38] 中还考虑仅带盒式约束的最小二乘问题:

$$\min_{x \in \mathbb{R}^n} \quad \frac{1}{2} \sum_{i=1}^m r_i(x)^2$$
s.t. $l \le x \le u$ (5.6.4)

先定义投影算子:

$$P(x,l,u)_{i} = \begin{cases} l_{i}, & \text{yi } \mathbb{R} & x_{i} < l_{i}, \\ x_{i}, & \text{yi } \mathbb{R} & x_{i} \in [l_{i},u_{i}], \\ u_{i}, & \text{yi } \mathbb{R} & x_{i} > u_{i}, \end{cases}$$
 (5.6.5)

这样一个简单的解问题 (5.6.4) 的算法为:

$$x_{k+1} = P(x_k + s_k, l, u),$$

其中 s_k 是无约束最小二乘问题:

$$\min_{x \in \mathbb{R}^n} \quad \frac{1}{2} \sum_{i=1}^m r_i(x)^2$$

的试探步,可以由 Gauss-Newton 法或 Levenberg-Marquardt 方法或拟牛顿法计算得到。

§5.7 可分非线性最小二乘问题与变量投影法

在这一节里我们介绍一种特殊结构的非线性最小二乘问题,其目标函数是线性和非 线性函数的组合,即

$$\min \phi(c, x) = \frac{1}{2} \|\Phi(x)c - y(x)\|_2^2 \quad , \tag{5.7.1}$$

其中 $\Phi: \mathbb{R}^p \to \mathbb{R}^{n \times l}$, $y: \mathbb{R}^p \to \mathbb{R}^n$ 连续可微。近年来有很多文献 [39, 40, 42, 43, 44, 45, 46] 都研究了这种问题,文献 [41] 还专门列举了它在数值分析、微分方程和动力系统、参数估计、神经网络、化学、机械系统、通讯、电子电气工程、医学成像等等一系列领域中的应用。我们统称这种类型的问题为可分非线性最小二乘问题。

直观上如果我们知道了非线性参数 x, 线性参数 c 就可以通过解线性最小二乘问题:

$$c = \Phi(x)^{\dagger} y(x) \tag{5.7.2}$$

得到, 其中 $\Phi(x)^{\dagger}$ 是 $\Phi(x)$ 的 Moore-Penrose 广义逆。然后将 (5.7.2) 代入 (5.7.1),则相当于求解:

$$\hat{\phi}(x) = \frac{1}{2} \| (I - \Phi(x)\Phi(x)^{\dagger})y(x) \|_{2}^{2} = \frac{1}{2} \| P_{\Phi}^{\perp} y(x) \|_{2}^{2} \quad . \tag{5.7.3}$$

这个函数称为变量投影函数, 其中 $P_{\Phi}^{\perp} = P_{\Phi}^{\perp}(x)$ 正交投影到 $\Phi(x)$ 的零空间。

变量投影算法 (Variable Projection Method, VP) 基本上分为两步,首先极小化变量投影函数 (5.7.3) 得到非线性参数 x^* ,然后用 x^* 通过式 (5.7.2) 恢复参数 c^* 。下面我们给出它的算法框架:

算法5.7.1. 变量投影算法 (VP)

步 1 给出初始值 xo

步 2 求解非线性最小二乘问题:

$$\min_{x} \hat{\phi}(x) = \frac{1}{2} \| (I - \Phi(x)\Phi(x)^{\dagger}) y(x) \|_{2}^{2} = \frac{1}{2} \| P_{\Phi}^{\perp} y(x) \|_{2}^{2}$$

得到 x*

步 3 由 $c^* = \Phi(x^*)^{\dagger} y(x^*)$ 恢复 c^* 。

变量投影算法的一个明显的优点是步 2 极小化变量投影函数 (5.7.3) 时是在约化空间中求极小,因此收敛速度一般比在整个空间中求极小快。而且无需对线性变量给定初值。[39] 还证明了变量投影算法的解即是原问题 (5.7.1) 的解。

变量投影算法关键是求变量投影函数 (5.7.3) 的极小点,当然也可以用前面提到的 Gauss-Newton 法或 Levenberg-Marquardt 方法或拟牛顿法来解。但 (5.7.3) 中含有广义逆,因此求其 Jacobian 矩阵要稍微复杂些。令 $\bar{r}(x) = P_{\pi}^{\perp} y(x)$ 表示 (5.7.3) 中的残量,J = J(x) 表示 $\bar{r}(x)$

的 Jacobian 矩阵。因此 $\hat{\phi}$ 的梯度 $\nabla \hat{\phi}$ 可以表示为 $\nabla \hat{\phi}(x) = J^T r(x)$ 。[39, 40] 中指出 Jacobian 矩阵 J 的第 j 列可以表示为:

$$J_{.,j}(x) = -[(P_{\Phi}^{\perp} \frac{\partial \Phi}{\partial x_j} \Phi^{\dagger}) + (P_{\Phi}^{\perp} \frac{\partial \Phi}{\partial x_j} \Phi^{\dagger})^T] y(x) + P_{\Phi}^{\perp} \frac{\partial y}{\partial x_j}(x). \tag{5.7.4}$$

那么 J 还可以表示成 J=K+L ,其中 $K_{.,j}=P_{\Phi}^{\perp}[\frac{\partial y}{\partial x_j}(x)-\frac{\partial \Phi}{\partial x_j}\Phi^{\dagger}y(x)]$ 。由于 K 的列属于 Φ^T 的零空间,而 L 的列属于 Φ 的值空间,因此: $K^TL=0,\nabla\hat{\phi}=K^Tr,L^Tr=0$,则有:

$$J^{T}J = K^{T}K + L^{T}L + K^{T}L + KL^{T} = K^{T}K + L^{T}L \quad .$$

Kaufman [42] 建议用 K 来近似 J,即用 K^TK 而不是 $K^TK + L^TL$ 作为 Hessian 阵的近似以减少计算量。

下面我们简单的介绍一下如何计算上面提到的广义逆和投影算子。对 $\Phi(x)$ 做 QR 分解:

$$\Phi = Q \begin{pmatrix} R & \bar{R} \\ 0 & 0 \end{pmatrix} Z \quad ,$$

其中 Q 是正交矩阵 (隐式的表示为一系列 Householder 矩阵和 Givens 变换的乘积), $R \in \mathbb{R}^{\rho \times \rho}$ 是秩为 $\rho \leq l$ 的上三角矩阵, Z 是置换矩阵, 则

$$P_{\Phi}^{\perp} = Q \begin{pmatrix} 0 & 0 \\ 0 & I_{n-\rho} \end{pmatrix} Q^{T} \quad , \quad \Phi^{\dagger} = Z^{T} \begin{pmatrix} R^{-1} \\ 0 \end{pmatrix} \begin{pmatrix} I_{\rho} & 0 \end{pmatrix} Q^{T}$$

