COMPUTATIONAL METHODS FOR LEAST SQUARES PROBLEMS AND CLINICAL TRIALS

A DISSERTATION
SUBMITTED TO THE PROGRAM IN SCIENTIFIC COMPUTING AND COMPUTATIONAL MATHEMATICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Zheng Su
June 2005

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Tze Leung Lai Principal Co-Advisor

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Michael Saunders Principal Co-Advisor

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Walter Murray

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Gene Golub

Approved for the University Committee on Graduate Studies.

Abstract

A common problem facing an analyst is how to interpret data collected from experiments or observations. A mathematical model involving unknown parameters is assumed to explain how a set of observations depends on certain other data. Under various assumptions, the model is used to estimate numerical values for the parameters.

In the first part of the thesis, parameters are estimated by the method of least squares. We study the "optimal backward error" problem of finding the smallest change to the data that would make the estimated parameter values fit the mathematical model exactly. Such knowledge helps judge the sensitivity of the data, the accuracy of the parameter estimates, and the stability of new algorithms for computing estimates.

While it is computationally intensive to compute optimal backward errors directly, an approximate formula has been studied analytically in recent years. We develop dense, sparse, and iterative methods to evaluate this formula numerically. We find that the computed estimate of the optimal backward error is very near the true optimal error. Algorithms for calculating sequences of upper and lower bounds for the estimate are also developed, based on Gauss quadrature theory. Numerical results show that the bounds converge quickly and are therefore useful in practice. This solves a twenty-five year old problem suggested by Stewart and Wilkinson.

When the data collected from experiments are not complete, they are called "censored". In this case, it is more natural to compute confidence intervals for the parameters (rather than estimating the sensitivity of the data).

Clinical trials generate much incomplete data. In the second part of the thesis we study clinical trials with time-to-event endpoints, in which the most important parameters are treatment effect and median survival. We use test-based approaches to compute confidence intervals and confidence regions for those parameters. Such knowledge is crucial for clinicians to make decisions from the results of a trial.

Importance resampling techniques are developed to compute tail probabilities of the tests, thereby reducing the variance of the Monte Carlo estimate of an error probability, and thus the number of simulations required to compute sample size and power in the design stage of a clinical trial, and to construct confidence intervals and regions from the trial's data.

Acknowledgements

First of all, I would like to thank my academic advisors, Professor Tze Leung Lai and Professor Michael Saunders. I feel very fortunate and honored for having had the opportunity to work with them. I thank Prof. Lai and Prof. Saunders from the bottom of my heart for their support, patience, and devotion, and for putting so much effort and time into guiding me.

I would also like to thank Dr. Joe Grcar sincerely for his initiative and guidance. Part of the dissertation is based on a paper with Joe (and Michael), and he brought to me many insightful views that benefited my research and writing.

I would like to thank Professor Walter Murray, the director of SCCM and my committee member, for his tremendous help and guidance throughout my years at Stanford. Prof. Murray made my life at Stanford much easier and happier.

I would like to thank Professor Philip Lavori for being the chair of my oral exam. Prof. Lavori is one of the nicest persons you could ever meet, and I look forward to doing postdoctoral research with him.

I would like to thank Professor Gene Golub, one of my closest friends. Despite Gene's senior position and world fame, he is one of the most kind and generous and down-to-earth persons I have ever met. Part of the dissertation is based on Gene's work on quadrature methods.

Finally, I want to express my deepest gratitude and love for my wonderful parents for their love, care, and support throughout the years, and to my dearest friend Xiao Wei for her additional love and care throughout my student days.

Contents

List of Tables

List of Figures

Part I

Estimates of Optimal Backward Perturbations for Linear Least Squares Problems

Chapter 1

Introduction

> "A great deal of thought, both by myself and by J. H. Wilkinson, has not solved this problem, and I therefore pass it on to you: *find easily computable statistics that are both necessary and sufficient for the stability of a least squares solution.*" — G. W. Stewart [32, pp. 6–7]

The purpose of this work is to examine the usefulness of a certain quantity as a practical backward error estimator for the least squares (LS) problem:

$$\min_x \|Ax - b\|_2 \qquad \text{where} \qquad b \in \mathbb{R}^m \text{ and } A \in \mathbb{R}^{m\times n}.$$

If the arbitrary vector x solves an LS problem for the data $A + \delta A$, then the perturbation δA is called a backward error for x. This name is borrowed from the context of Stewart and Wilkinson's remarks, backward rounding error analysis, which finds and bounds some δA when x is a computed solution. Since x may be chosen arbitrarily, it may be more appropriate to call δA a "data perturbation" or a "backward perturbation" rather than a "backward error." All three names have been used in the literature.

The size of the smallest backward error is

$$\mu(x) \;=\; \min_{\delta A} \|\delta A\|_{\mathrm{F}}\,.$$

A precise definition and more descriptive notation for this are

$$\mu(x) \;=\; \left\{ \begin{array}{l} \text{the size of data perturbation, for matrices in least squares} \\ \text{problems, that is optimally small in the Frobenius norm,} \\ \text{as a function of the approximate solution } x \end{array} \right\} \;=\; \mu_{\mathrm{F}}^{(\mathrm{LS})}(x)\,.$$

This level of detail is needed here only twice, so usually it is abbreviated to "optimal backward error" and written $\mu(x)$. The concept of optimal backward error originated with Oettli and Prager [27] in the context of linear equations.

If $\mu(x)$ can be estimated or evaluated inexpensively, then the literature describes three uses.

1. *Accuracy criterion.* When the data of a problem have been given with an error that is greater than $\mu(x)$, then x must be regarded as solving the problem, to the extent the problem is known. Conversely, if $\mu(x)$ is greater than the uncertainty in the data, then x must be rejected. These ideas originated with John von Neumann and Herman Goldstine [26] and were rediscovered by Oettli and Prager.

2. *Run-time stability estimation.* A calculation that produces x with small $\mu(x)$ is called backwardly stable. Stewart and Wilkinson [32, pp. 6–7], Karlson and Waldén [21, p. 862] and Malyshev and Sadkane [22, p. 740] emphasized the need for "practical" and "accurate and fast" ways to determine $\mu(x)$ for least squares problems.

3. *Exploring the stability of new algorithms.* Many fast algorithms have been developed for LS problems with various kinds of structure. Gu [18, p. 365] [19] explained that it is useful to examine the stability of such algorithms without having to perform backward error analyses of them.

When x is a computed solution, Wilkinson would have described these uses for $\mu(x)$ as "a posteriori" rounding error analyses.

The exact value of $\mu(x)$ was discovered by Waldén, Karlson and Sun [34] in 1995. To evaluate it, they recommended a formula that Higham had derived from their pre-publication manuscript [34, p. 275] [20, p. 405],

$$\mu(x) \;=\; \min\left\{ \frac{\|r\|}{\|x\|}\,,\, \sigma_{\min}[A\;\; B] \right\}, \qquad B \;=\; \frac{\|r\|}{\|x\|}\left(I - \frac{rr^t}{\|r\|^2} \right), \tag{1.1}$$

where $r = b - Ax$ is the residual for the approximate solution, $\sigma_{\min}$ is the smallest singular value of the $m \times (n+m)$ matrix in brackets, and $\|\cdot\|$ means the 2-norm unless otherwise specified. There are similar formulas when both A and b are perturbable. It is interesting to note that a prominent part of these formulas is the optimal backward error of the linear equations $Ax = b$, namely

$$\eta(x) \equiv \frac{\|r\|}{\|x\|} \;=\; \mu_{\mathrm{F}}^{(\mathrm{LE})}(x) \;=\; \mu_{2}^{(\mathrm{LE})}(x)\,. \tag{1.2}$$

The singular value in (1.1) is expensive to calculate by dense matrix methods, so other ways to obtain the backward error have been sought. Malyshev and Sadkane [22] proposed an iterative process based on Lanczos bidiagonalization to approximate $\mu(x)$. Other authors including Waldén, Karlson and Sun have derived explicit approximations for the backward error.

One estimate in particular has been studied in various forms by Karlson and Waldén [21], Gu

[18], and Grcar [17]. It can be written as

$$\widetilde{\mu}(x) \;=\; \left\| \left(\|x\|^2 A^t A + \|r\|^2 I \right)^{-1/2} A^t r \right\| . \tag{1.3}$$

For this quantity:

- Karlson and Waldén showed [21, p. 864, eqn. 2.5 with $y = y_{\text{opt}}$] that, in the notation of this paper,

$$\frac{2}{2+\sqrt{2}}\, \widetilde{\mu}(x) \;\leq\; f(y_{\text{opt}}) \,,$$

  where $f(y_{opt})$ is a complicated expression that is a lower bound for the smallest backward error in the spectral norm, $\mu_2^{(\text{LS})}(x)$. It is also a lower bound for $\mu(x) = \mu_{\text{F}}^{(\text{LS})}(x)$ because $\|\delta A\|_2 \leq \|\delta A\|_{\text{F}}$. Therefore Karlson and Waldén's inequality can be rearranged to

$$\frac{\widetilde{\mu}(x)}{\mu(x)} \;\leq\; \frac{2+\sqrt{2}}{2} \;\approx\; 1.707 \,. \tag{1.4}$$

- Gu [18, p. 367, cor. 2.2] established the bounds

$$\frac{\|r_*\|}{\|r\|} \;\leq\; \frac{\widetilde{\mu}(x)}{\mu(x)} \;\leq\; \frac{\sqrt{5}+1}{2} \;\approx\; 1.618 \,, \tag{1.5}$$

  where r_* is the unique, true residual of the LS problem. He used these inequalities to prove a theorem about the definition of numerical stability for LS problems. Gu derived the bounds assuming that A has full column rank. The lower bound in equation (1.5) should be slightly less than 1 because it is always true that $\|r_*\| \leq \|r\|$, and because $r \approx r_*$ when x is a good approximation to a solution.

- Finally, Grcar [17, thm. 4.4] proved that $\widetilde{\mu}(x)$ asymptotically equals $\mu(x)$ in the sense that

$$\lim_{x \,\to\, x_*} \frac{\widetilde{\mu}(x)}{\mu(x)} \;=\; 1 \,, \tag{1.6}$$

  where x_* is any solution of the LS problem. The hypotheses for this are that A, r_*, and x_* are not zero. This limit and both equations (1.1) and (1.3) do not restrict the rank of A or the relative sizes of m and n.

All these bounds and limits suggest that equation (1.3) is a robust estimate for the optimal backward error of least squares problems. However, this formula has not been examined numerically. It receives only brief mention in the papers of Karlson and Waldén, and Gu, and neither they nor Grcar performed numerical experiments with it. The aim of this paper is to determine whether $\widetilde{\mu}(x)$ is an acceptable estimate for $\mu(x)$ in practice, thereby answering Stewart and Wilkinson's question.

Chapter 2

Evaluating the Karlson and Waldén Estimate

Many ways to solve LS problems produce matrix factorizations that can be used to evaluate $\widetilde{\mu}(x)$ efficiently. If x is obtained in other ways, then the procedures described here still may be used to evaluate $\widetilde{\mu}(x)$ at the extra cost of calculating the factorizations just for this purpose.

2.1 SVD methods

When a singular value decomposition (SVD) is used to solve the LS problem, the economy size decomposition $A = U\Sigma V^t$ may be formed, where Σ and V are square matrices and U has orthonormal columns. With this notation and $\eta = \eta(x)$ in (1.2), it follows that

$$\begin{aligned}
\|x\|\,\widetilde{\mu}(x) &= \left\|\left(A^tA+\eta^2I\right)^{-1/2}A^tr\right\| \\
&= \left\|\left(V\Sigma^2V^t+\eta^2I\right)^{-1/2}V\Sigma\,U^tr\right\| \\
&= \left\|\left[V\left(\Sigma^2+\eta^2I\right)V^t\right]^{-1/2}V\Sigma\,U^tr\right\| \\
&= \left\|V\left(\Sigma^2+\eta^2I\right)^{-1/2}V^tV\Sigma\,U^tr\right\| \\
&= \left\|\left(\Sigma^2+\eta^2I\right)^{-1/2}\Sigma\,U^tr\right\| . \qquad (2.1)
\end{aligned}$$

Calculating $\widetilde{\mu}(x)$ has negligible cost once U, Σ and V have been formed. However, the most efficient SVD algorithms for LS problems accumulate U^tb rather than form U. This saving cannot be realized when U is needed to evaluate $\widetilde{\mu}(x)$. As a result, Table 2.1 shows the operations triple from roughly $2mn^2$ for x, to $6mn^2$ for both x and $\widetilde{\mu}(x)$. This is still much less than the cost of

Table 2.1: *Operation counts for solving LS problems by SVD methods with and without forming* $\widetilde{\mu}(x)$*. The work to evaluate* $\widetilde{\mu}(x)$ *includes that of* r*. Only leading terms are shown.*

task	operations	source
form U, Σ, V by Chan SVD	$6mn^2 + 20n^3$	[13, p. 175]
solve LS given U, Σ, V	$2mn + 2n^2$	
evaluate $\widetilde{\mu}(x)$ by equation (2.1) given U, Σ, V	$4mn + 10n$	
solve LS by Chan SVD	$2mn^2 + 11n^3$	[14, p. 248]

evaluating the exact $\mu(x)$ by equation (1.1) because about $4m^3 + 2m^2n$ arithmetic operations are needed to find all singular values of an $m \times (n+m)$ matrix [13, p. 175].

2.2 The KW problem, QR factors, and projections

Karlson and Waldén [21, p. 864] draw attention to the full-rank LS problem

$$K = \begin{bmatrix} A \\ \frac{\|r\|}{\|x\|} I \end{bmatrix}, \qquad v = \begin{bmatrix} r \\ 0 \end{bmatrix}, \qquad \min_y \|Ky - v\|, \tag{2.2}$$

which proves central to the computation of $\widetilde{\mu}(x)$. It should be mentioned that LS problems with this structure are called "damped", and have been studied in the context of Tikhonov regularization of ill-posed LS problems [4, pp. 101–102]. We need to study three such systems involving various A and r. To do so, we need some standard results on QR factorization and projections. We state these in terms of a full-rank LS problem $\min_y \|Ky - v\|$ with general K and v.

Lemma 1 *Suppose the matrix* K *has full column rank and QR factorization*

$$K = Q \begin{bmatrix} R \\ 0 \end{bmatrix} = YR, \qquad Q = \begin{bmatrix} Y & Z \end{bmatrix}, \tag{2.3}$$

where R *is upper triangular and nonsingular, and* Q *is square and orthogonal, so that* $Y^tY = I$*,* $Z^tZ = I$*, and* $YY^t + ZZ^t = I$*. The associated projection operators may be written as*

$$P = K(K^tK)^{-1}K^t = YY^t, \qquad I - P = ZZ^t. \tag{2.4}$$

Lemma 2 *For the quantities in Lemma 1, the LS problem* $\min_y \|Ky - v\|$ *has a unique solution*

and residual vector defined by $Ry = Y^t v$ and $t = v - Ky$, and the two projections of v satisfy

$$Pv = Ky = YY^t v, \qquad \|Ky\| = \|Y^t v\|, \tag{2.5}$$

$$(I - P)v = \quad t = ZZ^t v, \qquad \|t\| = \|Z^t v\|. \tag{2.6}$$

2.3 QR methods

We now find that $\widetilde{\mu}(x)$ in (1.3) is the norm of a certain vector's projection. Let K and v be as in the KW problem (2.2). From (1.3) and the definition of P in (2.4) we see that $\|x\|^2\,\widetilde{\mu}(x)^2 = v^t P v$, and from (2.5) we have $v^t P v = \|Y^t v\|^2$. It follows again from (2.5) that

$$\widetilde{\mu}(x) = \frac{\|Pv\|}{\|x\|} = \frac{\|Y^t v\|}{\|x\|} = \frac{\|Ky\|}{\|x\|}, \tag{2.7}$$

where Y and $\|Y^t v\|$ may be obtained from the reduced QR factorization $K = YR$ in (2.3). (It is not essential to keep the Z part of Q.) Alternatively, $\|Ky\|$ may be obtained after the KW problem is solved by any method.