我们再考虑带约束的可分非线性最小二乘问题:

min
$$\phi(c,x) = \frac{1}{2} \|\Phi(x)c - y(x)\|_2^2$$

s.t. $H(x)c = g(x)$ (5.7.5)

其中约束

$$H(x)c = g(x) \tag{5.7.6}$$

也具有可分的结构。

为了保证可行点的存在性,我们假设存在向量 x 使得线性系统 (5.7.6) 是相容的,即 $g(x) \in range(H(x))$,或者等价的 $g(x) = H(x)H^{\dagger}(x)g(x)$ 。记所有的这些向量 x 组成的集合为 A。对于固定的 $x \in A$,线性系统 (5.7.6) 的通解可写为:

$$c = H^{\dagger}(x)g(x) + Y(x)z$$
 , (5.7.7)

其中 $H^{\dagger}(x)^T Y(x) = 0, z \in \mathbb{R}^{n-r}, r$ 为 H 的秩。集合 $\{(x,c)|x \in A, c$ 满足(5.7.7)} 就是问题 (5.7.5) 的可行集。因此,问题等价于对:

$$s(z,x) = \|y - \Phi(x)H(x)^{\dagger}g(x) - \Phi(x)Y(x)z\|_{2}^{2}$$

$$= \|y - \zeta(x) - G(x)z\|_{2}^{2}$$
(5.7.8)

求极小,其中 $\zeta(x) = \Phi(x)H(x)^{\dagger}g(x)$, $G(x) = \Phi(x)Y(x)$ 。这就转化成了求解无约束的可分非线性最小二乘问题。

当然对于带盒式约束的可分最小二乘问题:

min
$$\phi(c, x) = \frac{1}{2} \|\Phi(x)c - y(x)\|_2^2$$

s.t. $l_1 \le x \le u_1$

我们同样可以利用第六节给出的投影算子 P(x,l,u) 来解。

§5.8 数值困难

在这一节里,我们说明一些常见的在解最小二乘问题时的数值困难。这些困难可能是由于模型本身的一些性质造成的,也有可能是由于数值算法的缺陷造成的(这里我们不讨论程序上的错误)。了解这些数值困难对我们解决问题,把握问题的本质将很有帮助。

§5.8.1 极小点

在前面的章节中,我们给出了许多解非线性最小二乘问题的算法(包括无约束和带约束问题),但这些算法并不能保证所得到的解就是全局最优解,除非目标函数和可行区域严格凸。在非凸的情况下,没有数学准则可以判别到底是局部极小点还是全局最优解。因而一般来说,能够期望的只是局部极小点,甚至在函数形态良好的情况下也只能找到局部极小点。当然有一些数值算法如模拟退火法,基因算法,随机搜索等等能克服困难,找到全局最优解。然而这些方法常常需要计算很多次函数值,且解带约束的问题可能会有实质的困难。因此,用这些方法来解最小二乘问题可能就不合适。唯一的一些建议是充分探索模型函数的性质,看看可不可以限制可行区域,能不能尽可能的增加试验数据,或者尝试多种不同的优化算法和初始点等等。

在实际应用中,下面两种情况经常导致局部极小点的出现:

- 1. 数据中存在噪声: 当模型无法拟和数据时,根据拟和准则,在试验数据的邻域内可能有不同的最优拟和曲线,这时多个局部极小点就出现了。
- 2. 数据精确,但模型超定: 对于超定的模型,即使数据精确,自由变量变化的多样性 也使得搜索方向的选择上有很大的自由度,因此可能得到很多不同的局部极小点。

§5.8.2 收敛速度慢

有很多因素可能使得最小二乘算法的收敛速度变慢,甚至大大减缓收敛速度。它们可能为:

1. 模型函数表达式病态;

- 2. 冗余的(超定)参数集合;
- 3. 拟和准则(停机准则)上存在误差;
- 4. 试验数据存在误差,内部计算存在误差(如数值积分,微分方程的数值解);
- 5. 找不到好的初始值;
- 6. 数据或参数的尺度很不一致;
- 7. 模型函数不可微:
- 8. 实际中提出的约束可能不相容。

值得注意的是任何一种算法都有其适应范围,当我们遇到收敛速度慢时,就往往需要再仔细研究一下模型,对症下药,直到找到一个满意的解。

§5.8.3 病态性

函数的病态性表现多种多样。首先我们讨论一下在迭代过程或解处 Jacobian 矩阵变为 亏秩或病态的情形。广义的说,我们可以分为五种情况:

- 1. 问题几乎处处是满秩的,但在迭代过程中的某一点处 Jacobian 矩阵亏秩;
- 2. 问题几乎处处是满秩的, 但在解处 Jacobian 矩阵亏秩;
- 3. 问题几乎处处(包括在解处)以一固定的数值秩亏秩:
- 4. 问题几乎处处(包括在解处)是几乎亏秩的;
- 5. 问题病态。

前三种情况比较好解决,如在信赖域框架内求解。当然在解子问题时也可以采用一些特殊的技巧如截断 QR 方法,截断 SVD 方法等等,包括第五节中提出的 M 范数方法也一定程度上克服奇异性。后两种情况就比较困难,因为本身就不好定义什么是好的解。对于线性最小二乘问题的不适定性,已经有了很多很好的工作,我们在本文第一部分中也介绍了不少有效的方法和应用。然而对于非线性最小二乘,仍有很多工作要做。

此外即使是精确的试验数据也会导致极不精确的估计,这种情况当遇到超定的模型时更为严重。差的试验设计可能使事情变得更糟糕,但即便是好的设计和充足的数据有时也不足以消除病态。

笛六音 非线性最小二乘法在热中子测井技术中的应用

专门测量热中子的测井技术有热中子空间测井技术和热中子时间测井技术。如单探 测器中子孔隙度测井和补偿中子测井都是属于空间测井,用的中子源为同位素稳态中子 源,一般用 ²⁴¹Am - Be 中子源。热中子时间测井是利用脉冲中子源的中子脉冲间隔时间 (周期)测量热中子数随时间的分布,即测量热中子时间谱的测井技术。中子脉冲宽度一 般为 $100\mu s$, 中子脉冲的周期 $800 \sim 2150\mu s$ 。该问题是庞巨丰等在 [52] 提出来的。

热中子时间谱测井可以获得热中子的寿命或宏观俘获截面,同时可以确定井眼、侵入 带和地层的热中子寿命及地层孔隙度,包括侵入带的孔隙度。孔隙度的响应主要受地层 含氢指数的影响,对地层颗粒密度不灵敏,特别是提高了用中子孔隙度(用热中子时间谱 解析得到的参数)识别气层的可靠性。