If QR factors of A are available (e.g., from solving the original LS problem), the required projection may be evaluated in two stages. Let the factors be denoted by subscript A. Applying Q_A^t to the top parts of K and v yields an equivalent LS problem

$$K' = \begin{bmatrix} R_A \\ 0 \\ \frac{\|r\|}{\|x\|} I \end{bmatrix}, \qquad v' = \begin{bmatrix} Y_A^t r \\ Z_A^t r \\ 0 \end{bmatrix}, \qquad \min_y \|K'y - v'\|\,. \tag{2.8}$$

The middle rows of K' and v' can now be removed and the problem becomes

$$K'' = \begin{bmatrix} R_A \\ \frac{\|r\|}{\|x\|} I \end{bmatrix}, \qquad v'' = \begin{bmatrix} Y_A^t r \\ 0 \end{bmatrix}, \qquad \min_y \|K''y - v''\|\,. \tag{2.9}$$

(If A has low column rank, we would still regard R_A and Y_A as having n columns.) Either way, a second QR factorization gives

$$\widetilde{\mu}(x) \;=\; \frac{\|Y_{K''}^t\, v''\|}{\|x\|}\,. \tag{2.10}$$

This formula could use two reduced QR factorizations. Of course, $Y_{K''}$ needn't be stored because $Y_{K''}^t\, v''$ can be accumulated as K'' is reduced to triangular form.

Table 2.2 shows that the optimal backward error can be estimated at little additional cost over that of solving the LS problem when $m \gg n$. Since K'' is a $2n \times n$ matrix, its QR factorization

Table 2.2: *Operation counts for solving LS problems by QR methods and then evaluating $\widetilde{\mu}(x)$ when $m \geq n$. The work to evaluate $Y_A^t r$ includes that of r. Only leading terms are shown.*

task	operations	source
solve LS by Householder QR, retaining Y_A	$2mn^2$	[14, p. 248]
form $Y_A^t r$	$4mn$	
apply $Y_{K''}^t$ to v''	$\frac{8}{3}n^3$	[21, p. 864]
finish evaluating $\widetilde{\mu}(x)$ by equation (2.10)	$2n$	

needs only $\mathcal{O}(n^3)$ operations compared to $\mathcal{O}(mn^2)$ for the factorization of A. Karlson and Waldén [21, p. 864] considered this same calculation in the course of evaluating a different estimate for the optimal backward error. They noted that sweeps of plane rotations most economically eliminate the lower block of K'' while retaining the triangular structure of R_A.

2.4 Operation counts for dense matrix methods

Table 2.3 summarizes the operation counts of solving the LS problem and estimating its optimal backward errors by the QR and SVD solution methods for dense matrices. It is clear that evaluating the estimate is negligible compared to evaluating the true optimal backward error. *Obtaining the estimate is even negligible compared to solving the LS problem by QR methods.*

The table shows that the QR approach also gives the most effective way to evaluate $\widetilde{\mu}(x)$ when the LS problem is solve by SVD methods. Chan's algorithm for calculating the SVD begins by performing a QR factorization. Saving this intermediate factorization allows equation (2.10) to evaluate the estimate with the same, small marginal cost as in the purely QR case of Table 2.3.

2.5 Sparse QR methods

Equation (2.10) uses both factors of A's QR decomposition: Y_A to transform r, and R_A occurs in K'. Although progress has been made towards computing both QR factors of a sparse matrix, notably by Adlers [1], it is considerably easier to work with just the triangular factor, as described by Matstoms [24]. Therefore methods to evaluate $\widetilde{\mu}(x)$ are needed that do not presume Y_A.

The simplest approach may be to evaluate equation (2.7) directly by transforming K to upper triangular form. Notice that A^tA and K^tK have identical sparsity patterns. Thus the same elimination analysis would serve to determine the sparse storage space for both R_A and R. Also, Y^tv can be obtained from QR factors of $\begin{bmatrix} K & v \end{bmatrix}$. The following MATLAB code [23] is often effective

Table 2.3: *Summary of operation counts to solve LS problems, to evaluate the estimate $\widetilde{\mu}(x)$, and to evaluate the exact $\mu(x)$. Only leading terms are considered.*

task	operations	$m = 1000$ $n = 100$	source
solve LS by QR	$2mn^2$	20,000,000	Table 2.2
solve LS by QR and evaluate $\widetilde{\mu}(x)$ by equation (2.10)	$2mn^2 + \frac{8}{3}n^3$	22,666,667	Table 2.2
solve LS by Chan SVD	$2mn^2 + 11n^3$	31,000,000	Table 2.1
solve LS by Chan SVD and evaluate $\widetilde{\mu}(x)$ by equation (2.10)	$2mn^2 + \frac{41}{3}n^3$	33,666,667	Tables 2.1, 2.2
solve LS by Chan SVD and evaluate $\widetilde{\mu}(x)$ by equation (2.1)	$6mn^2 + 20n^3$	80,000,000	Table 2.1
evaluate $\mu(x)$ by equation (1.1)	$4m^3 + 2m^2n$	4,200,000,000	[13, p. 175]

for computing $\widetilde{\mu}(x)$ for a sparse matrix A and sparse or dense vector b:

```
[m,n] = size(A);
r     = b - A*x;
normx = norm(x);
eta   = norm(r)/normx;
p     = colamd(A);
K     = [   A(:,p)                                    (2.11)
           eta*speye(n)];
v     = [   r
           zeros(n,1)];
[c,R] = qr(K,v,0);
muKW  = norm(c)/normx;
```

Note that `colamd` returns a good permutation `p` without forming `A'*A`, and `[c,R] = qr(K,v,0)` computes an "economy size" sparse `R` without storing any Q. The vector `c` is the required projection $Y^t v$.

Another approach is to evaluate equation (2.10) but with the substitution $Y_A^t r = Y_A^t(b - Ax)$, which gives

$$\widetilde{\mu}(x) \;=\; \frac{\left\| Y_{K''}^t \begin{bmatrix} Y_A^t b - R_A x \\ 0 \end{bmatrix} \right\|}{\|x\|} \,. \tag{2.12}$$

This simply recognizes that $Y_A^t r$ is the residual of the triangular linear equations used to solve

the LS problem. Solving that problem produces $Y_A^t b$ as an intermediary that can be saved for the backward error estimation. Factorization of K'' is still required, but again the orthogonal factor needn't be saved because it suffices to accumulate $Y_{K''}^t \begin{bmatrix} Y_A^t b - R_A x \\ 0 \end{bmatrix}$.

2.6 Iterative methods

If A is too large to permit the use of direct methods, we may consider iterative solution of the original problem $\min \|Ax - b\|$ as well as the KW problem (2.2):

$$\min_y \|Ky - v\| \;\equiv\; \min_y \left\| \begin{bmatrix} A \\ \eta I \end{bmatrix} y - \begin{bmatrix} r \\ 0 \end{bmatrix} \right\|, \qquad \eta \equiv \eta(x) = \frac{\|r\|}{\|x\|}. \tag{2.13}$$

In particular, LSQR [28, 29, 31] takes advantage of the damped least squares structure in (2.13). Using results from Saunders [30], we show here that the required projection norm is available within LSQR at negligible additional cost.

For problem (2.13), LSQR uses the Golub-Kahan bidiagonalization of A to form matrices U_k and V_k with theoretically orthonormal columns and a lower bidiagonal matrix B_k at each step k. With $\beta_1 = \|r\|$, a damped LS subproblem is defined and transformed by a QR factorization:

$$\min_{w_k} \left\| \begin{bmatrix} B_k \\ \eta I \end{bmatrix} w_k - \begin{bmatrix} \beta_1 e_1 \\ 0 \end{bmatrix} \right\|, \qquad Q_k \begin{bmatrix} B_k & \beta_1 e_1 \\ \eta I & 0 \end{bmatrix} = \begin{bmatrix} R_k & z_k \\ & \bar{\zeta}_{k+1} \\ & q_k \end{bmatrix}. \tag{2.14}$$

The kth estimate of y is defined to be $y_k = V_k w_k = (V_k R_k^{-1}) z_k$. From [30, pp. 99–100], the kth estimate of the required projection is given by

$$Ky \approx Ky_k \equiv \begin{bmatrix} A \\ \eta I \end{bmatrix} y_k \;=\; \begin{bmatrix} U_{k+1} & \\ & V_k \end{bmatrix} Q_k^t \begin{bmatrix} z_k \\ 0 \end{bmatrix}. \tag{2.15}$$

Orthogonality (and exact arithmetic) gives $\|Ky_k\| = \|z_k\|$. If LSQR terminates at iteration k, $\|z_k\|$ may be taken as the final estimate of $\|Ky\|$ for use in (2.7). Thus, $\widetilde{\mu}(x) \approx \|z_k\|/\|x\|$. Since z_k differs from z_{k-1} only in its last element, only k operations are needed to accumulate $\|z_k\|^2$.

LSQR already forms monotonic estimates of $\|y\|$ and $\|v - Ky\|$ for use in its stopping rules, and the estimates are returned as output parameters. We see that the estimate $\|z_k\| \approx \|Ky\|$ is another useful output. Experience shows that the estimates of such norms retain excellent accuracy even though LSQR does not use reorthogonalization.

2.7 When both A and b are perturbed

The case where only A is perturbed has been discussed. A practical estimate for the optimal backward error when both A and b are perturbed is also of interest.

In this case, the optimal backward error is defined as

$$\min_{\Delta A, \Delta b} \{||\Delta A, \theta \Delta b||_F : ||(A + \Delta A)y - (b + \Delta b)||_2 = \min\},$$

where θ is a weighting parameter. (Taking the limit $\theta \to \infty$ forces $\Delta b = 0$, giving the case where only A is perturbed.) The exact backward error, $\mu_{A,b}(x)$, is given as [20, p. 393]

$$\mu_{A,b}(x) = \min\{\sqrt{\nu}\eta,\ \sigma_{\min}[A\ \ B]\},$$

where

$$\eta = ||r||/||x||,\ B = \sqrt{\nu}\eta\left(I - \frac{rr^t}{||r||^2}\right),\ \text{and}\ \nu = \frac{\theta^2||x||^2}{1 + \theta^2||x||^2}.$$

Using the estimate $\tilde{\mu}(x) \approx \mu(x)$ (with only A perturbed), we can derive an analogous estimate $\tilde{\mu}_{A,b}(x) \approx \mu_{A,b}(x)$ as follows:

$$\begin{aligned}
\tilde{\mu}_{A,b}(x) &= \min\{\sqrt{\nu}\eta,\ \sigma_{\min}[A\ \ B]\} \\
&= \sqrt{\nu}\min\left\{\eta, \sigma_{\min}\left[\frac{1}{\sqrt{\nu}}A \quad \eta\left(I - \frac{rr^t}{||r||^2}\right)\right]\right\} \\
&= \sqrt{\nu}\left\|\left(\frac{||x||^2}{\nu}A^tA + ||r||^2 I\right)^{-1/2}\frac{1}{\sqrt{\nu}}A^tr\right\| \\
&= \sqrt{\nu}\|(||x||^2A^tA + \nu||r||^2I)^{-1/2}A^tr\|.
\end{aligned}$$

Note that to proceed from the line beginning $\sqrt{\nu}\min$, we simply replaced A by $A/\sqrt{\nu}$ in the formula for $\tilde{\mu}(x)$.

The asymptotic property (1.6) again follows because $\tilde{\mu}_{A,b}(x)$ and $\tilde{\mu}(x)$ have the same essential structure. All the evaluation methods for $\tilde{\mu}(x)$ can be carried out for $\tilde{\mu}_{A,b}(x)$ in a similar way.

Chapter 3

Numerical Tests

3.1 Description of the test problems

This section presents numerical tests of the optimal backward error estimate. For this purpose it is most desirable to make many tests with problems that occur in practice. Since large collections of test problems are not available for least squares, it is necessary to compromise by using many randomly generated vectors, b, with a few matrices, A, that are related to real-world problems.

Table 3.1 describes the test matrices. They originated in the least-squares analysis of gravity-meter observations. They are available from the Harwell-Boeing sparse matrix collection [9] and the Matrix Market [5].

Table 3.1: *Matrices used in the numerical tests.*

matrix	rows m	columns n	$\kappa_2(A)$
(a) `well1033`	1033	320	$1.7e{+}2$
(b) `illc1033`	1033	320	$1.9e{+}4$

3.2 Description of the calculations

For the factorization methods, 1000 sample problems are considered for each type of matrix in Table 3.1. For each sample problem, the solution x, the backward error estimate $\widetilde{\mu}(x)$ and the optimal backward error $\mu(x)$ from Higham's equation (1.1) are evaluated using MATLAB.

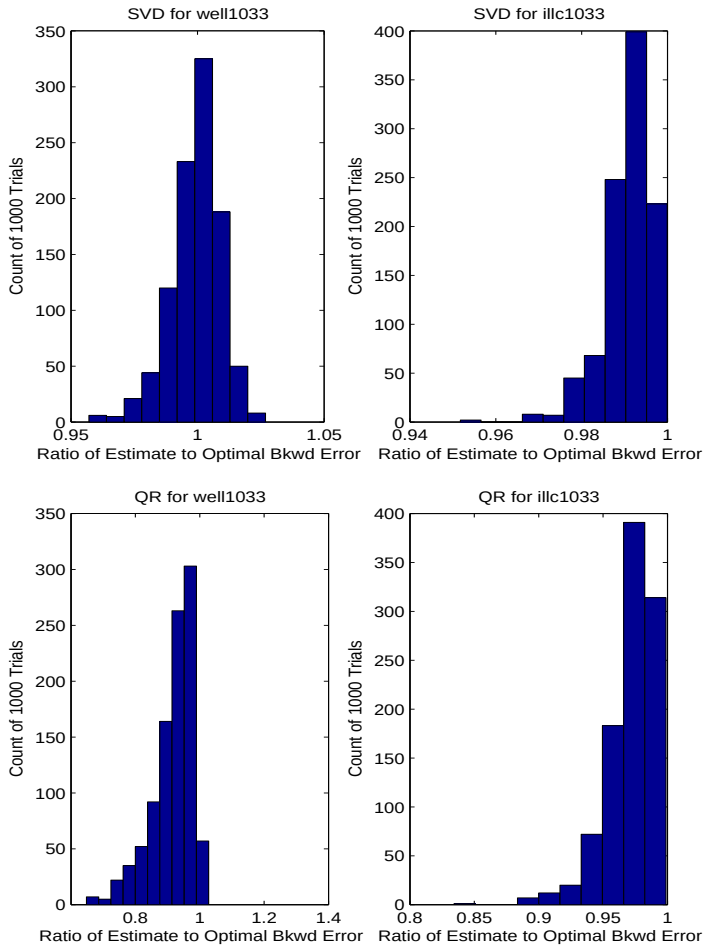

Figure 3.1: *Histograms for the ratios of estimate to true optimal backward error for all the test cases solved by dense matrix factorizations. The SVD and QR solution methods use the estimates in equation (2.1) and (2.10), respectively.*

3.3 Test results for SVD, QR, and sparse methods

Figure 3.1 displays the ratios of estimate $\widetilde{\mu}(x)$ to optimal backward error $\mu(x)$ for all the test cases solved by dense matrix factorizations. The SVD and QR solution methods use the estimates in equation (2.1) and (2.10), respectively. Figure 3.2 displays the ratios for the same x obtained by QR methods in Figure 3.1 but with $\widetilde{\mu}(x)$ evaluated by equation (2.11). This formula is the first of the two approaches suggested for use with sparse matrices. The figures show that $\widetilde{\mu}(x)$ evaluated by these formulas is a reasonable estimate for the optimal backward error.