我们在这一章介绍用多参数的数学模型生成热中子时间谱(仿真谱):同时用该数学 模型对仿真的热中子时间谱进行解析。

§6.1 中子时间谱

中子时间谱包括热中子时间谱和超热中子时间谱,它们的数学模型可以由 (6.1.1) 来 描述,两者的区别仅在于谱特征参数的数值不同。

$$n(t) = \begin{cases} A_0 \sum_{i=1}^{P} A_i \int_0^t \mu(t - t')(t - t')^{\alpha_i} e^{-\frac{t - t'}{\tau_i}} dt', & \stackrel{\underline{\Psi}}{=} t < \tau_n \text{ fr}, \\ A_0 \sum_{i=1}^{P} A_i \int_0^{\tau_n} \mu(t - t')(t - t')^{\alpha_i} e^{-\frac{t - t'}{\tau_i}} dt', & \stackrel{\underline{\Psi}}{=} t \ge \tau_n \text{ fr}. \end{cases}$$
(6.1.1)

其中,各参数的意义如下:

$$n(t)$$
: t 时刻的中子记述率, $n(t) \ge 0$; $\mu(t)$: 单位阶越函数, $\mu(t) = \begin{cases} 0, & t < 0, \\ 1, & t \ge 0. \end{cases}$

 A_0 : 归一化系数, $A_0^{-1} = \int_0^\infty n(t)dt$;

 A_i : 第 i 中介质的贡献份额。 $A_i \geq 0$;

 α_i : 与第 i 种介质的中子减速能力相关的参数;

 τ_i : 与第 i 种介质的热中子寿命或超热中子寿命:

 τ_n : 快中子脉冲宽度。

热中子时间谱的解析是指在已知谱数据 n(t) 的情况下, 识别谱参数集

$$x = \{A_1, \alpha_1, \tau_1, \dots, A_i, \alpha_i, \tau_i, \dots, A_P, \alpha_P, \tau_P\},$$
(6.1.2)

其中, τ_i 是最重要的,其次是 A_i 。首先我们来考虑仿真谱,即时间谱的第 j 道计数 N_j 通过下列两步生成:

1. 先按数学模型 (6.1.1) 依给定的谱特征参数 x 生成该道的计数平均数 \bar{N}_i , 即:

$$\bar{N}_j = n(t_j)\Delta t_j$$
 , $j = 1, \cdots, M$,

上式中, \bar{N}_j 是第 j 道计数平均值; $n(t_j)$ 表示当 $t=t_j$ 时,由模型 (6.1.1) 计算给出的计数率; Δt_j 表示第 j 道的道宽。该道宽的中心点位于 $t=t_j$ 处。各道宽可能相等,也可能不等;M 表示总道数,M<256。

2. 然后再以均值和方差都是 \bar{N}_j 的柏松分布 ($\bar{N}_j < 15$ 时) 或正态分布 ($\bar{N}_j \ge 15$ 时) 为概率密度分布函数随机抽样得到 N_j 。或者简单的我们直接给 \bar{N}_j 加上一定水平的噪声得到 N_j 。

这是一个典型的数据拟合问题,我们的目的是要根据中子计数 N_j 反演恢复谱特征参数 A_i 、 α_i 和 τ_i 。不妨令:

$$r_j(x) = n(t_j) - \frac{N_j}{\Delta t_j}$$
 , $j = 1, \dots, M$, (6.1.3)

其中 x 由式 (6.1.2) 给出,记 $r(x) = (r_1(x), \cdots, r_M(x))^T$ 。于是通过求解非线性最小二乘问题:

$$\min f(x) = \frac{1}{2} ||r(x)||_2^2 \quad . \tag{6.1.4}$$

就可以得到所需要的谱参数。

为了计算上的方便,我们先来对函数 n(t) 进行一些分析。注意到如果做变量代换 s = t - t',那么:

$$n(t) = \begin{cases} A_0 \sum_{i=1}^{P} A_i \int_0^t s^{\alpha_i} e^{-\frac{s}{\tau_i}} ds, & \stackrel{\text{def}}{=} t < \tau_n \text{ ft}, \\ A_0 \sum_{i=1}^{P} A_i \int_{t-\tau_n}^t s^{\alpha_i} e^{-\frac{s}{\tau_i}} ds, & \stackrel{\text{def}}{=} t \ge \tau_n \text{ ft}. \end{cases}$$
(6.1.5)

我们还可以显式的计算出 n(t) 对参数 A_i , α_i , τ_i 的偏导:

$$\frac{\partial n(t)}{\partial A_i} = \begin{cases}
A_0 \sum_{i=1}^P \int_0^t s^{\alpha_i} e^{-\frac{s}{\tau_i}} ds, & \stackrel{\text{"}}{=} t < \tau_n \text{ iff}, \\
A_0 \sum_{i=1}^P \int_{t-\tau_n}^t s^{\alpha_i} e^{-\frac{s}{\tau_i}} ds, & \stackrel{\text{"}}{=} t \ge \tau_n \text{ iff}.
\end{cases}$$
(6.1.6)

$$\frac{\partial n(t)}{\partial \alpha_{i}} = \begin{cases}
A_{0} \sum_{i=1}^{P} A_{i} \int_{0}^{t} s^{\alpha_{i}} \ln(s) e^{-\frac{s}{\tau_{i}}} ds, & \stackrel{\underline{\square}}{=} t < \tau_{n} & \stackrel{\underline{\square}}{=}, \\
A_{0} \sum_{i=1}^{P} A_{i} \int_{t-\tau_{n}}^{t} s^{\alpha_{i}} \ln(s) e^{-\frac{s}{\tau_{i}}} ds, & \stackrel{\underline{\square}}{=} t \ge \tau_{n} & \stackrel{\underline{\square}}{=}.
\end{cases} (6.1.7)$$

$$\frac{\partial n(t)}{\partial \tau_i} = \begin{cases}
A_0 \sum_{i=1}^{P} \frac{A_i}{\tau_i^2} \int_0^t s^{\alpha_i + 1} e^{-\frac{s}{\tau_i}} ds, & \stackrel{\underline{\square}}{=} t < \tau_n \text{ iff}, \\
A_0 \sum_{i=1}^{P} \frac{A_i}{\tau_i^2} \int_{t - \tau_n}^t s^{\alpha_i + 1} e^{-\frac{s}{\tau_i}} ds, & \stackrel{\underline{\square}}{=} t \ge \tau_n \text{ iff}.
\end{cases}$$
(6.1.8)

如果取等距节点:

$$\Delta t_1 = \Delta t_2 = \dots = \Delta t_M = \Delta t \quad ,$$

而且如果还有 $\tau_n = (N_{\tau_n} + \frac{1}{2})\Delta t$,其中 N_{τ_n} 是正整数 (这在试验上是容易做到的),另记 $t_0 = 0$,那么:

$$n(t_{j}) = \begin{cases} n(t_{j-1}) + A_{0} \sum_{i=1}^{P} A_{i} \int_{t_{j-1}}^{t_{j}} s^{\alpha_{i}} e^{-\frac{s}{\tau_{i}}} ds, & \stackrel{\underline{\text{w}}}{=} t_{j} < \tau_{n} \text{ ft}, \\ n(t_{j-1}) - n(t_{j-N_{\tau_{n}}}) + A_{0} \sum_{i=1}^{P} A_{i} \int_{t_{j-1}}^{t_{j}} s^{\alpha_{i}} e^{-\frac{s}{\tau_{i}}} ds, & \stackrel{\underline{\text{w}}}{=} t_{j} \ge \tau_{n} \text{ ft}. \end{cases}$$
(6.1.9)

事实上式 (6.1.5) 至式 (6.1.8) 中的积分无法显式积出 (有点类似 Gamma 积分),计算起来比较困难。每计算一次残量 r(x) 和它的 Jacobian 矩阵,共需计算 $3M \times P$ 次此类型的积分。因此即使 M 取成几十时,总的计算量也是相当的大。倘若能在试验设计稍做些调整,利用关系式 (6.1.9) 就可以将每次积分的运算量降下来,而不再需要重新在整个区间上进行计算。

此外偏导 (6.1.8) 要除以 τ_i^2 ,而 τ_i 的一般在 $10 \sim 1000$ 之间,这样使得它在量级上比偏导 (6.1.6) 和 (6.1.7) 往往要小很多,很容易导致 Jacobian 矩阵出现奇异。如果做变量代换

$$\theta_i = \frac{1}{\tau_i} \quad , \quad \tau_i \neq 0 \quad , \quad i = 1, \dots, P \quad ,$$
 (6.1.10)

那么 n(t) 等价于:

而且

$$\frac{\partial \hat{n}(t)}{\partial \theta_i} = \begin{cases}
A_0 \sum_{i=1}^{P} -A_i \int_0^t s^{\alpha_i + 1} e^{-s\theta_i} ds, & \stackrel{\text{def}}{=} t < \tau_n \text{ pt}, \\
A_0 \sum_{i=1}^{P} -A_i \int_{t - \tau_n}^t s^{\alpha_i + 1} e^{-s\theta_i} ds, & \stackrel{\text{def}}{=} t \ge \tau_n \text{ pt}.
\end{cases}$$
(6.1.12)

如此变换之后,偏导 $\frac{\partial \hat{n}(t)}{\partial \theta_i}$ 与 $\frac{\partial \hat{n}(t)}{\partial A_i}$, $\frac{\partial \hat{n}(t)}{\partial \alpha_i}$ 在量级上就不会差很多。

这样我们可以重新考虑目标函数,直接取:

$$\bar{r}_j(x) = \bar{n}(t_j) - \hat{N}_j \quad , \quad j = 1, \cdots, M \quad ,$$
 (6.1.13)

其中:

$$\bar{n}(t_j) = A_0 \sum_{i=1}^{P} A_i \int_{t_{j-1}}^{t_j} s^{\alpha_i} e^{-s\theta_i} ds \quad , \tag{6.1.14}$$

而 \hat{N}_j 是由式 (6.1.14) 模拟得到的 (真实测量的数据经过简单的变换也能化成这个形式)。 此外,残量 $\bar{r}_j(x)$ 一般比 $r_j(x)$ 来得小,因此再用最小二乘算法在计算上也比较有优势。

§6.2 中子时间谱分析

实际上,即使目标函数化成了式 (6.1.13) 的形式,求解优化问题:

$$\min \bar{f}(x) = \frac{1}{2} \|\bar{r}(x)\|_2^2 \quad , \quad \not \exists \, \dot{T} \quad \bar{r}(x) = (\bar{r}_1(x), \cdots, \bar{r}_M(x))^T \quad , \tag{6.2.1}$$

还是有一定的困难。因为每一个介质对中子计数率的贡献在表达式上可互换的,即都是积分:

$$A_i \int_{t_{j-1}}^{t_j} s^{\alpha_i} e^{-s\theta_i} ds \quad . \tag{6.2.2}$$

因此若一组参数:

$$x_P = \{A_1, \alpha_1, \theta_1, \cdots, A_i, \alpha_i, \theta_i, \cdots, A_P, \alpha_P, \theta_P\}$$

$$(6.2.3)$$

是问题 (6.2.1) 的解,则只需将不同介质参数的顺序调整一下,如第 P 个介质的参数与 1 个介质的参数交换一下顺序,其它的顺序不变,则新的参数集:

$$\bar{x}_P = \{A_P, \alpha_P, \theta_P, \cdots, A_i, \alpha_i, \theta_i, \cdots, A_1, \alpha_1, \theta_1\}$$

$$(6.2.4)$$

也是问题 (6.2.1) 的解。也就是说模型本身是超定的。我们在第五章第八节提到过,如果模型超定,则最小二乘问题可能会有很多极小点。而且积分 (6.2.2) 的被积函数是指数函数的形式,这就给数值求解问题 (6.2.1) 带来了很大的困难。值得注意的是即使求解类似:

$$\min \sum_{i=1}^{m} \left(\sum_{i=1}^{P} x_{2i-1} e^{x_{2i}t_j} - y_j\right)^2 \tag{6.2.5}$$

的问题,在很多情况下也会有很大的困难[34,41,42,45,1]。

当然由于该问题是一个实际的物理系统,其参数在一定范围内才有意义,所以我们可以对变量加界约束,即考虑问题:

min
$$\bar{f}(x) = \frac{1}{2} ||\bar{r}(x)||_2^2$$

s.t. $l_{A_i} \leq A_i \leq u_{A_i}$, $i = 1, \dots, P$

$$l_{\alpha_i} \leq \alpha_i \leq u_{\alpha_i}$$

$$l_{\theta_i} \leq \theta_i \leq u_{\theta_i}$$

$$(6.2.6)$$

这样把变量限制在一定范围内,可以在一定程度上减少极小点的出现。

为了避免参数可交换的情况,我们还可以对 α_i 或 θ_i 添加序约束。试验表明 θ_i 对算法的性态影响比较大,我们把序约束加在 θ_i 上,即:

$$l_{\theta} < \theta_1 < \dots < \theta_i < \dots < \theta_P < u_{\theta} \quad . \tag{6.2.7}$$

约束 (6.2.7) 看起来比较复杂,但通过简单的变换,仍可以化成简单的界约束。我们令:

$$\beta_1 = \theta_1 - l_\theta$$
, $\beta_2 = \theta_2 - \theta_1$, \cdots , $\beta_P = \theta_P - \theta_{P-1}$