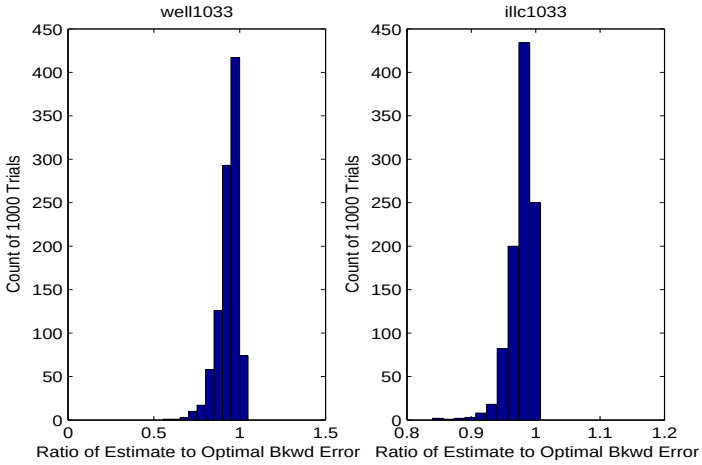

Figure 3.2: *Histograms for the ratios of estimate to true optimal backward error for all the test cases solved by QR methods. Equation (2.11) is used to evaluate the estimate.*

3.4 Test results for iterative methods

The preceding results have been for accurate estimates of the LS solution. Applying LSQR to a problem $\min_x \|Ax-b\|$ generates a sequence of approximate solutions $\{x_k\}$. For the `well` and `illc` test problems we used the MATLAB code (2.11) to compute $\widetilde{\mu}(x_k)$ for each x_k. To our surprise, these values proved to be monotonically decreasing, as illustrated by the lower curve in Figures 3.3 and 3.4. (To make it scale-independent, this curve is really $\widetilde{\mu}(x_k)/\|A\|_{\mathrm{F}}$.)

For each x_k, let $r_k = b - Ax_k$ and $\eta(x_k) = \|r_k\|/\|x_k\|$. Also, let K_k, v_k and y_k be the quantities in (2.2) when $x = x_k$. The LSQR iterates have the property that $\|r_k\|$ and $\|x_k\|$ are decreasing and increasing respectively, so that $\eta(x_k)$ is monotonically decreasing. Also, we see from (2.7) that

$$\widetilde{\mu}(x_k) = \frac{\|Y_k^t v_k\|}{\|x_k\|} < \frac{\|v_k\|}{\|x_k\|} = \frac{\|r_k\|}{\|x_k\|} = \eta(x_k),$$

so that $\eta(x_k)$ forms a monotonically decreasing bound on $\widetilde{\mu}(x)$. However, we can only note empirically that $\widetilde{\mu}(x_k)$ itself appears to decrease monotonically also.

The stopping criterion for LSQR is of interest. It is based on a non-optimal backward error $\|E_k\|_{\mathrm{F}}$ derived by Stewart [32], where

$$E_k = -\frac{1}{\|r_k\|^2} r_k r_k^t A.$$

(If $\tilde{A} = A + E_k$ and $\tilde{r} = b - \tilde{A}x_k$, then $(x_k, \tilde{r}_k)$ are the exact solution and residual for $\min_x \|\tilde{A}x - b\|$.) Note that $\|E_k\|_{\mathrm{F}} = \|E_k\|_2 = \|A^t r_k\|/\|r_k\|$. On incompatible systems, LSQR terminates when its

estimate of $\|E_k\|_2/\|A\|_{\text{F}}$ is sufficiently small; i.e., when

$$\texttt{test2}_k \equiv \frac{\|A^t r_k\|}{\|A\|_k \|r_k\|} \leq \texttt{atol}, \tag{3.1}$$

where $\|A\|_k$ is a monotonically increasing estimate of $\|A\|_{\text{F}}$ and `atol` is a user-specified tolerance.

Figures 3.3 and 3.4 show $\|r_k\|$ and three relative backward error quantities for problems `well1033` and `illc1033` when LSQR is applied to $\min_x \|Ax - b\|$ with $\texttt{atol} = 10^{-12}$. From top to bottom, the curves plot the following ($\log_{10}$):

- $\|r_k\|$ (monotonically decreasing).
- $\texttt{test2}_k$, LSQR's relative backward error estimate (3.1).
- $\eta(x_k)/\|A\|_{\text{F}}$, the optimal relative backward error for $Ax = b$ (monotonic).
- $\widetilde{\mu}(x_k)/\|A\|_{\text{F}}$, the KW relative backward error estimate for $\min_x \|Ax - b\|$ (apparently monotonic).

The last curve is extremely close to the optimal relative backward error for LS problems. We see that LSQR's $\texttt{test2}_k$ is two or three orders of magnitude larger for most x_k, and it is far from monotonic. Nevertheless, its trend is downward in broad synchrony with $\widetilde{\mu}(x_k)/\|A\|_{\text{F}}$. We take this as an experimental approval of Stewart's backward error E_k and confirmation of the reliability of LSQR's cheaply computed stopping rule.

3.5 Iterative computation of $\widetilde{\mu}(x)$

Here we use an iterative solver twice: first on the original LS problem to obtain an approximate solution x, and then on the associated KW problem to estimate the backward error for x.

1. Apply LSQR to $\min_x \|Ax - b\|$ with iteration limit $kmax$. This generates a sequence $\{x_k\}$, $k = 1 : kmax$. Define $x = x_{kmax}$. We want to estimate the backward error for that final point x.
2. Define $r = b - Ax$ and $\texttt{atol} = 0.01\|A^t r\|/(\|A\|_{\text{F}}\|x\|)$.
3. Apply LSQR to the KW problem $\min_y \|Ky - v\|$ (2.13) with convergence tolerance `atol`. As described in section 2.6, this generates a sequence of estimates $\widetilde{\mu}(x) \approx \|z_\ell\|/\|x\|$ using $\|z_\ell\| \approx \|Ky\|$ in (2.14)–(2.15).

To avoid ambiguity we use k and ℓ for LSQR's iterates on the two problems. Also, the following figures plot relative backward errors $\widetilde{\mu}(x)/\|A\|_{\text{F}}$, even though the accompanying discussion doesn't mention $\|A\|_{\text{F}}$.

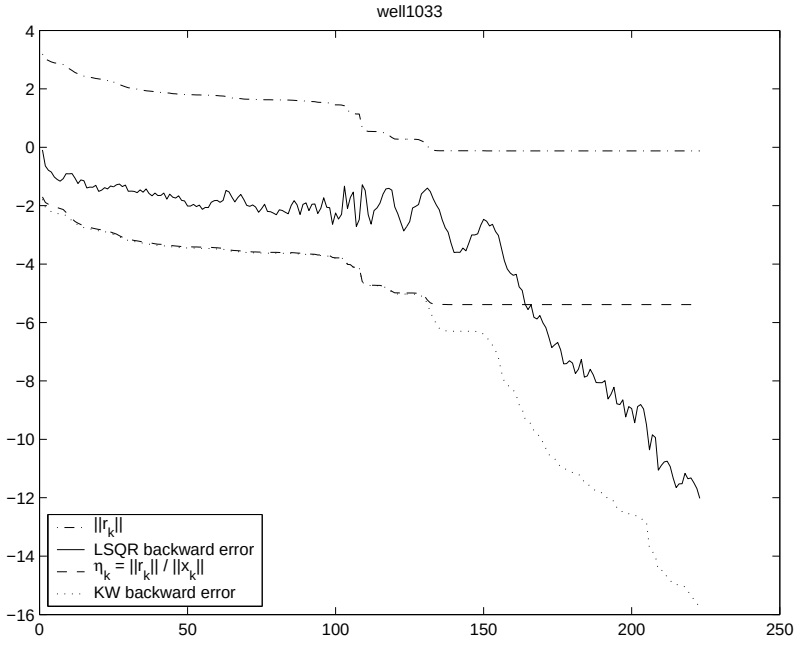

Figure 3.3: *Backward error estimates for each LSQR iterate x_k during the solution of* `well1033` *with* `atol` $= 10^{-12}$.

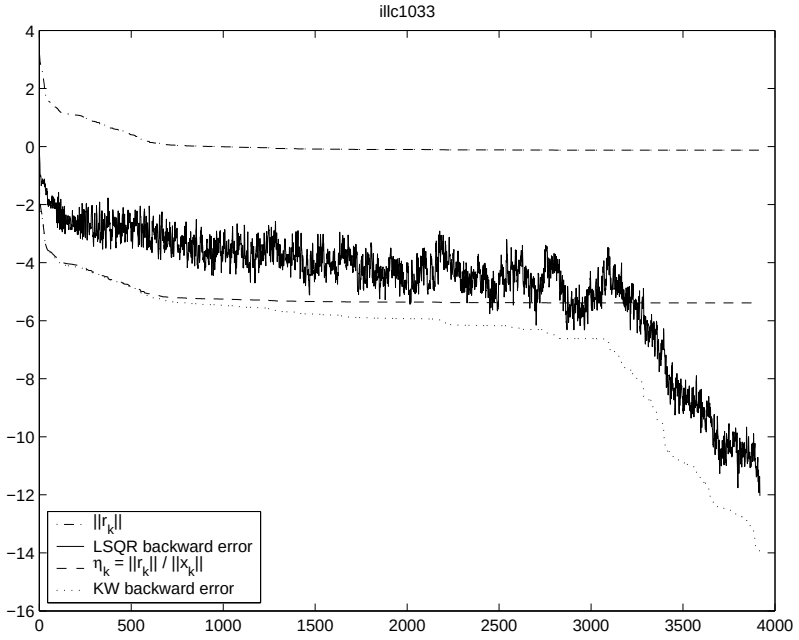

Figure 3.4: *Backward error estimates for each LSQR iterate x_k during the solution of* `illc1033` *with* `atol` $= 10^{-12}$.

For problem `well1033` with $kmax = 50$, Figure 3.5 shows $\widetilde{\mu}(x_k)$ for $k = 1:50$ (the same as the beginning of Figure 3.3). The right-hand curve then shows about 130 estimates $\|z_\ell\|/\|x\|$ converging to $\widetilde{\mu}(x_{50})$ with about 2 digits of accuracy (because of the choice of `atol`).

Similarly with $kmax = 160$, Figure 3.6 shows $\widetilde{\mu}(x_k)$ for $k = 1:160$ (the same as the beginning of Figure 3.3). The final point x_{160} is close to the LS solution, and the subsequent KW problem converges more quickly. About 20 LSQR iterations give a 2-digit estimate of $\widetilde{\mu}(x_{160})$.

For problem `illc1033`, similar effects were observed. In Figure 3.7 about 2300 iterations on the KW problem give a 2-digit estimate of $\widetilde{\mu}(x_{2000})$, but in Figure 3.8 only 280 iterations are needed to estimate $\widetilde{\mu}(x_{3500})$.

3.6 Comparison with Malyshev and Sadkane's method

Malyshev and Sadkane [22] show how to use the bidiagonalization of A with starting vector r to estimate $\sigma_{\min}[A\ \ B]$ in (1.1). This is the same bidiagonalization that LSQR uses on the KW problem (2.2) to estimate $\widetilde{\mu}(x)$. The additional work per iteration is nominal in both cases. A numerical comparison is therefore of interest. We use the results in Tables 5.2 and 5.3 of [22] corresponding to LSQR's iterates x_{50} and x_{160} on problems `well1033` and `illc1033`. Also, MATLAB gives us accurate values for $\widetilde{\mu}(x_k)$ and $\sigma_{\min}[A\ \ B]$ via sparse `qr` (2.11) and dense `svd` respectively.

In Tables 3.2–3.4, the true backward error is $\mu(x) = \sigma_{\min}[A\ \ B]$, the last line in each table.

In Tables 3.2–3.3, σ_ℓ denotes Malyshev and Sadkane's $\sigma_{\min}(\bar{B}_\ell)$ [22, (3.7)]. Note that the iterates σ_ℓ provide decreasing upper bounds on $\sigma_{\min}[A\ \ B]$, while the LSQR iterates $\|z_\ell\|/\|x\|$ are increasing lower bounds on $\widetilde{\mu}(x)$, but they do not bound $\sigma_{\min}$.

We see that all of the Malyshev and Sadkane estimates σ_ℓ bound $\sigma_{\min}$ to within a factor of 2, but they have no significant digits in agreement with $\sigma_{\min}$. In contrast, $\eta(x_k)$ agrees with $\sigma_{\min}$ to 3 digits in three of the cases, and indeed it provides a tighter bound whenever it satisfies $\eta < \sigma_\ell$. The estimates σ_ℓ are therefore more valuable when $\eta > \sigma_{\min}$ (i.e., when x_k is close to a solution x_*).

However, we see that LSQR computes $\widetilde{\mu}(x_k)$ with 3 or 4 correct digits in all cases, and requires fewer iterations as x_k approaches x_*. The bottom-right values in Tables 3.2 and 3.4 show Grcar's limit (1.6) taking effect. LSQR can compute these values to high precision with reasonable efficiency.

The primary difficulty with our iterative computation of $\widetilde{\mu}(x)$ is that when x is not close to x_*, rather many iterations may be required, and there is no warning that $\widetilde{\mu}$ may be an underestimate of μ.

Ironically, solving the KW problem for $x = x_k$ is akin to restarting LSQR on a slightly modified problem. We have observed that if ℓ iterations are needed on the KW problem to estimate $\widetilde{\mu}(x_k)/\|A\|_{\text{F}}$, continuing the original LS problem a further ℓ iterations would have given a point

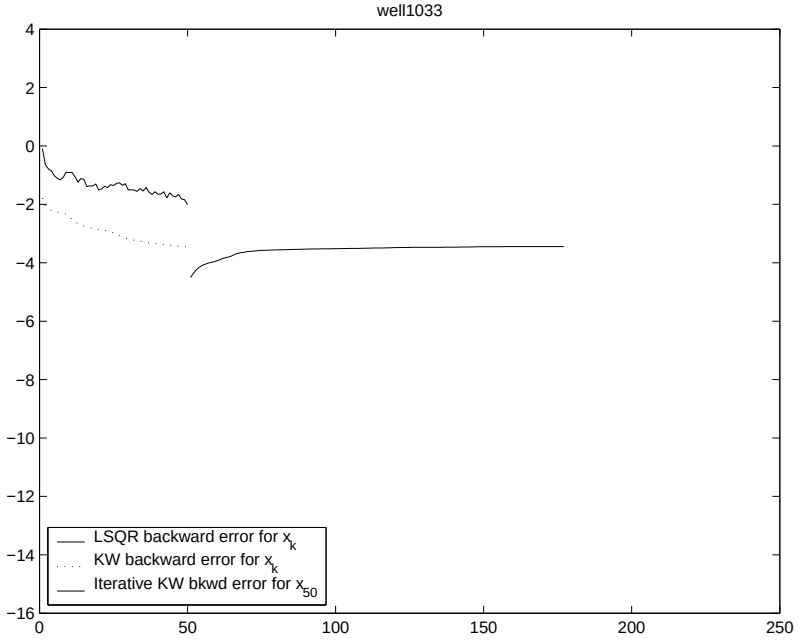

Figure 3.5: *Problem* `well1033`*: Iterative solution of KW problem after LSQR is terminated at* x_{50}.