则约束 (6.2.7) 等价于:

$$\beta_i > 0 \quad , \quad i = 1, \cdots, P \tag{6.2.8}$$

$$\sum_{i=1}^{P} \beta_i + l_{\theta} < u_{\theta} \tag{6.2.9}$$

实际上问题 (6.2.6) 中关于 θ_i 的界约束完全可以用式 (6.2.7) 来代替,而且更条件更强。

我们还是从直观上再来感觉一下函数 (6.1.13) 和 (6.1.14)。首先考察在精确参数下,函数 $\bar{n}(t_j)$ 随时间 t_j 的变化关系。图 6.2 给出了一个介质的时候,中子计数率 $\bar{n}(t_j)$ 随时间 t_j 的变化情形。在这三幅图中, $A_1=1,\alpha_1=0.07$,但在第一幅图中 $\tau_1=30$,第二幅图中 $\tau_1=100$,第三幅图中 $\tau_1=500$ 。并且我们给 $\bar{n}(t_j)$ 简单的加上水平分别为 0.005,0.01 的噪声,其中光滑的曲线为精确输出,带 * 的为噪声水平 0.05 的输出,点线为噪声水平为 0.01 的输出。可以看出当 τ_1 比较小时, $\bar{n}(t_j)$ 的衰减速度比较快,而 τ_1 比较大时,噪声影响比较大。对于多个介质的情形,函数 $\bar{n}(t_j)$ 随时间 t_i 的变化关系还是类似,我们这里就不再列出图像。

我们再来考察函数 $\bar{f}(x) = \frac{1}{2} \|\bar{r}(x)\|_2^2$ 随着参数 α_1, τ_1 的变化关系。模拟过程如下: 先给定一组精确参数产生 \hat{N}_j ,然后取 $A_1 = 1$,让 α_1 在区间 [0.001, 0.1] 之间变化,而让 τ_1 在区间 [10, 1000] 之间变化,计算对应的 \bar{f} ,生成三维图像。在图 6.2 中,第一幅图的 \hat{N}_j 由参数 $A_1 = 1, \alpha_1 = 0.07, \tau_1 = 100$ 生成;第二幅图的 \hat{N}_j 由参数 $A_1 = 1, \alpha_1 = 0.07, \tau_1 = 500$ 生成;第一幅图的 \hat{N}_j 由参数 $A_1 = 1, \alpha_1 = 0.07, \tau_1 = 800$ 生成。从这几幅图来看, $\bar{f}(x)$ 在极小点附近是一条狭长的谷,谷底顺着 α 方向,而顺着 τ 方向在极小点两边是陡的悬崖,当然随着精确解的大小,悬崖的陡峭程度不一样。

₹6.3 数值试验

下面我们针对问题 (6.2.6), 即:

$$\min \quad \bar{f}(x) = \frac{1}{2} \|\bar{r}(x)\|_2^2$$
s.t.
$$l_{A_i} \le A_i \le u_{A_i} \quad , i = 1, \cdots, P$$

$$l_{\alpha_i} \le \alpha_i \le u_{\alpha_i}$$

$$l_{\theta_i} \le \theta_i \le u_{\theta_i}$$

进行数值试验,为了叙述的方便,我们仍记参数集为:

$$X_P = \{A_1, \alpha_1, \tau_1, \cdots, A_i, \alpha_i, \tau_i, \cdots, A_P, \alpha_P, \tau_P\},\$$

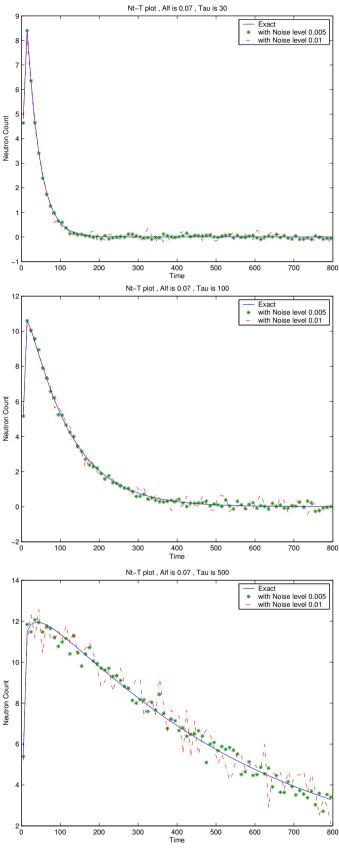
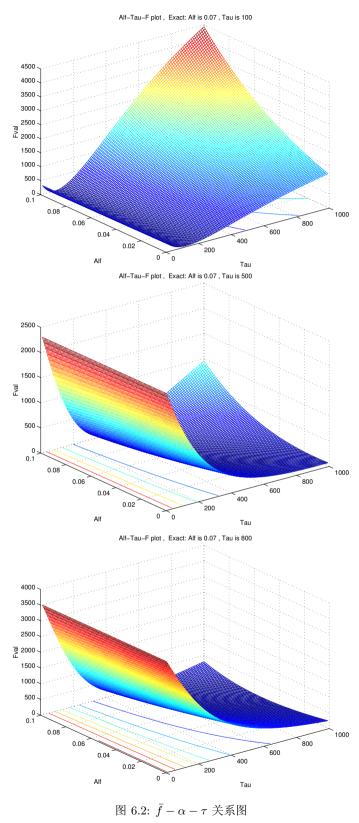


图 6.1: 中子计数率一时间关系图

这三幅图中, $A_1=1,\alpha_1=0.07$,但第一幅图中 $\tau_1=30$,第二幅图中 $\tau_1=100$,第三幅图中 $\tau_1=500$



在第一幅图中, \hat{N}_j 由参数 $A_1=1,\alpha_1=0.07,\tau_1=100$ 生成;在第二幅图中, \hat{N}_j 由参数 $A_1=1,\alpha_1=0.07,\tau_1=500$ 生成;在第一幅图中, \hat{N}_j 由参数 $A_1=1,\alpha_1=0.07,\tau_1=800$ 生成;

只不过具体计算的时候参数是经过变换 (6.1.10) 了的。我们使用 Port 数学库 [38] 中的程序 N2GB 进行计算,它是 NL2SOL 的一个变形,根据迭代表现自动选择到底走拟牛顿步还是 Gauss-Newton 步或 Lenvenberg-Marquardt 步。

总的来说,对于精确的数据,在介质控制在3个以内的时候下,计算效果一般比较好。 尤其是只有一个介质的时候,一般在10步以内就找到所要模拟的精确解。但由于问题的 病态性,我们还是有一些困难。下面主要从极小点,模型超定和初值选取等几个方面进行 一些说明。

a) 极小点存在

由于问题的特殊形式,可能有极小点的存在。这是目前任何一个优化算法避免不了的。

例6.3.1. 用三个介质的参数

$$X_3^* = \{0.5385, 0.03, 100, 0.3846, 0.06, 200, 0.0769, 0.08, 400\}$$

产生试验数据。初始点取为:

$$X_3^0 = \{0.5392, 0.02, 100.0096, 0.3624, 0.04, 298.4192, 0.0984, 0.06, 397.1700\}$$
.

经过6次迭代,目标函数值为 2.5312×10^{-12} ,但前后两次迭代点之间的误差达到 2.13×10^{-14} . 迭代中止,此时解为:

 $\hat{X}_3 = \{0.5410, 0.0393, 100.0604, 0.3992, 0.0586, 301.2279, 0.0599, 0.1000, 404.6759\}$

此时梯度的模为: $\|g_{\hat{X}_3}\| = 1.5 \times 10^{-7}$, 因此可以认为 \hat{X}_3 是一个极小点。

b) 模型超定

还是由于问题的特殊形式,模型可能是超定的。这里暂不考虑函数形式可交换。**例6.3.2.** 用三个介质的参数

$$X_3^* = \{0.5385, 0.03, 100, 0.3846, 0.06, 200, 0.0769, 0.08, 400\}$$

产生试验数据。只取两个介质逼进,初始点取为:

$$X_2^0 = \{0.3750, 0.02, 150, 0.25, 0.04, 450\}$$

,经过 11 次迭代,目标函数值为 4.195×10^{-5} ,但两次相邻迭代的函数值的相对误差达到 6.5×10^{-11} , 迭代中止,此时解为:

$$\hat{X}_2 = \{0.7355, 0.0242, 112.8515, 0.2751, 0.0843, 277.0748\}$$

此时梯度的模为: $\|g_{\hat{X}_2}\| = 4.9 \times 10^{-7}$ 。这说明即使是精确的数据,但由于参数的自由度过大,仍可能得到非常不同的解。

例6.3.3. 在上例中, 我们用 2 个介质的参数就可以模拟 3 个介质的数据。事实上, 如果我们在上例解 \hat{X}_2 的基础上加入第 3 个介质, 其参数为: $\{A_3=0,\alpha_3,\tau_3\}$, 其中 α_3,τ_3 可为任意实数, 我们就能得到任意多的极小点: $\hat{X}_3=\{\hat{X}_2,0,\alpha_3,\tau_3\}$

c) 初值影响

对于基于梯度的算法,不同的初值,将极大的影响算法的表现。即使是精确的数据, 精确的梯度,坏的初值将显著的减慢收敛速度,甚至使算法不能收敛到极小点。

例6.3.4. 用参数 $X_2^* = \{0.5833, 0.04, 200, 0.4167, 0.06, 300\}$ 产生试验数据。初始点取为: $X_2^{01} = \{0.6, 0.02, 50, 0.4, 0.04, 450\}$,经过 284 次迭代绝对收敛到 X_2^* ,此时函数值和 Jacobi 矩阵的计算次数分别为 342 次和 285 次。而初始点取为: $P_2^{02} = \{0.6, 0.02, 190, 0.4, 0.04, 350\}$,经过 79 次迭代绝对收敛到 X_2^* ,此时函数值和 Jacobi 矩阵的计算次数分别为 92 次和 80 次。可以看出迭代次数和函数值,Jacobi 矩阵的计算次数减少了将近 4 倍。

实际上我们可以分为两个阶段:第一阶段给定一初始值 X^0 ,先调用最小二乘算法进行计算得到 \hat{X}^0 。经过这次调用,某些参数可能得到了更好的估计。第二个阶段,我们可以用 \hat{X}^0 的一部分替换 X^0 作为新的初始值再次调用最小二乘算法进行计算。由于第一次计算只是计算一个估计值,我们可以控制迭代步数,如取 10 步,或者减小收敛的判断条件。经过实际计算,我们发现这种调整对于两个介质的情况特别有效,基本上能在 100 个迭代步内算得所取的测试精确解(最优解)。三个介质时调整增加第一阶段的迭代步,很多情况也能取得较好的效果。

例6.3.5. 对于上例的测试数据,我们控制第一阶段的函数值的计算次数最多不超过为 10次,当初始值取 X_2^{01} 时,第二阶段只需经过 5次迭代就绝对收敛到 X_2^* ,此时函数值和 Jacobi 矩阵的计算次数分别为 6次和 6次。两个阶段的迭代次数之和也只有 15次。当初始值取 X_2^{02} 时,第二阶段只需经过 6次迭代就绝对收敛到 X_2^* ,此时函数值和 Jacobi 矩阵的计算次数分别为 7次和 7次。两个阶段的迭代次数之和也只有 14次。可以看出经过两次调用,迭代次数和函数值,Jacobi 矩阵的计算次数大大的减少了。

§6.4 讨论和总结

在这一章里我们研究了最小二乘法在热中子时间谱中的应用。着重分析了该问题的一些特性,如目标函数的选取,变量代换,及约束的添加等等。最后我们做了一些数值算例。对于精确的数据,在介质控制在3个以内的时候下,计算效果一般比较好。但由于超定的模型函数以及极小点的存在,使得计算上还是存在些一些困难,尤其是对于带噪声的数据。这是我们以后还要考虑的重点。

第七章 结论与展望

本论文主要研究了最小二乘问题中的一些新的内容。

在线性最小二乘问题方面,我们重点研究了其不适定性,回顾了经典的正则化方法,提出了截断 SVD 分解算法和带信赖域技巧的截断 Lanczos 方法。其中在信号处理的应用中,我们研究了傅立叶矩阵的特殊结构,得到了傅立叶矩阵实部和虚部的一个简单易行的奇异值分解,从而使得截断 SVD 分解算法在实际计算上成为可能,数值试验表明我们的算法是有效可行。而在数字图像恢复的应用中,我们利用带信赖域技巧的截断 Lanczos 方法进行求解。同时根据二次模型函数的特点对信赖域半径的调整和试探步的接受准则进行了一些有意义的调整。在遥感图像恢复中的应用表明我们的这一方法是成功的。

在非线性最小二乘问题方面,我们在信赖域框架内分析了 Levenberg-Marquardt 方法和拟牛顿法。接着对于解非线性最小二乘的信赖域子问题,我们提出了 M 范数方法,化成一个等价的对角系统来解。此外还概述了带约束的非线性最小二乘问题的一些解法。特别的介绍了可分非线性最小二乘问题。最后还分析了最小二乘问题的一些数值困难。作为应用,我们介绍了热中子时间谱测井谱的解析。

但研究过程中,我们感到仍有很多问题需要解决,下面我们只列举一些:

- 1. 如何更有效的求解线性不适定最小二乘问题:
- 2. 在具体的应用当中,如何有效的利用问题的特殊结构;
- 3. 关于非线性最小二乘问题的不适定性需要做进一步研究;
- 4. 如何设计带约束非线性最小二乘问题更有效的算法:
- 5. 对于目标函数不能显式表达时,如是一个动力系统或微分方程的解,如何有效求解;
- 6. 对于超定问题需做进一步研究。