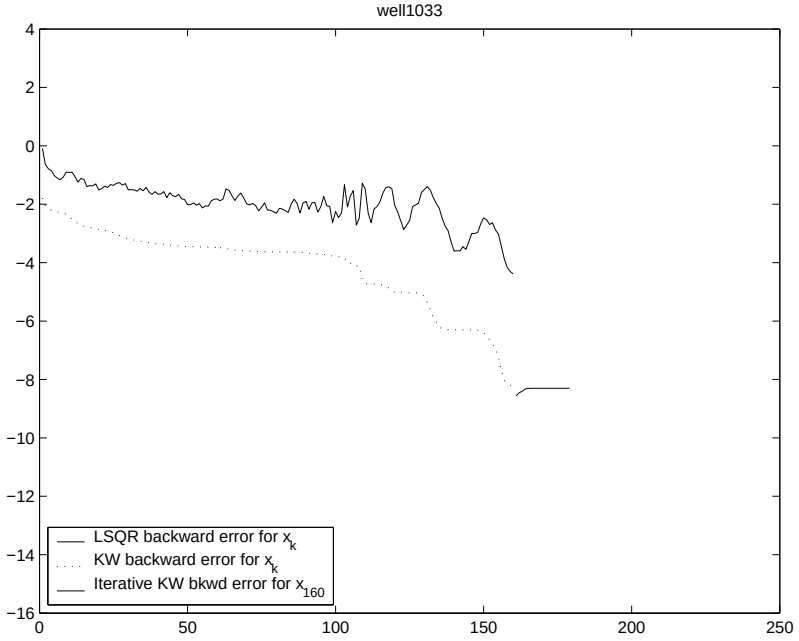

Figure 3.6: *Problem* `well1033`*: Iterative solution of KW problem after LSQR is terminated at* x_{160}.

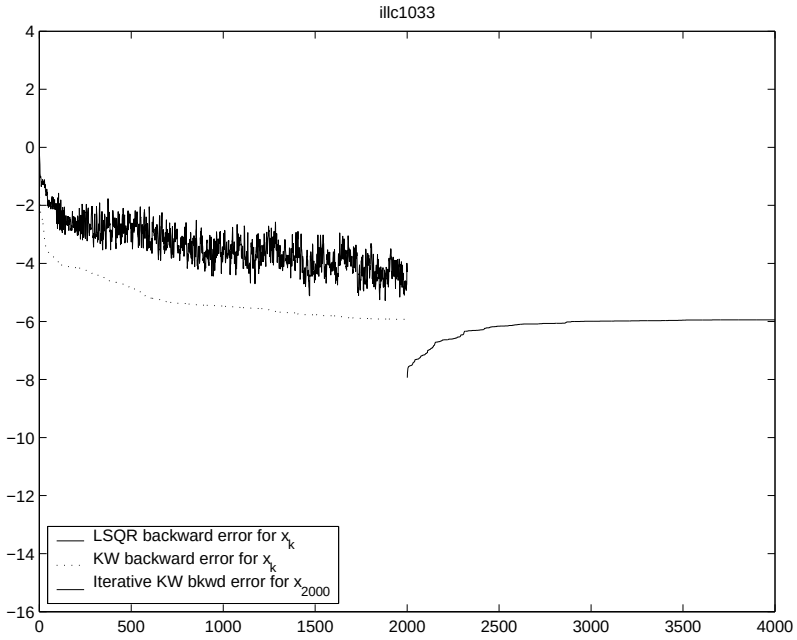

Figure 3.7: *Problem* `illc1033`*: Iterative solution of KW problem after LSQR is terminated at* x_{2000}.

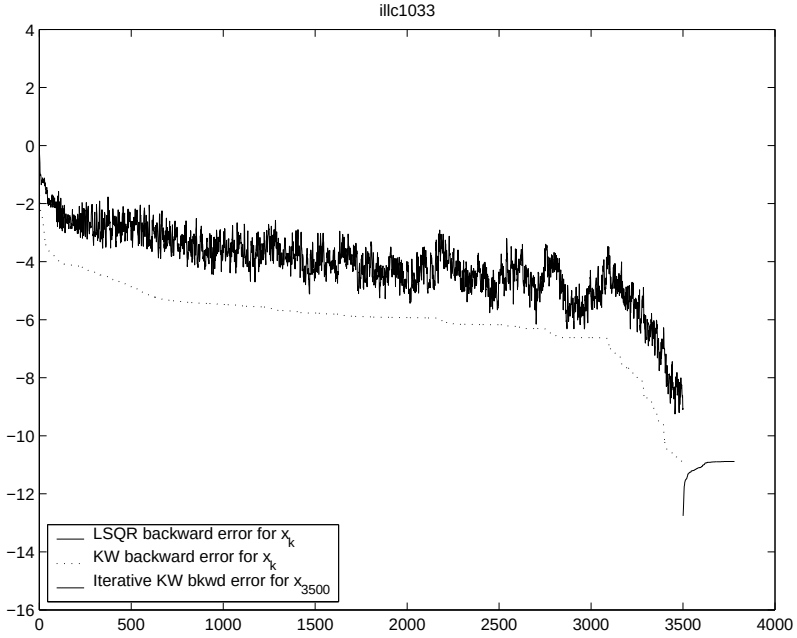

Figure 3.8: *Problem* `illc1033`*: Iterative solution of KW problem after LSQR is terminated at* x_{3500}.

Table 3.2: *Comparison of* σ_ℓ *and* $\|z_\ell\|/\|x_k\|$ *for problem* `well1033`.

$k = 50$			$k = 160$		
$\|r_k\|$	6.35e+1		$\|r_k\|$	7.52e+1	
$\|A^t r_k\|$	5.04e+0		$\|A^t r_k\|$	4.49e−4	
$\eta(x_k)$		7.036807e−3	$\eta(x_k)$		7.3175e−5
`atol`	4.44e−5		`atol`	3.34e−7	
ℓ	σ_ℓ	$\|z_\ell\|/\|x_k\|$	ℓ	σ_ℓ	$\|z_\ell\|/\|x_k\|$
10	2.35e−2	2.11e−3	10	3.79e−5	8.9316e−8
50	1.51e−2	5.43e−3	19		8.9381e−8
100	1.22e−2	6.32e−3	50	2.95e−7	
127		6.379461e−3	100	1.21e−7	
$\widetilde{\mu}(x_k)$		6.379462e−3	$\widetilde{\mu}(x_k)$		8.9386422278e−8
$\sigma_{\min}[A\ B]$		7.036158e−3	$\sigma_{\min}[A\ B]$		8.9386422275e−8

Table 3.3: *Comparison of* σ_ℓ *and* $\|z_\ell\|/\|x_k\|$ *for problem* `illc1033`.

$k = 50$			$k = 160$		
$\|r_k\|$	3.67e+1		$\|r_k\|$	1.32e+1	
$\|A^t r_k\|$	3.08e+1		$\|A^t r_k\|$	3.78e−1	
$\eta(x_k)$		4.6603e−3	$\eta(x_k)$		1.6196e−3
`atol`	4.69e−5		`atol`	1.60e−5	
ℓ	σ_ℓ	$\|z_\ell\|/\|x_k\|$	ℓ	σ_ℓ	$\|z_\ell\|/\|x_k\|$
10	3.04e−2	1.62e−3	10	1.10e−2	2.09e−4
50	1.84e−2	3.71e−3	50	4.63e−3	4.92e−4
100	1.02e−2	4.11e−3	100	3.40e−3	8.45e−4
200		4.25e−3	200		1.23e−3
300		4.28e−3	300		1.34e−3
310		4.2825e−3	400		1.38e−3
$\widetilde{\mu}(x_k)$		4.2831e−3	500		1.3841e−3
$\sigma_{\min}[A\ B]$		4.6576e−3	542		1.3843e−3
			$\widetilde{\mu}(x_k)$		1.3847e−3
			$\sigma_{\min}[A\ B]$		1.6144e−3

Table 3.4: $\|z_\ell\|/\|x_k\|$ *for problem* `illc1033`.

$k = 2000$		$k = 3500$	
$\|r_k\|$	7.89e−1	$\|r_k\|$	7.52e−1
$\|A^t r_k\|$	2.45e−3	$\|A^t r_k\|$	5.54e−8
$\eta(x_k)$	7.82e−5	$\eta(x_k)$	7.30e−5
`atol`	1.73e−6	`atol`	4.11e−11
ℓ	$\|z_\ell\|/\|x_k\|$	ℓ	$\|z_\ell\|/\|x_k\|$
500	1.22e−5	10	4.41e−11
1000	1.81e−5	50	1.11e−10
1500	1.97e−5	100	1.54e−10
2000	2.02e−5	200	2.28e−10
2330	2.08e−5	280	2.32006e−10
$\widetilde{\mu}(x_k)$	2.10e−5	$\widetilde{\mu}(x_k)$	2.3209779030e−10
$\sigma_{\min}[A\ B]$	2.12e−5	$\sigma_{\min}[A\ B]$	2.3209779099e−10

$x_{k+\ell}$ for which the Stewart-type backward error $\texttt{test2}_{k+\ell}$ is generally at least as small. (Compare Figures 3.4 and 3.8.) Thus, the decision to estimate optimal backward errors by iterative means must depend on the real need for optimality.

3.7 Test results for perturbed b

Figure 3.9 displays the ratios of the estimate $\widetilde{\mu}(x)$ to the optimal backward error $\mu(x)$ for the SVD and the sparse QR methods for evaluating $\tilde{\mu}(x)$. $\theta = 0.02$ for the SVD method and $\theta = 0.001$ for the sparse QR method. θ is chosen such that ν won't be too close to 1. The figures show that $\widetilde{\mu}(x)$ evaluated by these formulas is a good estimate for the optimal backward error $\mu(x)$.

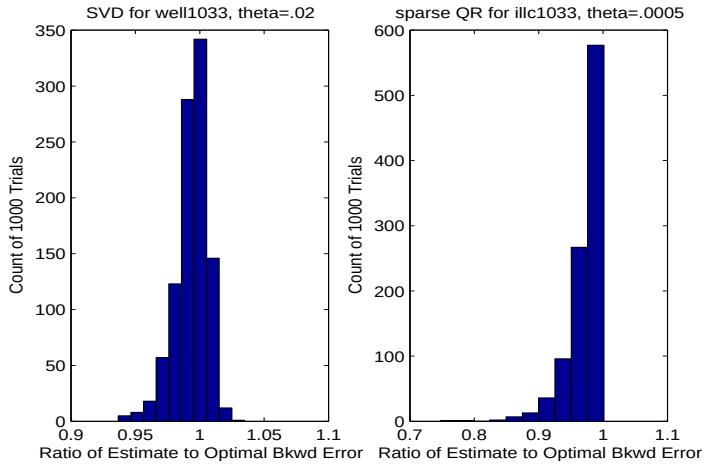

Figure 3.9: *Histograms for the ratios of estimate to true optimal backward error for SVD and sparse methods.* $\theta = 0.02$ *for SVD and* $\theta = 0.001$ *for sparse method.* θ *is chosen such that* ν *won't be too close to 1.*

Chapter 4

Upper and Lower Bounds for $\tilde{\mu}$

Another way of evaluating $\tilde{\mu}$ is to find a sequence of decreasing upper bounds and another sequence of increasing lower bounds for $\tilde{\mu}$, and we will have a good estimate of $\tilde{\mu}$ when the upper and lower bounds get close enough. This can be done by using Gauss, Gauss-Radau and Gauss-Lobbato quadrature formulas. In order to find upper and lower bounds for $\tilde{\mu}$, we only need to find upper and lower bounds for $\bar{\mu} = z^t(A^tA + \eta^2 I)^{-1}z$, where $z = A^tr/||A^tr||$. We have

$$\tilde{\mu} = \frac{||A^tr||^2}{||x||^2}\bar{\mu}.$$

4.1 Matrix functions

Given a symmetric positive definite matrix A, we may write $A = Q\Lambda Q^t$, where Q is the orthonormal matrix whose columns are the normalized eigenvectors of A, and Λ is a diagonal matrix whose diagonal elements are the eigenvalues λ_i, which we order as $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$.

If $f(A)$ is an analytic function of A (such as a polynomial in A), we have

$$f(A) = Qf(\Lambda)Q^t.$$

Therefore, for arbitrary vectors u and v,

$$u^tf(A)v \;=\; u^tQf(\Lambda)Q^tv \;=\; \alpha^tf(\Lambda)\beta \;=\; \sum_{i=1}^{n} f(\lambda_i)\alpha_i\beta_i,$$

where $\alpha = Q^tu$, $\beta = Q^tv$. This last sum can be considered as a Riemann-Stieltjes integral:

$$I[f] = u^tf(A)v = \int_a^b f(\lambda)\,d\alpha(\lambda),$$

where the measure α is piecewise constant and defined by

$$\alpha(\lambda) = \begin{cases} 0 & \text{if } \lambda < a = \lambda_1 \\ \sum_{j=1}^{i} \alpha_j \beta_j & \text{if } \lambda_i \leq \lambda < \lambda_{i+1} \\ \sum_{j=1}^{n} \alpha_j \beta_j & \text{if } b = \lambda_n \leq \lambda. \end{cases}$$

Note that α is an increasing positive function. We are looking for methods to obtain upper and lower bounds L and U for $I[f]$:

$$L \leq I[f] \leq U.$$

In the next section, we review and describe some basic results from Gauss quadrature theory following Golub and Meurant [12], as this plays a fundamental role in estimating the integrals and computing bounds.

4.2 Bounds on matrix functions as integrals

A way to obtain bounds for the Stieltjes integrals is to use Gauss, Gauss-Radau and Gauss-Lobatto quadrature formulas. The general formula we use is

$$\int_a^b f(\lambda)\, d\alpha(\lambda) = \sum_{j=1}^{N} w_j f(t_j) + \sum_{k=1}^{M} v_k f(z_k) + R[f],$$

where the weights $[w_j]_{j=1}^N, [v_k]_{k=1}^M$ and the nodes $[t_j]_{j=1}^N$ are unknowns and the nodes $[z_k]_{k=1}^M$ are prescribed. The remainder term is

$$R[f] = \frac{f^{(2N+M)}(\xi)}{(2N+M)!} \int_a^b \prod_{k=1}^{M} (\lambda - z_k) \left[\prod_{j=1}^{N} (\lambda - t_j) \right]^2 d\alpha(\lambda), \quad a < \xi < b.$$

If $M = 0$, this leads to the Gauss rule with no prescribed nodes. If $M = 1$ and $z_1 = a$ or $z_1 = b$ we have the Gauss-Radau formula. If $M = 2$ and $z_1 = a$, $z_2 = b$, this is the Gauss-Lobatto formula.

Let us recall briefly how the nodes and weights are obtained in the Gauss, Gauss-Radau and Gauss-Lobatto rules. For the measure α, it is possible to define a sequence of polynomials $p_0(\lambda)$, $p_1(\lambda), \ldots$ that are orthonormal with respect to α:

$$\int_a^b p_i(\lambda) p_j(\lambda)\, d\alpha(\lambda) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases}$$

and p_k is of exact degree k. Moreover, the roots of p_k are distinct, real and lie in the interval $[a, b]$.

If $\int d\alpha = 1$, this set of orthonormal polynomials satisfies a three-term recurrence relationship:

$$\gamma_j p_j(\lambda) = (\lambda - w_j)p_{j-1}(\lambda) - \gamma_{j-1}p_{j-2}(\lambda), \quad j = 1, 2, \ldots, N$$

with $p_{-1} \equiv 0$, $p_0(\lambda) \equiv 1$. In matrix form, this can be written as

$$\lambda p(\lambda) = T_N p(\lambda) + \gamma_N p_N(\lambda) e_N,$$

where

$$p(\lambda)^t = [p_0(\lambda)\ p_1(\lambda)\ \cdots\ p_{N-1}(\lambda)],$$

$$e_N^t = (0\ 0\ \cdots\ 1),$$

and

$$T_N = \begin{pmatrix} \omega_1 & \gamma_1 & & & \\ \gamma_1 & \omega_2 & \gamma_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \gamma_{N-2} & \omega_{N-1} & \gamma_{N-1} \\ & & & \gamma_{N-1} & \omega_N \end{pmatrix}.$$

The eigenvalues of T_N (which are the zeroes of p_N) are the nodes of the Gauss quadrature rule (i.e. $M = 0$). The weights are the squares of the first elements of the normalized eigenvectors of T_N. We note that all the eigenvalues of T_N are real and simple. A natural and elegant way to compute the orthonormal polynomials, or equivalently the tridiagonal matrices, is to use the Lanczos algorithm.