我们将在以后的工作中继续考虑这些问题。

参考文献

- [1] G.A.F. Seber and C.J.Wild, Nonlinear Regression, Wiley Interscience, New York, 1989
- [2] Golub G.H. and Van Loan C.F., *Matrix Computations (Third Edition)*, The Johns Hopkins University Press, 1996.
- [3] 徐树方,《矩阵计算的理论与方法》,北京大学出版社,1995
- [4] D.Colton, R.Kress, Integrel Equation Methods for Scattering Theory, John Wiley, New York, 1983
- [5] I.J.D.Craig, J.C.Brown, Inverse Problems in Astronomy, Adam Hilger, Bristol, 1986
- [6] Carrara W.G., Goodman R.S. and Majewski R.M., Spotlight synthetic Aperture Rada Signal Processing Algorithms, Boston, MA: Artech House, 1995.
- [7] Çetin M. and Karl W.C., Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization, *IEEE Trans. Image Processing*, Vol.10, No.4, pp.623-631, 2001.
- [8] Degraaf S.R., SAR imaging via mordern 2-D spectral estimation methods, *IEEE Trans. Image Processing*, Vol.17, No.5, pp.729-761, 1998.
- [9] Gerchberg R.W., Super-resolution through energy reduction, Optica Acta, Vol.21, No.9, pp.709-720, 1974.
- [10] Hu X., Levin D.N., Lauterbur and Spraggins T.A., SLIM: Spectral localization by imaging, Magnetic Resonance in Medicine, Vol.8, pp.314-322, 1988.
- [11] Jain A.K. and Ranganath S., Extrapolation algorithms for discrete signals with application in spectral estimation, IEEE Trans. Acoust. Speech Signal Process. ASSP-29, pp.830-845, 1981.
- [12] Li J. and Stoica P., An adaptive filtering approach to spectral estimation and SAR imaging, IEEE Trans. Signal Processing, Vol. 44, pp.1469-1484, 1996.
- [13] Reeves S.J. and Heck L.P., Selection of obervations in signal reconstruction, IEEE Transactions on Signal Processing, Vol.43, No.3, pp.788-791, 1995.
- [14] Skodras A.N., Fast Discrete Cosine Transform Pruning, IEEE Transaction on Signal Processing, vol.42, no.7, pp.1833-1837, 1994.
- [15] Sullivan B.J. and Liu B., On the use of singular value decomposition and decimation in discrete-time bandlimited signal extrapolation, IEEE Trans. Acoust. Speech Signal Process. ASSP-32, pp.1201-1212, 1984.
- [16] Sun M.T. and Pao I.M., Statistical computation of discrete cosine transform in video encoders, Journal of Visual Communication and Image Representation, Vol.9, No.2, pp.163-170, 1998.
- [17] Varah J. M., On the numerical solution of ill-conditioned linear systems with application to ill-posed problems, SIAM J. Numer. Anal., Vol.10, pp.549-565, 1973.
- [18] Walsh D. O. and Nielsen-Delaney P. A., Direct method for superresolution, Journal of the Optical Society of America, A Vol.11, pp.572-579, 1994.
- [19] Wingham D.J., The reconstruction of a band-limited function and its Fourier transform from a finite number of samples at arbitrary locations by singular value decomposition, *IEEE Transactions on Signal Processing*

参考文献 80

- , Vol.40, pp.559-570, 1992.
- [20] 王彦飞,数字图像信息处理的心方法,博士后论文,中国科学院遥感应用研究所,2003.7
- [21] 王彦飞,反问题的优化与正则化方法,博士论文,中国科学院数学与系统科学研究院,2002.5
- [22] 肖庭延、于慎根、王彦飞、《反问题的数值解法》,科学出版社,2003
- [23] M.Roggemann and B.Welsh, Imaging through Turbulence, Boca Raton: CRC Press, 1996.
- [24] M.Hanke, Iterative Regularization Techniques in Image Reconstruction, in Proceedings of the Conference Mathematical Methods in Inverse Problems for Partial Differential Equations, Mt.Holyoke: Springer-Verlag, 1998.
- [25] C.R. Vogel and M.E. Oman, Fast Robust Total Variation-based Reconstruction of Noisy, Blurred Images, IEEE Transactions on Image Processing, Vol. 7, pp.813-824, 1998.
- [26] A.N.Tikhonov and V.Y. Arsenin, Solutions of Ill-Posed Problems, New York: Wiley, 1977.
- [27] Y.F.Wang, Y.X.Yuan and H.C.Zhang, A Trust Region-CG Algorithm for Deblurring Problem in Atmospheric Image Reconstruction, Science in China, Vol. 45, pp.731-740, 2002.
- [28] Y.F.Wang, On the Regularity of Trust Region-CG Algorithm: with Application to Image Deconvolution Problem, Science in China, Vol. 46, pp.312-325, 2003.
- [29] T.Steihaug, The Conjugate Gradient Method and Trust Regions in Large Scale Optimization, SIAM J.Numer.Anal., Vol 20, pp.626-637, 1983.
- [30] P.L.Toint, Towards an Efficient Sparsity Exploiting Newton Method for Minimization, in I.Duff ed.: Sparse Matrices and Their Uses, Berlin: Academic Press, pp.57-88, 1981.
- [31] N.Gould, S.Lucidi, M.Roma and Ph.L.Toint, Solving The Trust Region Subproblem Using The Lanczos Method, SIAM Journal on Optimization Vol. 9, pp. 504-525, 1999.
- [32] Raymond H.Chan and Michael K.Nagy, Conjugate Gradient Methods For Toeplitz Systems, SIAM Review Vol.38.No.3, pp.427-482, 1996.
- [33] P.C.Hansen, Regularization Tools: A Matlab Package for Analysis and Solution of Discrete Ill-Posed Problems, *Numer. Algorithms*, Vol. 6, pp.1-35, 1994. Software is available from Netlib as well as from the author's homepage: http://www.imm.dtu.dk/~pch.
- [34] Klaus Schittkowski, Numerical Data Fitting in Dynamical Systems, Kluwer Academic Publishers, 2002
- [35] J.E.Dennis, Jr., D.M.Gay, and R.E.Welsch, An Adaptive Nonlinear Least Squares Algorithm, ACM Trans. Math. Software 7(1981), pp.348-368
- [36] Per Lindström, Gauss-Newton Based Algorithm for Constrained Nonlinear Least Squares Problems, Report, Institute of Information Processing, University Umea, Sweden
- [37] R.J. Hanson, Fred T.Krogh, A Quadratic-Tensor Model Algorithm for Nonlinear Least-Squares Problems with Linear Constraints, ACM Transaction on Mathematical Software, Vol.18, No.2, 115-133
- $[38] \begin{tabular}{l} The Port Mathematical Subroutine Library \\ , AT\&T Bell Laboratories, Murray Hill, NJ.1984. \\ Third edition \\]$
- [39] G.H.Golub and V.Pereyra, The Differentiation of Pseudo-Inverses and Nonlinear Least-Squares Problems