4.2.1 Gauss quadrature rule for lower bounds

For the Gauss quadrature rule (renaming the weights and nodes w_j^G and t_j^G) we have

$$\int_a^b f(\lambda)\, d\alpha(\lambda) = \sum_{j=1}^{N} w_j^G f(t_j^G) + R_G[f],$$

with

$$R_G[f] = \frac{f^{(2N)}(\xi)}{(2N)!} \int_a^b \left[\prod_{j=1}^{N} (\lambda - t_j^G) \right]^2 d\alpha(\lambda), \quad a < \xi < b.$$

Golub and Meurant [12] showed that, given $u = v$,

$$\sum_{j=1}^{N} w_j^G f(t_j^G) = e_1^t f(T_N) e_1,$$

where T_N arises from applying the Lanczos tridiagonalization algorithm to the symmetric positive definite matrix A with $u/||u||$ as the starting vector.

In our case the function f has the form $f(x) = (x + \eta^2)^{-1}$. Thus,

$$\frac{f^{(2N)}(\xi)}{(2N)!} = (\xi + \eta^2)^{-(2N+1)} > 0.$$

We have $R_G[f] > 0$, and hence $\sum_{j=1}^{N} w_j^G f(t_j^G)$ gives a lower bound for $I[f]$. As a result, $e_1^t(T_N + \eta^2 I)^{-1} e_1$ gives a lower bound for $\bar{z}$, $N = 1, 2, \ldots$, where T_N is from applying the Lanczos tridiagonalization algorithm to $A^t A$ with z as the starting vector.

Here we describe a procedure to evaluate $e_1^t(T_N + \eta^2 I)^{-1} e_1$ that avoids forming $A^t A$. First, we apply the Golub-Kahan bidiagonalization algorithm to A with starting vector r to get lower bidiagonal matrices

$$B_N = \begin{pmatrix} b_{11} & & & & & \\ b_{21} & b_{22} & & & & \\ & b_{32} & \ddots & & & \\ & & \ddots & b_{N-1,N-1} & & \\ & & & b_{N,N-1} & b_{NN} & \\ & & & & & b_{N+1,N} \end{pmatrix}, \quad N = 1, 2, \ldots.$$

We have

$$T_N = B_N^t B_N.$$

Second, we calculate the QR factorization of B_N:

$$Q_N B_N = \begin{pmatrix} R_N \\ 0 \end{pmatrix}.$$

Third, we calculate the QR factorization of $\begin{pmatrix} R_N \\ \eta I \end{pmatrix}$:

$$\bar{Q}_N \begin{pmatrix} R_N \\ \eta I \end{pmatrix} = \begin{pmatrix} \bar{R}_N \\ 0 \end{pmatrix}.$$

Then

$$e_1^t(T_N + \eta^2 I)^{-1} e_1 \;=\; e_1^t(\bar{R}_N^t \bar{R}_N)^{-1} e_1 \;=\; \|w_N\|^2,$$

where $\bar{R}_N^t w_N = e_1$. We now have a sequence of lower bounds as $N = 1, 2, \ldots$.

4.2.2 Gauss-Radau rule for upper and lower bounds

To obtain the Gauss-Radau rule ($M = 1$), we should extend the matrix T_N in such a way that it has one prescribed eigenvalue z_1. We wish to construct p_{N+1} such that $p_{N+1}(z_1) = 0$. From the recurrence relation, we have

$$0 = \gamma_{N+1} p_{N+1}(z_1) = (z_1 - w_{N+1}) p_N(z_1) - \gamma_N p_{N-1}(z_1).$$

This gives

$$w_{N+1} = z_1 - \gamma_N \frac{p_{N-1}(z_1)}{p_N(z_1)}.$$

Let us denote $\delta(z_1) = [\delta_1(z_1), \cdots, \delta_N(z_1)]^t$ with

$$\delta_l(z_1) = -\gamma_N \frac{p_{l-1}(z_1)}{p_N(z_1)}, \quad l = 1 \ldots, N.$$

This gives $\hat{w}_{N+1} = z_1 + \delta_N(z_1)$, where

$$(T_N - z_1 I)\delta(z_1) = \gamma_N^2 e_N.$$

For the Gauss-Radau rule the remainder R_{GR} is

$$R_{GR}[f] = \frac{f^{(2N+1)}(\xi)}{(2N+1)!} \int_a^b (\lambda - z_1) \left[\prod_{j=1}^{N} (\lambda - t_j) \right]^2 d\alpha(\lambda).$$

In our case,

$$\frac{f^{(2N+1)}(\xi)}{(2N+1)!} = -(\xi + \eta^2)^{-(2N+2)} < 0.$$

As a result, the sign of the remainder is determined by the sign of

$$\int_a^b (\lambda - z_1) \left[\prod_{j=1}^{N} (\lambda - t_j) \right]^2 d\alpha(\lambda).$$

Note that a and b are the smallest and largest eigenvalues of the symmetric positive definite matrix A, so they are positive. If we fix z_1 at 0, then the integral is positive. As a result, the remainder is negative and $\sum_{j=1}^{N} w_j f(t_j) + v_1 f(z_1)$ is an upper bound for $I[f]$. If we fix z_1 at $\sqrt{||A||_1 ||A||_\infty}$ ($> b$), then the integral is negative. As a result, the remainder is positive and $\sum_{j=1}^{N} w_j f(t_j) + v_1 f(z_1)$ is a lower bound.

The tridiagonal matrix $\hat{T}_{N+1}$ defined as $\hat{T}_{N+1} = \begin{pmatrix} T_N & \gamma_N e_N \\ \gamma_N e_N^t & \hat{w}_{N+1} \end{pmatrix}$ will have z_1 as an eigenvalue and give the weights and nodes of the corresponding quadrature rule. Therefore, the recipe is to

compute as for the Gauss quadrature rule and then to modify the last diagonal element to obtain the prescribed node.

Golub and Meurant [12] proved that

$$\sum_{j=1}^{N} w_j f(t_j) + v_1 f(z_1) = e_1^t f(\hat{T}_N) e_1.$$

In order to find upper bounds for $\bar{\mu}$, we set $z_1 = 0$. Next, we solve

$$T_N \delta(z_1) = \gamma_N^2 e_N,$$

where T_N arises from applying Lanzos tridiagonalization to $A^t A$ with z as the starting vector. Set $\hat{w}_{N+1} = z_1 + \delta_N(z_1)$ and then $e_1^t(\hat{T}_{N+1} + \eta^2 I)^{-1} e_1$ gives an upper bound for $\bar{\mu}$, $N = 1, 2, \ldots$. Similar to the Gauss rule case, we can do Golub-Kahan bidiagonalization for A instead of Lanczos tridiagonalization for $A^t A$.

In order to find lower bounds for $\bar{\mu}$, we set $z_1 = \sqrt{||A^t A||_1 ||A^t A||_\infty}$. Next, we solve

$$(T_N - z_1 I)\delta(z_1) = \gamma_N^2 e_N.$$

Set $\hat{w}_{N+1} = z_1 + \delta_N(z_1)$ and then $e_1^t(\hat{T}_{N+1} + \eta^2 I)^{-1} e_1$ gives a lower bound for $\bar{\mu}$, $N = 1, 2, \ldots$.

4.2.3 Gauss-Lobatto rule for upper bounds

Consider the Gauss-Lobatto rule ($M = 2$), with z_1 and z_2 as prescribed nodes. Again, we should modify the matrix of the Gauss quadrature rule. Here, we would like to have

$$p_{N+1}(z_1) = p_{N+1}(z_2) = 0.$$

Using the recurrence relation for the polynomials, we obtain a linear system of order 2 for the unknowns $\hat{w}_{N+1}$ and $\hat{\gamma}_N$:

$$\begin{pmatrix} p_N(z_1) & p_{N-1}(z_1) \\ p_N(z_2) & p_{N-1}(z_2) \end{pmatrix} \begin{pmatrix} \hat{w}_{N+1} \\ \hat{\gamma}_N \end{pmatrix} = \begin{pmatrix} z_1\, p_N(z_1) \\ z_2\, p_N(z_2) \end{pmatrix}.$$

Let δ and μ be defined as vectors with components

$$\delta_l = -\frac{p_{l-1}(z_1)}{\gamma_N p_N(z_1)}, \quad \mu_l = -\frac{p_{l-1}(z_2)}{\gamma_N p_N(z_2)}.$$

Then

$$(T_N - z_1 I)\delta = e_N, \quad (T_N - z_2 I)\mu = e_N,$$

and the linear system can be written as

$$\begin{pmatrix} 1 & -\delta_N \\ 1 & -\mu_N \end{pmatrix} \begin{pmatrix} \hat{w}_{N+1} \\ \hat{\gamma}_N^2 \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix},$$

giving the unknowns we need. The tridiagonal matrix $\hat{T}_{N+1}$ is obtained by replacing γ_N and w_{N+1} with $\hat{\gamma}_N$ and $\hat{w}_{N+1}$.

For the Gauss-Lobatto rule the remainder R_{GL} is

$$R_{GL}[f] = \frac{f^{(2N+2)}(\xi)}{(2N+2)!} \int_a^b (\lambda - z_1)(\lambda - z_2) \left[\prod_{j=1}^{N} (\lambda - t_j) \right]^2 d\alpha(\lambda).$$

In our case,

$$\frac{f^{(2N+2)}(\xi)}{(2N+2)!} = (\xi + \eta^2)^{-(2N+3)} > 0.$$

As a result, the sign of the remainder is determined by the sign of

$$\int_a^b (\lambda - z_1)(\lambda - z_2) \left[\prod_{j=1}^{N} (\lambda - t_j) \right]^2 d\alpha(\lambda).$$

Recall that a and b are the smallest and largest eigenvalues of the symmetric positive definite matrix A. If we set $z_1 = 0$ and $z_2 = \sqrt{||A||_1 ||A||_\infty} (> b)$, then

$$\int_a^b (\lambda - z_1)(\lambda - z_2) \left[\prod_{j=1}^{N} (\lambda - t_j) \right]^2 d\alpha(\lambda) < 0.$$

As a result, the remainder is negative and

$$\sum_{j=1}^{N} w_j f(t_j) + v_1 f(z_1) + v_2 f(z_2)$$

is an upper bound for $I[f]$, $N = 1, 2, \ldots$.

Golub and Meurant [12] proved that

$$\sum_{j=1}^{N} w_j f(t_j) + v_1 f(z_1) + v_2 f(z_2) = e_1^t f(\hat{T}_N) e_1.$$

In order to find upper bounds for $\bar{\mu}$, we set $z_1 = 0$ and $z_2 = \sqrt{||A^t A||_1 ||A^t A||_\infty}$. Next, we solve

$$(T_N - z_1 I)\delta = e_N, \quad (T_N - z_2 I)\mu = e_N,$$

and

$$\begin{pmatrix} 1 & -\delta_N \\ 1 & -\mu_N \end{pmatrix} \begin{pmatrix} \hat{w}_{N+1} \\ \hat{\gamma}_N^2 \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}.$$

$\hat{T}_{N+1}$ is defined as

$$\hat{T}_{N+1} = \begin{pmatrix} T_N & \hat{\gamma}_N e_N \\ \hat{\gamma}_N e_N^t & \hat{\omega}_{N+1} \end{pmatrix}$$

and $e_1^t(\hat{T}_{N+1} + \eta^2 I)^{-1} e_1$ gives an upper bound for $\bar{\mu}$, $N = 1, 2, \ldots$.

4.3 Numerical results for bounds

Lower bounds are studied for four different approximate solutions of test problems well1033 and illc1033 using Gauss and Gauss-Radau rules. The first approximate x_1 is the solution from using the QR method to solve the least squares problems. We perturb each element of x_1 by 0.01 times a random number generated from $U(0, 1)$ to get the second estimate x_2. Similarly, we perturb each element of x_1 by 0.1 and 10 times a random number generated from $U(0, 1)$ to get the estimates x_3 and x_4. We have the following three goals:

- See if the lower bounds are monotonically increasing.
- If so, see if the lower bounds converge to the true value.
- If so, see how the convergence behavior changes as the estimates become less accurate.

Upper bounds are studied for the same four approximate solutions using Gauss-Radau and Gauss-Lobatto rules, and again we try to answer the same three questions.

4.3.1 Lower bounds for the Gauss rule

Figure 4.1 displays the lower bounds calculated using the Gauss rule for these four approximate solutions for test problem well1033. The top left plot is for the approximate solution x_1. The top right plot is for x_2, the bottom left plot is for x_3 and the bottom right plot is for x_4. Figure 4.2 displays the lower bounds calculated using Gauss rule for these four approximate solutions for test problem illc1033. Together, the figures show that

- The lower bounds are monotonically increasing.
- They converge to the true value.
- They converge faster for less accurate approximates of the least squares problems.
- The bounds converge to the true value in just a few steps when the approximate solution is quite inaccurate.

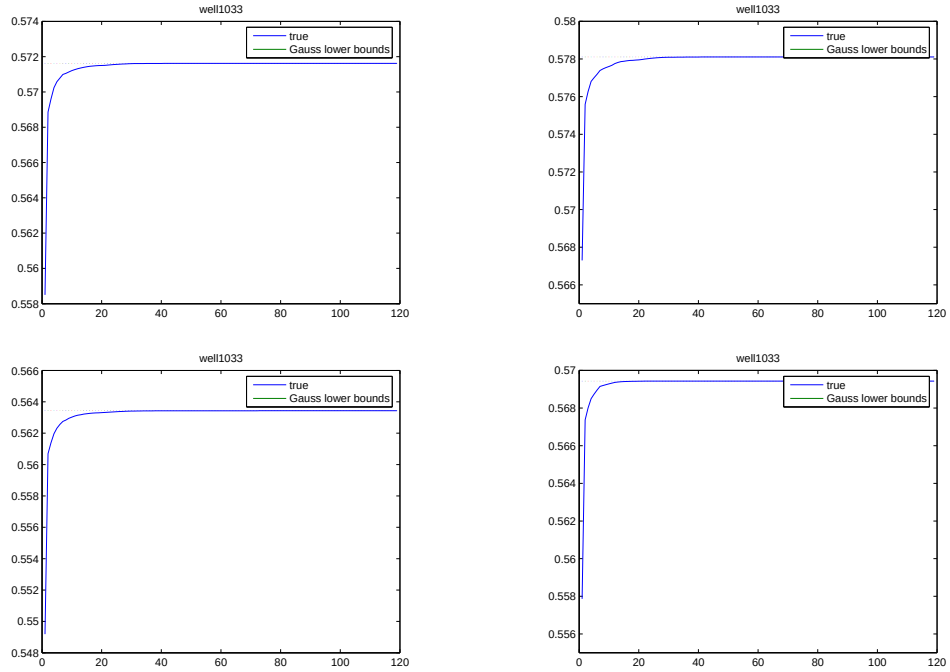

Figure 4.1: *Lower bounds calculated using Gauss rule for* x_i, $i = 1, 2, 3, 4$ *for well1033.*

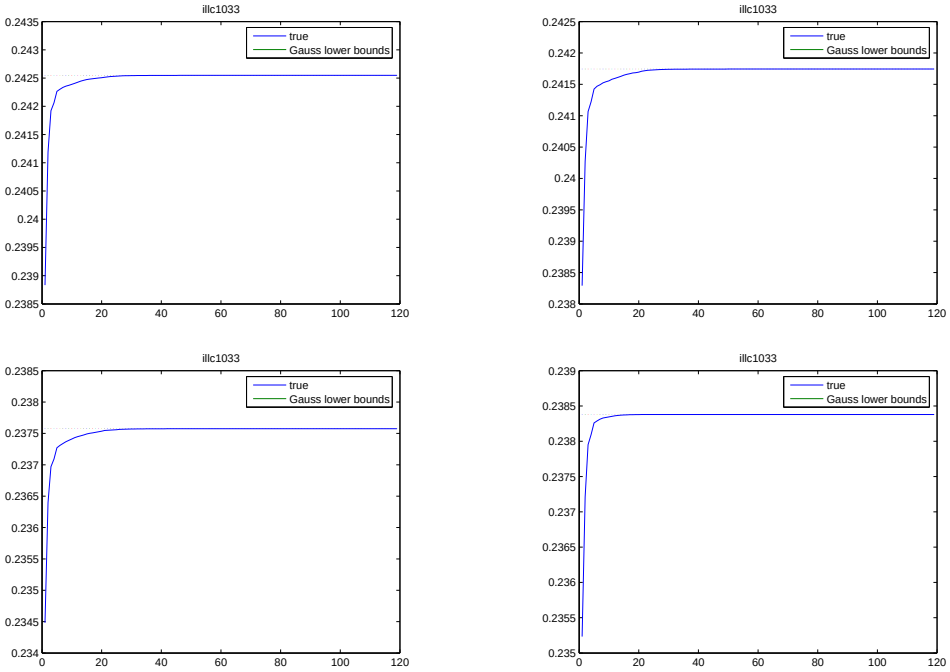

Figure 4.2: *Lower bounds calculated using Gauss for* x_i, $i = 1, 2, 3, 4$ *for illc1033.*

4.3.2 Upper and lower bounds for the Gauss-Radau rule

Figure 4.3 displays the upper bounds calculated using Gauss-Radau rule for these four approximate solutions for test problem well1033.

Figure 4.4 displays the upper bounds calculated using Gauss-Radau rule for these four approximate solutions for test problem illc1033.

Figure 4.5 displays the lower bounds calculated using Gauss-Radau rule for these four approximate solutions for test problem well1033.

Figure 4.6 displays the lower bounds calculated using Gauss-Radau rule for these four approximate solutions for test problem illc1033.

The figures show that

- The lower bounds are monotonically increasing.
- The upper bounds are monotonically decreasing.
- They converge to the true value.
- They converge faster for less accurate approximates of the least squares problems.
- The bounds converge to the true value in just a few steps when the approximate solution is quite inaccurate.

4.3.3 Upper bounds for the Gauss-Lobatto rule

Figure 4.7 displays the upper bounds calculated using Gauss-Lobatto rule for these four approximate solutions for test problem well1033.

Figure 4.8 displays the upper bounds calculated using Gauss-Lobatto rule for these four approximate solutions for test problem illc1033.

Again, the figures show that

- The upper bounds are monotonically decreasing.
- They converge to the true value.
- They converge faster for less accurate approximates of the least squares problems.
- The bounds converge to the true value in just a few steps when the approximate solution is quite inaccurate.

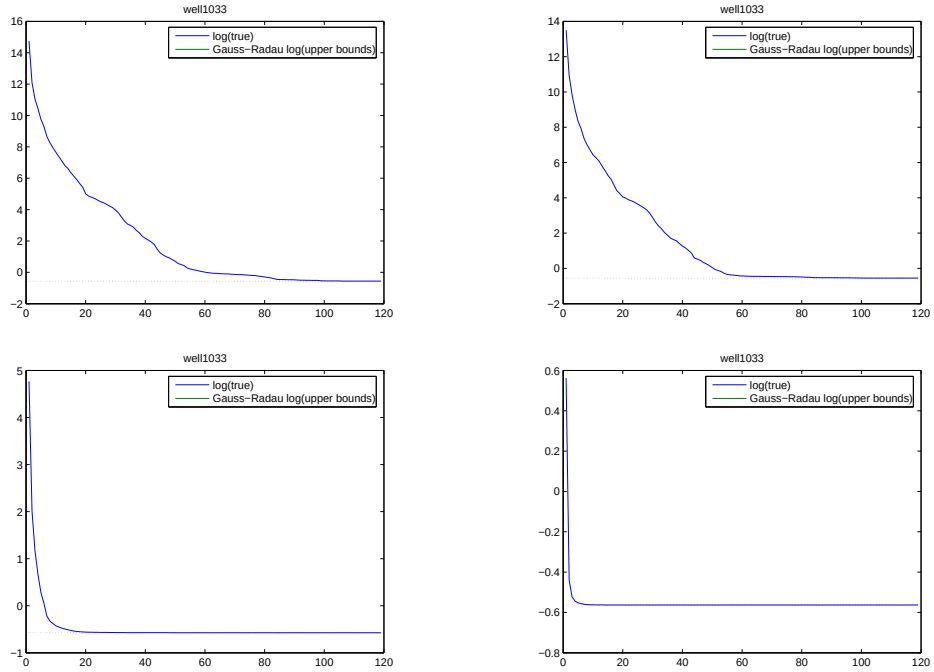

Figure 4.3: *Upper bounds calculated using Gauss-Radau rule for* x_i, $i = 1, 2, 3, 4$ *for well1033.*

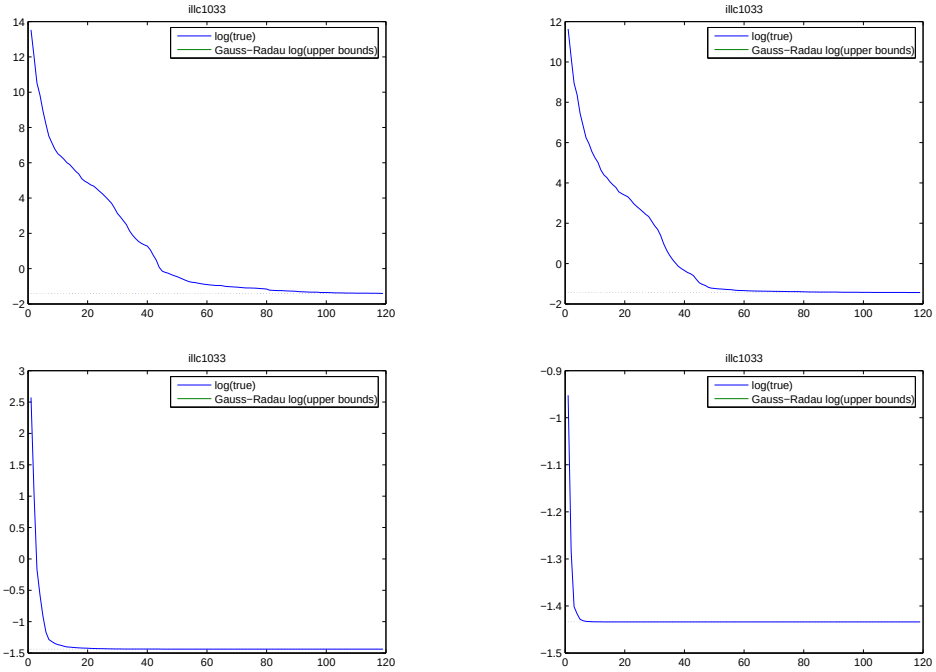

Figure 4.4: *Upper bounds calculated using Gauss-Radau rule for* x_i, $i = 1, 2, 3, 4$ *for illc1033.*

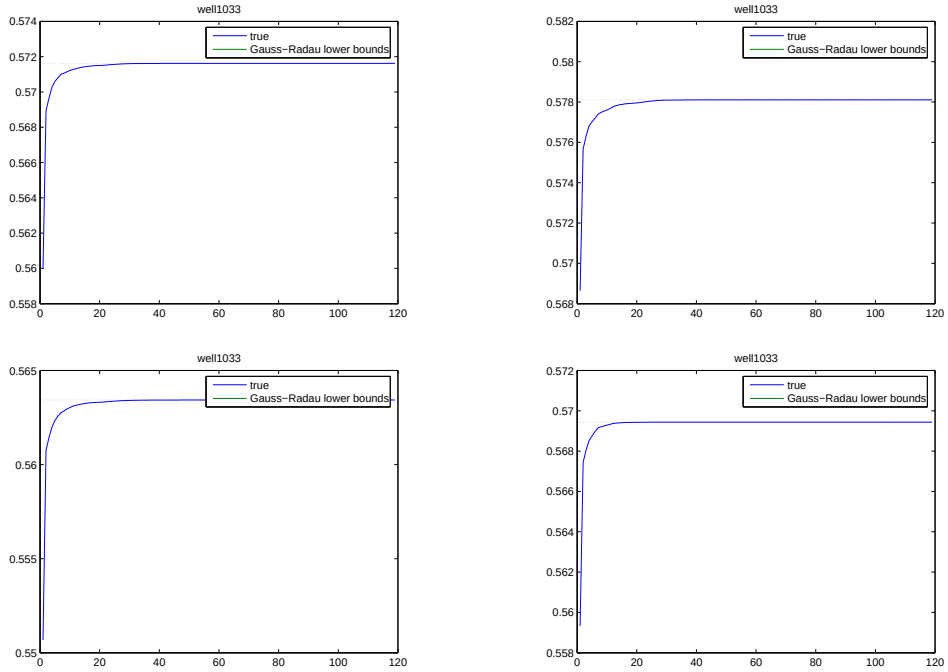

Figure 4.5: *Lower bounds calculated using Gauss-Radau rule for* x_i, $i = 1, 2, 3, 4$ *for well1033.*

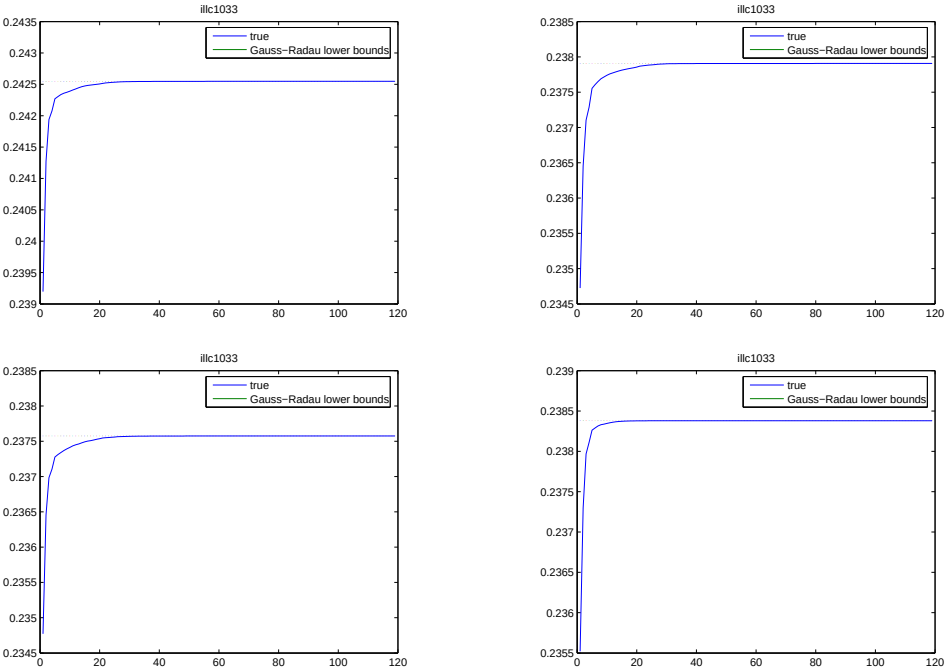

Figure 4.6: *Lower bounds calculated using Gauss-Radau rule for* x_i, $i = 1, 2, 3, 4$ *for illc1033.*

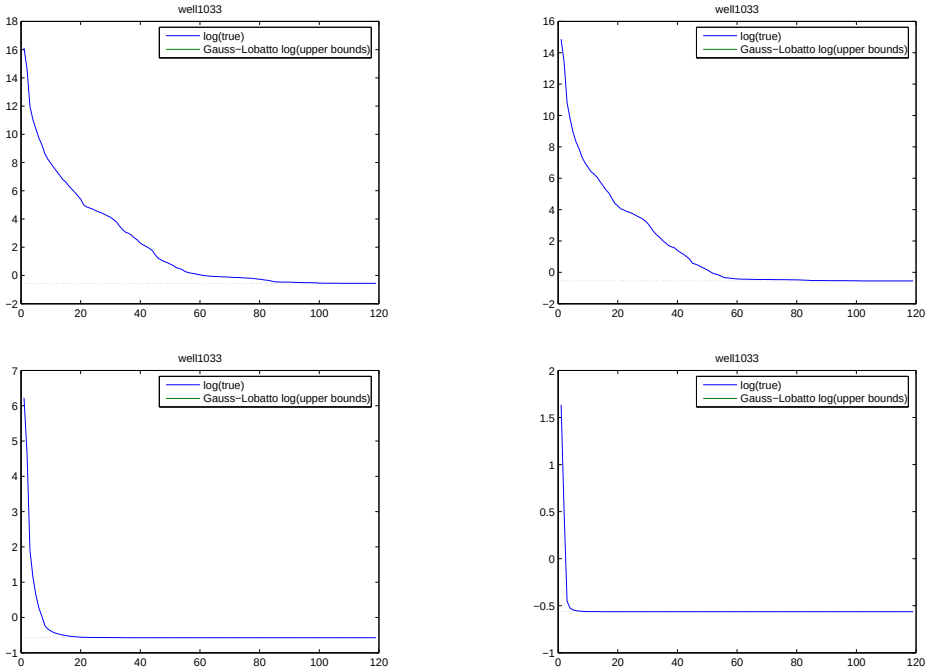

Figure 4.7: *Upper bounds calculated using Gauss-Lobatto rule for* x_i, $i = 1, 2, 3, 4$ *for well1033.*

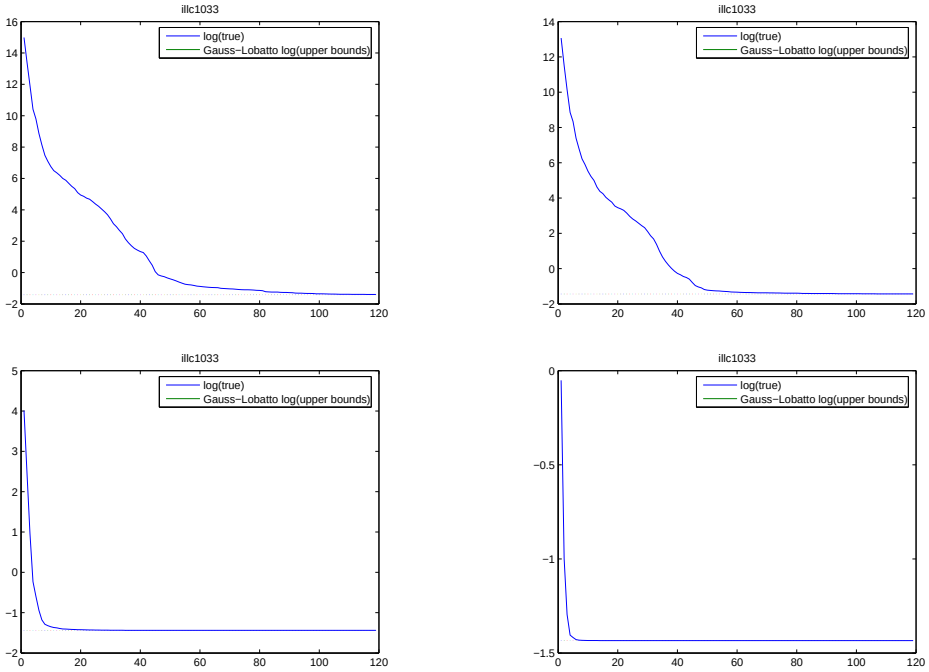

Figure 4.8: *Upper bounds calculated using Gauss-Lobatto rule for* $x_i, i = 1, 2, 3, 4$ *for illc1033.*

4.3.4 z_1 and z_2 fixed at the extreme eigenvalues of A^tA

We set

$$z_1 = 0 \text{ and } \sqrt{||A^tA||_1||A^tA||_\infty}$$

in the Gauss-Radau rule, and

$$z_1 = 0, \quad z_2 = \sqrt{||A^tA||_1||A^tA||_\infty}$$

in the Gauss-Lobatto rule. The latter values 0 and $\sqrt{||A^tA||_1||A^tA||_\infty}$ are the best lower and upper bounds we know for a and b, the smallest and largest eigenvalues of A^tA. It would be interesting to see if there is a big difference if we fix $z_1 = a$ and b for upper and lower bounds in Gauss-Radau rule and $z_1 = a$, $z_2 = b$ for upper bounds in Gauss-Lobatto rule.