参考文献 81

- Whose Variables Separate, SIAM J.Nuer.Anal. 10(1973), pp.413-432
- [40] G.H.Golub and V.Pereyra, The Differentiation of Pseudo-Inverses, Separable Nonlinear Least-Squares Problems and Other tales, pp.303-324 in *Generalized Inverses and Applications*, ed.M.Z.Nashed, Academic Press(1976)
- [41] G.H.Golub and V.Pereyra, Separable nonlinear least squares: the variable projection method and its applications, *Inverse Problems*, VOL.19, NO.2(April 2003), R1-R26
- [42] L.Kaufman, A Variable Projection Method for Solving Separable Nnonlinear Least Squares Problems, BIT 15(1975), pp.49-57
- [43] David M.Gay, L.Kaufman, Tradeoffs in Algorithms For Separable Nonliear Least Squares, technical report.
- [44] L.Kaufman and Victor Pereyra, A Method for Separable Nonlinear Least Squares Problems with Separable Nonlinear Equality, SIAM J.Nuer.Anal. VOL.15,NO.1,February(1978),pp.12-20
- [45] H. B. Nielsen, Separable NonLinear Least Squares. Report IMM-REP-00-01, Department of Mathematical Modelling, DTU (2000), 16 pages.
- [46] Ruhe, A., and P.A. Wedin, Algorithms for separable nonlinear least squares problems, SIAM RIEW, 22:318-337(1980)
- [47] A.R.Conn, N.I.M.Gould and P. L.Toint, Trust Region Methods, CGT publications, 1999.
- [48] Douglas M.Bates and Donald G.Watts, Nonlinear Regression Analysis and It's Applications, JOHN WILEY & SONS, Inc. , 1988
- $\left[49\right]$ Jorge Nocedal, Stephen J.Wright, Numerical Optimization, Springer ,1999
- [50] Wen Z.W, Wang Y.F, New Trust Region Algorithm for Image Restoration, to appear in Science in China
- [51] Wang Y.F, Wen Z.W, Z. Nashed, Q. Sun, A Direct Fast Method for Time-limited Signal Reconstruction, submitted to applied optics
- [52] 庞巨丰,李敏,韩继勇,热中子时间谱测井谱的解析,《测井技术》,第 24 卷,第一期,2000.
- [53] 袁亚湘,孙文瑜,《最优化理论与方法》,科学出版社,1999.
- [54] 陈景良, 陈向辉, 《特殊矩阵》, 清华大学出版社, 2001

致谢

首先,我要诚挚的感谢我的导师袁亚湘研究员。在我攻读硕士学位期间,他一直给予我热心的指导和悉心的关怀,并引导我进入最优化理论与算法这一充满活力的研究领域,使我得以顺利的完成学业与论文的写作。袁亚湘先生渊博的学识、活跃的思维、严谨的治学态度以及把握问题本质的能力给我留下了深刻的印象,先生的睿智、如朋友般亲切的教导将使我终生受益。

其次,我要感谢中科院遥感所的王彦飞副研究员在图像恢复和信号处理等等方面对 我毫无保留的指导。每当我遇到问题请教时他总是优先处理,并给了我很多中肯的意见。 与他的合作很愉快。

十分感谢实验室的张林波研究员、白中治研究员、戴彧虹副研究员和北京邮电大学的 艾文宝教授,和他们的讨论对我帮助甚大。

感谢计算数学所的白英、吴继平、樊建荣、张文琼以及人教处的绍欣等老师对我的关心和帮助。感谢我的同学张磊、孙元昌、何长州、陈闵昕、何平、殷俊锋、靳聪明、戴桂东、王颖、杨军、张一等等平常对我的鼓励和支持,和他们一起学习生活很开心。特别要感谢常珂琪同学在生活上给予的无微不至的关心和帮助,使得我得以安心的学习和顺利的完成学业。

感谢优化讨论班的师兄师姐师弟师妹们,他们有李工农、严涛、周庆华、王丽萍、李 改弟、李多全、王周宏、胡毅庆、夏勇、郑真真、马士谦等等,也感谢已毕业的黑龙、孙 凯、韩兴楼、马国选等师兄,能经常与他们一起探讨问题是我的荣幸。

非常感谢数学与系统研究院、计算数学与工程数学研究所提供的舒适的学习环境和 营造的良好学术氛围以及科学与工程计算国家重点实验室提供的先进设备和优越的上机 条件。我的大部分时间是在实验室度过的。

最后,衷心的感谢我的父母,是他们的辛勤劳动使我能有幸接受高等教育,他们的憨厚,正直,从不为困难所压倒的精神给我打下了深深烙印。感谢我的哥哥,他一直是我学习的榜样,没有他的鼎立支持和这些年来对家庭的支撑,我不可能来北京读研继续从事研究工作。再次衷心感谢所有帮助过我的善良的人们!

攻读硕士学位期间部分工作

- 1. Wen Z.W., Wang Y.F., A trust region method for large scale inverse problems in atmospheric image restotation, accepted by *Proceeding of the 2003' International Conference on Nonlinear Programming and Numerical Linear Algebra*
- Zhouhong Wang , Yaxiang Yuan , Zaiwen Wen , A Subspace Trust Region Method for Large Scale Unconstrained Optimization, accepted by Proceeding of the 2003' International Conference on Nonlinear Programming and Numerical Linear Algebra
- 3. Wen Z.W. , Wang Y.F , A New Trust Region Algorithm for Image Restoration, to appear in *Science* in China
- 4. Wang Y.F , Wen Z.W. , Z. Nashed, Q. Sun , A Direct Fast Method for Time-limited Signal Reconstruction , submitted to *Applied Optics* (SCI impact > 1.4)
- 5. 文再文, 热中子时间谱测井谱的解析, 草稿

个人简历

基本资料

姓 名: 文再文

性别: 男

籍 贯: 湖南省沅江市

电子邮件: wenzw@lsec.cc.ac.cn , wendoublewen@hotmail.com

学习经历

2001.9-2004.7 中国科学院数学与系统科学研究院 硕士

1997.9—2001.7 上海交通大学应用数学系 学士

1997.9 上海交通大学少年班

科研情况

研究兴趣 最优化方法及其应用,数值代数,并行算法,数字图像处理,信号处理

科研项目 (1)《鱼眼镜头成像技术研究》,已完成:该项目为上海交通大学应用

数学系与浦东软件园电子出版社合作项目

(2)《最优化算法在数字图像复原和超分辨率计算中的应用》,进行中,

中国科学院研究生科学创新与社会实践专项资助

学术论文 攻读硕士学位期间已投稿一系列论文

获奖和荣誉情况

2003.9 中科院数学与系统科学研究院首届院长奖学金

2001.7 上海交通大学优秀毕业生

2000.9 上海交通大学优秀奖学金

1999.9 上海交通大学优秀奖学金

1998.9 上海交通大学优秀奖学金