Figure 4.9 displays the lower bounds calculated using the Gauss-Radau rule for test problem illc1033 with $z_1 = b$. Figure 4.10 displays the upper bounds calculated using the Gauss-Radau rule for test problem well1033 with $z_1 = a$. Figure 4.11 displays the upper bounds calculated using the Gauss-Labotto rule for test problem illc1033 with $z_1 = a$ and $z_2 = b$.

The figures show that no big improvement can be achieved by using $z_1 = a$ and b for the Gauss-Radau rule and $z_1 = a$, $z_2 = b$ for the Gauss-Lobatto rule instead of using $z_1 = 0$, and $\sqrt{||A||_1||A||_\infty}$ for the Gauss-Radau rule and $z_1 = 0$, $z_2 = \sqrt{||A||_1||A||_\infty}$ for the Gauss-Lobatto rule.

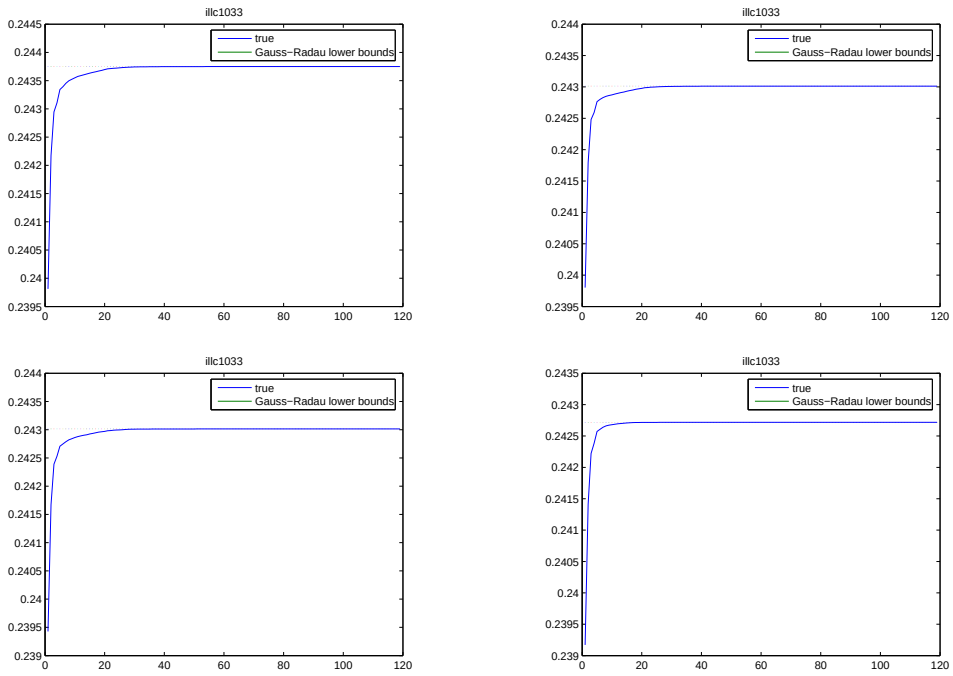

Figure 4.9: *Lower bounds calculated using the Gauss-Radau rule for illc1033 with* $z_1 = b$.

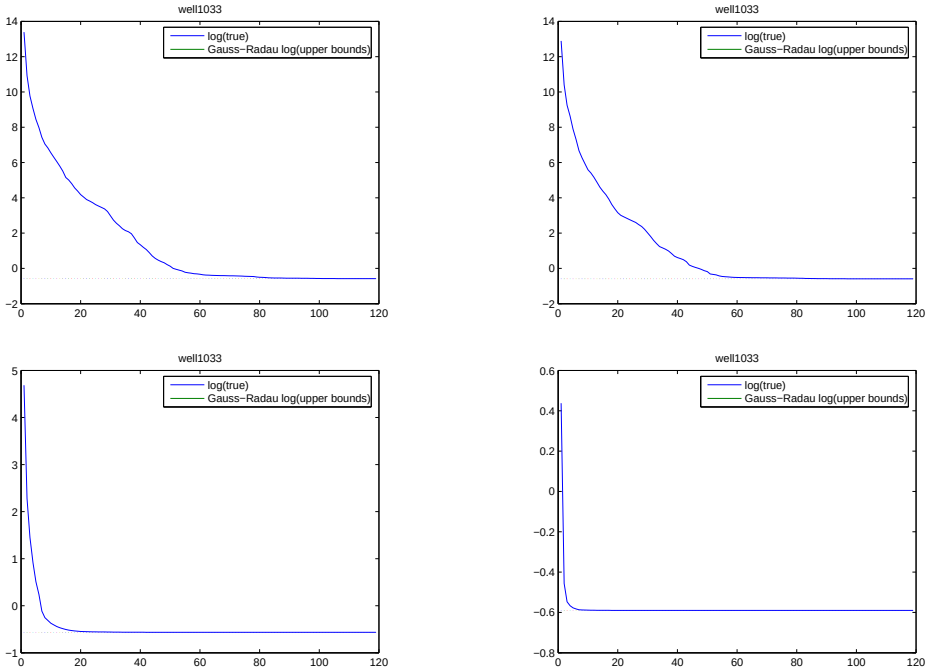

Figure 4.10: *Upper bounds calculated using the Gauss-Radau rule for well1033 with* $z_1 = a$.

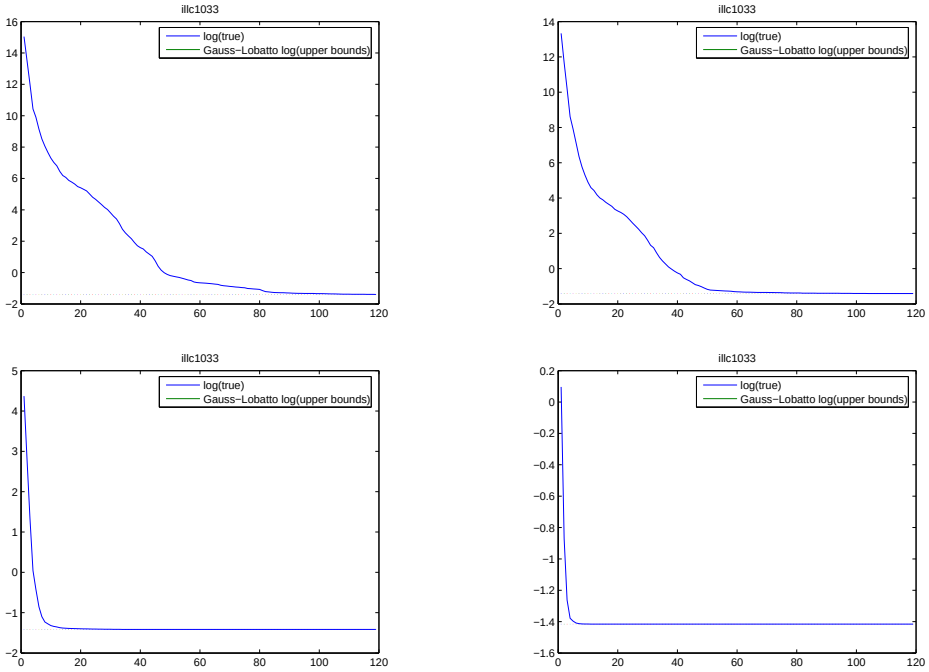

Figure 4.11: *Upper bounds calculated using Gauss-Lobatto for illc1033 with* $z_1 = a$ *and* $z_2 = b$.

Chapter 5

Conclusions

Several approaches are suggested and tested to evaluate an estimate for the optimal (that is, the minimal Frobenius norm) size of backward errors for LS problems. The numerical tests support the following conclusions.

Regarding the estimates:

1. The computed estimate of the optimal backward error is very near the true optimal backward error in all but a small percent of the tests.

   (a) Grcar's limit (1.6) for the ratio of the estimate to the optimal backward error appears to approach 1 very quickly.

   (b) The greater part of the fluctuation in the estimate is caused by rounding error in its evaluation.

2. Gu's lower bound (1.5) for the ratio of the estimate to the optimal backward error often fails in practice because of rounding error in evaluating the estimate.

3. As the computed solution of the LS problem becomes more accurate, the estimate may become more difficult to evaluate accurately due to the unavoidable rounding error in forming the residual.

4. For QR methods, evaluating the estimate is insignificant compared to the cost of solving a dense LS problem. A version of the estimate that does not either retain or recompute the orthogonal decomposition is less accurate.

5. When iterative methods become necessary, applying LSQR to the KW problem is a practical alternative to the bidiagonalization approach of Malyshev and Sadkane [22], particularly when x is close to x_*. No special coding is required (except a few new lines in LSQR to

compute $\|z_k\| \approx Ky$ as in section 2.6), and LSQR's normal stopping rules ensure at least some good digits in the computed $\widetilde{\mu}(x)$.

6. The smooth lower curves in Figures 3.3 and 3.4 suggest that when LSQR is applied to an LS problem, the backward errors for the sequence of approximate solutions $\{x_k\}$ are (unexpectedly) monotonically decreasing.

Regarding the bounds obtained from quadrature rules:

7. The computed lower bounds are monotonically increasing and the computed upper bounds are monotonically decreasing.

8. The computed bounds converge to the true value.

9. The computed bounds converge faster for less accurate approximations of the least squares solutions.

10. The computed bounds converge to the true value in just a few steps when the approximate solution is quite inaccurate.

11. No big improvement can be achieved by using $z_1 = a$ and b for the Gauss-Radau rule and $z_1 = a$, $z_2 = b$ for the Gauss-Lobatto rule instead of using $z_1 = 0$, and $\sqrt{||A||_1||A||_\infty}$ for Gauss-Radau and $z_1 = 0$, $z_2 = \sqrt{||A||_1||A||_\infty}$ for Gauss-Lobatto.

12. The LSQR based algorithm and the Gauss quadrature based algorithm give complementary results. The former converges faster for more accurate solutions while the latter converges faster for less accurate solutions.

Bibliography

[1] M. Adlers. *Topics in Sparse Least Squares Problems*. PhD thesis, Linköping University, Sweden, 2000.

[2] E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov, and D. Sorensen. *LAPACK Users' Guide*. Society for Industrial and Applied Mathematics, Philadelphia, 1992.

[3] S. Basu and N. K. Bose. Matrix Stieltjes series and network models. *SIAM J. Math Anal.*, 14(2):209–222, 1983.

[4] A. Björck. *Numerical Methods for Least Squares Problems*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1996.

[5] R. F. Boisvert, R. Pozo, K. Remington, R. Barrett, and J. Dongarra. The Matrix Market: a web repository for test matrix data. In R. F. Boisvert, editor, *The Quality of Numerical Software, Assessment and Enhancement*, pages 125–137. Chapman & Hall, London, 1997. The web address of the Matrix Market is `http://math.nist.gov/MatrixMarket/`.

[6] G. Calafiore, F. Dabbene, and R. Tempo. Radial and uniform distributions in vector and matrix spaces for probabilistic robustness. In D. E. Miller et al., editor, *Topics in Control and Its Applications*, pages 17–31. Springer, 2000. Papers from a workshop held in Toronto, Canada, June 29–30, 1998.

[7] G. Dahlquist, S. C. Eisenstat, and G. H. Golub. Bounds for the error of linear systems of equations using the theory of moments. *J. Math. Anal. Appl.*, 37:151–166, 1972.

[8] J. J. Dongarra, J. R. Bunch, C. B. Moler, and G. W. Stewart. *LINPACK Users' Guide*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1979.

[9] I. S. Duff, R. G. Grimes, and J. G. Lewis. Sparse matrix test problems. *ACM Transactions on Mathematical Software*, 15(1):1–14, 1989.

[10] G. H. Golub. Some modified matrix eigenvalue problems. *SIAM Review*, 15(2):318–334, 1973.

[11] G. H. Golub. Bounds for matrix moments. *Rocky Mnt. J. of Math.*, 4(2):207–211, 1974.

[12] G. H. Golub and G. Meurant. Matrices, moments and quadrature. *SCCM Report*, 1994.

[13] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, first edition, 1983.

[14] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, second edition, 1989.

[15] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, third edition, 1996.

[16] G. H. Golub and J. H. Welsch. Calculation of Gauss quadrature rule. *Math. Comp.*, 23:221–230, 1969.

[17] J. F. Grcar. Optimal sensitivity analysis of linear least squares. Technical Report LBNL-52434, Lawrence Berkeley National Laboratory, 2003. Submitted for publication.

[18] M. Gu. Backward perturbation bounds for linear least squares problems. *SIAM Journal on Matrix Analysis and Applications*, 20(2):363–372, 1999.

[19] M. Gu. New fast algorithms for structured linear least squares problems. *SIAM Journal on Matrix Analysis and Applications*, 20(1):244–269, 1999.

[20] N. J. Higham. *Accuracy and Stability of Numerical Algorithms*. Society for Industrial and Applied Mathematics, Philadelphia, second edition, 2002.

[21] R. Karlson and B. Waldén. Estimation of optimal backward perturbation bounds for the linear least squares problem. *BIT*, 37(4):862–869, December 1997.

[22] A. N. Malyshev and M. Sadkane. Computation of optimal backward perturbation bounds for large sparse linear least squares problems. *BIT*, 41(4):739–747, December 2002.

[23] MATLAB matrix laboratory, http://www.mathworks.com.

[24] P. Matstoms. Sparse QR factorization in MATLAB. *ACM Trans. Math. Software*, 20:136–159, 1994.

[25] G. Meurant. A review of the inverse of tridiagonal and block tridiagonal matrices. *SIAM J. Matrix Anal. Appl.*, 13(3):707–728, 1992.

[26] J. von Neumann and H. H. Goldstine. Numerical inverting of matrices of high order. *Bulletin of the American Mathematical Society*, 53(11):1021–1099, November 1947. Reprinted in [33, v. 5, pp. 479–557].

[27] W. Oettli and W. Prager. Compatibility of approximate solution of linear equations with given error bounds for coefficients and right-hand sides. *Numerische Mathematik*, 6:405–409, 1964.

[28] C. C. Paige and M. A. Saunders. LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Trans. Math. Software*, 8(1):43–71, 1982.

[29] C. C. Paige and M. A. Saunders. Algorithm 583; LSQR: Sparse linear equations and least-squares problems. *ACM Trans. Math. Software*, 8(2):195–209, 1982.

[30] M. A. Saunders. Computing projections with LSQR. *BIT*, 37(1):96–104, 1997.

[31] M. A. Saunders. lsqr.f, lsqr.m: Implementations of LSQR. http://www.stanford.edu/group/SOL/software/lsqr.html, 2002.

[32] G. W. Stewart. Research development and LINPACK. In J. R. Rice, editor, *Mathematical Software III*, pages 1–14. Academic Press, New York, 1977.

[33] A. H. Taub, editor. *John von Neumann Collected Works*. Macmillan, New York, 1963.

[34] B. Waldén, R. Karlson, and J.-G. Sun. Optimal backward perturbation bounds for the linear least squares problem. *Numerical Linear Algebra With Applications*, 2(3):271–286, 1995.

Part II

Monte Carlo Methods in Survival Analysis: Error Probabilities, Confidence Intervals, and Time-Sequential Clinical Trials

Chapter 6

Introduction

Part II of the dissertation explores test-based approaches to constructing confidence intervals in clinical trials with survival time as the primary response. It also develops importance resampling techniques to compute tail probabilities of the tests, thereby reducing the variance of the Monte Carlo estimate of an error probability and thus the number of simulations required to compute sample size and power in the design stage of a clinical trial and to construct confidence intervals from the trial's data. In this chapter we first review some basic methods and long-standing problems in survival analysis and then give an outline of the dissertation.

6.1 Censored rank statistics

Suppose we have K groups of censored data. Let $t_1 < t_2 < \ldots < t_D$ be the distinct death times in the pooled data. At time t_i we observe d_{ij} deaths in the jth group out of Y_{ij} individuals at risk, $j = 1, \ldots, K$, $i = 1, \ldots, D$. Let $d_i = \sum_{j=1}^{K} d_{ij}$ and $Y_i = \sum_{j=1}^{K} Y_{ij}$ be the number of deaths and the number at risk in the pooled data at time t_i, $i = 1, \ldots, D$.

Let $W_j(t)$ be a positive weight function with the property that $W_j(t_i)$ is zero whenever Y_{ij} is zero. We have the following censored rank statistics:

$$S_j = \sum_{i=1}^{D} W_j(t_i) \left\{ \frac{d_{ij}}{Y_{ij}} - \frac{d_i}{Y_i} \right\}, \quad j = 1, \ldots, K.$$

In practice, all commonly used statistics have a weight function $W_j(t_i) = Y_{ij} W(t_i)$. With this choice of weight functions,

$$S_j = \sum_{i=1}^{D} W(t_i) \left\{ d_{ij} - Y_{ij} \frac{d_i}{Y_i} \right\}, \quad j = 1, \ldots, K.$$

The weight function $W(t_i) = Y_i$ yields Gehan's (1965) generalization of the two-sample Mann-Whitney form of the Wilcoxon statistic, and the weight function $W(t) \equiv 1$ leads to the logrank statistic. These two statistics are used in Halpern and Brown's (1987) simulation program for the design of fixed-duration clinical trials, which will be reviewed in Section 7.1.

6.2 Estimation of the distribution function and its quantiles

In this section we review some techniques for drawing an inference about the distribution function and its quantiles based on a sample of censored survival data.

The survival function, i.e., the distribution of lifetime X, is

$$S(t) = P(X \geq t).$$

The hazard function is defined as

$$h(t) = P(X = t \mid X \geq t) = -\frac{d}{dt} \log S(t).$$

The cumulative hazard is the integral of the hazard function:

$$\Lambda(t) = \int_0^t h(u)\, du = -\log S(t).$$

A standard estimator of the survival function, proposed by Kaplan and Meier (1958), is called the product-limit estimator (also called the Kaplan-Meier estimator). This estimator is defined as

$$\hat{S}(t) = \begin{cases} 1 & \text{if } t < t_1, \\ \prod_{t_i \leq t} [1 - \frac{d_i}{Y_i}] & \text{if } t_1 \leq t, \end{cases}$$

where $t_1 < t_2 < \cdots < t_D$ are the D death times, d_i is the number of deaths at time t_i, and Y_i is the number of individuals who are at risk at time t_i. The Kaplan-Meier estimator is a step function with jumps at the observed death times. The size of these jumps depends not only on the number of deaths observed at each death time t_i but also on the pattern of the censored observations prior to t_i. The variance of the Kaplan-Meier estimator is estimated by Greenwood's formula:

$$\hat{V}[\hat{S}(t)] = \hat{S}(t)^2 \sum_{t_i \leq t} \frac{d_i}{Y_i(Y_i - d_i)}.$$

The pth quantile $\hat{\mu}_p$ of the Kaplan-Meier estimator $\hat{S}(t)$ is given as follows:

$$\hat{\mu}_p = \inf\{t : \hat{S}(t) \leq 1 - p\}.$$

In the case of $p = \frac{1}{2}$, $\hat{\mu}_{\frac{1}{2}}$ is the estimated median survival time.

6.3 Cox's proportional hazards model

Let $T_1, \ldots, T_n$; $C_1, \ldots, C_n$ be independent random variables. C_i is the censoring time associated with the survival time T_i. We observe $(Y_1, \delta_1), \ldots, (Y_n, \delta_n)$ where

$$Y_i = T_i \wedge C_i, \quad \delta_i = I(Y_i = T_i).$$

Let z_i be the vector of covariates associated with T_i.

Cox (1972) proposed the following proportional hazards model:

$$h(t; z) = e^{\beta' z} h_0(t),$$

where β is the vector of regression coefficients and $h_0(t)$ is the baseline hazard function.

Breslow's (1974) estimator of the cumulative hazard function from all the data is as follows:

$$\hat{\Lambda}(s) = \sum_{i:Y_i \leq s} \left\{ \delta_i \Big/ \left(\sum_{j \in R_i} e^{\hat{\beta}' z_j} \right) \right\},$$

in which $\hat{\beta}$ is Cox's (1972) estimate of β that maximizes the partial likelihood and $R_i = \{j : Y_j \geq Y_i\}$. Cox's partial likelihood can be expressed as

$$L(\beta) = \prod_{i=1}^{n} \left[\frac{e^{\beta' z_i}}{\sum_{j \in R_i} e^{\beta' z_j}} \right]^{\delta_i}.$$

We can estimate the baseline survival function by

$$\hat{S}(s) = \exp(-\hat{\Lambda}(s)).$$

This in turn yields the following estimate of the survival function of a patient with covariate z:

$$\hat{S}^z(s) = \exp(-\hat{\Lambda}(s) e^{\hat{\beta} z}).$$

6.4 Design of clinical trials with failure-time endpoints

A typical clinical trial tests whether a new treatment is better than an established one, which is called the control by clinicians. To be approved by the Food and Drug Administration (FDA), a new drug or treatment must perform as well as or no worse than the currently available one. The

test needs to satisfy a prescribed type I error probability, following the guidelines of the FDA.

An important part of planning a trial comparing two treatments is the determination of the approximate number of patients required to achieve adequate sensitivity of power for detecting a specified difference between the two treatments, if such difference exists. To design a clinical trial with failure-time endpoints, the null and alternative survival distributions need to be specified. Also, a censored rank statistic needs to be chosen. The goal of the design is to find the combination of accrual and follow-up times most attractive given the Type I error and the power. Monte Carlo simulations are often used in the design stage to compute power and sample size of clinical trials; see Halpern & Brown (1987).

For clinical trials with failure-time endpoints, it is often the case that certain of the observations are right censored at the time of analysis. Typically, this situation arises when the patients enter the clinical trial in sequence rather than all starting at the same time. At the end of the study, survival times are observed on those who have died, but for the survivors, the observations are censored and the time to censoring varies with the date of entry. In this case the determination of sample size at the design stage involves how long the accrual of patients should continue, given the rate of entry, in order to accrue and follow enough patients to obtain the desired power.

6.5 Data monitoring, interim analysis, and time-sequential designs

Clinical trials with time-to-event as the primary response are usually conducted for a period of several years. It is very advantageous if a definitive conclusion can be reached earlier than originally planned. There are substantial savings in the cost of the trial and the time saved can be used to put the drug on the market earlier and to be allocated to trials for other new treatments. Moreover, if there is enough evidence showing one treatment is better than the other, it is unethical to continue the trial in which some patients are randomly allocated to receive the inferior treatment. Typically, a Data and Safety Monitoring Board meets periodically to examine the treatment and adverse effects of the new drug. If the adverse effect of any drug or treatment turns out to be excessive, the trial is stopped for safety reasons.

To design a group sequential test, one needs to specify the stopping rule and the terminal decision rule that satisfy a prescribed type I error probability of the test, following the guidelines of the FDA.

Since the trial is typically monitored at prescribed calendar times, there are two time-scales in the problem. One is calendar time, while the other is "information time", which is related to how much information has been accrued at the calendar time of interim analysis. As explained in Section 8.1, these two time-scales create substantial difficulties in the analysis of group sequential clinical trials with time-to-event endpoints.

Table 6.1: Coverage errors in % for lower (L) and upper (U) confidence limits for normal mean for different values of $\sqrt{15}\mu$. Methods: N, naive normal; B, bootstrap.

	0		0.5		1		2.5		5	
Method	L	U	L	U	L	U	L	U	L	U
N	6.20	5.60	10.55	4.80	8.45	4.50	5.40	3.25	4.05	4.30
B	4.75	4.60	9.45	2.65	9.40	4.05	7.15	4.25	6.65	9.00

6.6 Confidence intervals following group sequential tests

We review in this section some methods for constructing confidence intervals following group sequential tests.

Let S_n be the partial sum of n independent and identically distributed normal random variables $X_1, \ldots, X_n$ with unknown mean μ and known variance 1. The stopping rule T for a group sequential test is a random variable taking values in the set $J = \{n_1, n_1 + n_2, \ldots, n_1 + \cdots + n_k\}$, where n_j is the jth group size. Consider stopping rules of the form

$$T = \min\{n \in J : S_n \geq b_n \quad \text{or} \quad S_n \leq a_n\}, \text{ with } a_n < b_n.$$

The special case $b_j = -a_j = c\sqrt{n_j}$ corresponds to Pocock's (1977) boundary.

If we ignore the group sequential nature and treat the experiment as if it were obtained from a sample of fixed size, then we have the naive confidence interval

$$(\bar{X}_T - z_{1-\alpha}/\sqrt{T}, \bar{X}_T - z_\alpha/\sqrt{T}),$$

where z_p is the pth quantile of the standard normal distribution. However, the confidence intervals thus constructed are biased toward the extremes and the coverage probabilities are not correct. In a group sequential setting, $T^{\frac{1}{2}}(\bar{X}_T - \mu)$ differs substantially from a standard normal random variable; see Fig. 1 of Chuang & Lai (1998).

The bootstrap method is known to give second-order accurate confidence intervals when the stopping rule T is replaced by a fixed sample size n. In the group sequential case, since $T^{\frac{1}{2}}(\bar{X}_T - \mu)$ is no longer an approximate pivot, the coverage errors of the bootstrap confidence intervals can differ substantially from the nominal values.

For example, if X_i are i.i.d. normal, $J = \{15j : j = 1, \ldots, 5\}$, and $T = \min\{n \in J : |S_n| \geq 2.413\sqrt{n}\}$, Table 6.1 taken from Chuang & Lai (1998) shows that both the naive method and the bootstrap method give coverage errors that differ substantially from the nominal values.

6.6.1 Ordering scheme for (T, S_T)

Exact confidence intervals have been developed by making use of various orderings of the sample space (T, S_T) when the group sizes are pre-determined constants. Under a total ordering $\leq$ of the sample space, an exact $1-2\alpha$ confidence interval for μ is $\mu_\alpha < \mu < \mu_{1-\alpha}$, where μ_c is the value of μ for which

$$\mathrm{P}_\mu\{(T, S_T) > (t_o, s_o)\} = c, \tag{6.1}$$

in which (t_o, s_o) denotes the observed value of (T, S_T).

Such confidence intervals were first introduced by Siegmund (1978) for stopping rules of the form

$$T = \min\{n \in J : S_n \geq b_n \quad \text{or} \quad S_n \leq a_n\}, \text{ with } a_n < b_n. \tag{6.2}$$

Siegmund used the following ordering of the sample space of (T, S_T) : $(t, s) > (\widetilde{t}, \widetilde{s})$ if and only if one of the following holds:

- $t = \widetilde{t}$ and $s > \widetilde{s}$,
- $t < \widetilde{t}$ and $s \geq b_t$,
- $t > \widetilde{t}$ and $\widetilde{s} \leq a_{\widetilde{t}}$.

Rosner & Tsiatis (1988) and Chang (1989) used an alternative ordering that is based on the signed root likelihood ratio statistic for testing a given value of μ. It is called the "likelihood ratio ordering", for which $(t, s) > (\widetilde{t}, \widetilde{s})$ whenever $t^{1/2}(s-\mu) > \widetilde{t}^{1/2}(\widetilde{s}-\mu)$. Emerson & Fleming (1990) proposed ordering (T, S_T) according to S_T/T. Under their "sample mean ordering", $(t, s) > (\widetilde{t}, \widetilde{s})$ whenever $s/t > \widetilde{s}/\widetilde{t}$.

In practice, however, the group sizes are often unpredictable instead of being pre-assigned constants; see §7.1 of Jennison & Turnbull (2000). Interim analyses are usually scheduled at fixed calendar times for administrative reasons, but patients are recruited at an uneven rate, so n_j is a random variable that is unobservable if $n_1 + \cdots + n_j$ exceeds the stopping time T. Since randomness of the n_j is due to the accrual pattern, which is unrelated to the X_i, we can assume that $\{n_1, \ldots, n_k\}$ is independent of $\{X_1, X_2, \ldots\}$.

For Siegmund's ordering, $(T, S_T) > (t_o, s_o)$ if and only if $S_{T\wedge t_o} > s_{T\wedge t_o}$, it only involves sample points that stop before or at t_o, and the group sizes n_j need only be specified for $j \leq j(t_o)$. We can therefore condition on $n_1, \ldots, n_{j(t_o)}$ in evaluating the probability in (6.1) when Siegmund's ordering is used, and thereby still obtain an exact $1-2\alpha$ confidence interval for μ even when it is not known how the n_j are generated for $j > j(t_o)$; see Lai & Li (2004).

This important property of Siegmund's ordering is not shared by the likelihood ratio and mean orderings. Under the last two orderings, the event $\{(T, S_T) > (t_o, s_o)\}$ contains sample points with $T > t_o$ when t_o is smaller than the largest allowable sample size $N = n_1 + \cdots + n_k$. Therefore,

unless one imposes assumptions on the (typically unknown) probability mechanism generating the group sizes after t_o, one cannot evaluate the probability in (6.1); see Lai & Li (2004).

6.6.2 A hybrid resampling method under Siegmund's ordering

A fundamental technique we use in our approach is the hybrid resampling method developed by Chuang & Lai (1998; 2000), which "hybridizes" the essential features of the bootstrap and exact methods. Following Chuang & Lai (2000), we give a brief description of the exact, bootstrap, and hybrid resampling methods for constructing confidence intervals.

Exact Method: The family of distributions is known except for the parameter of interest. If $\mathcal{F} = \{F_\theta : \theta \in \Theta\}$ is indexed by a real-valued parameter θ, an exact method can use test inversion to construct confidence intervals. Specifically, suppose that $R(\mathbf{X}, \theta_\mathbf{0})$ is the test statistic for the null hypothesis $H_0 : \theta = \theta_0$. Let $u_\alpha(\theta_0)$ be the α-quantile of the distribution of $R(\mathbf{X}, \theta_\mathbf{0})$ under distribution F_{θ_0}. An exact $1 - 2\alpha$ confidence set is given by

$$\{\theta : u_\alpha(\theta) < R(\mathbf{X}, \theta) < u_{1-\alpha}(\theta)\}.$$

Bootstrap method: A basic underlying assumption of the exact method is that there are no nuisance parameters, but this is rarely satisfied in practice. The bootstrap method replaces $F \in \mathcal{F}$ by an estimate $\hat{F}$ and θ by $\hat{\theta} = \theta(\hat{F})$, so that $u_\alpha(\theta)$ and $u_{1-\alpha}(\theta)$ can be approximated by u^*_α and $u^*_{1-\alpha}$, where u^*_p is the pth quantile of the distribution of $R(\mathbf{X}^*, \hat{\theta})$ with $\mathbf{X}^*$ generated from $\hat{F}$. The bootstrap method yields an approximate $1 - 2\alpha$ confidence set of the form $\{\theta : u^*_\alpha < R(\mathbf{X}, \theta) < u^*_{1-\alpha}\}$.

Hybrid resampling method: Whereas the bootstrap method replaces the family $\mathcal{F}$ in the exact method by the singleton $\{\hat{F}\}$, the hybrid resampling method replaces it by a one-parameter resampling family $\{\hat{F}_\theta, \theta \in \Theta\}$, where θ is the parameter of interest. Let $\hat{u}_\alpha(\theta)$ be the α-quantile of the sampling distribution of $R(\mathbf{X}, \theta)$ under the assumption that $\mathbf{X}$ has distribution $\hat{F}_\theta$. The hybrid resampling method yields an approximate $1 - 2\alpha$ confidence set of the form $\{\theta : \hat{u}_\alpha(\theta) < R(\mathbf{X}, \theta\} < \hat{u}_{1-\alpha}(\theta)\}$. It therefore involves two issues, the selection of the root $R(\mathbf{X}, \theta)$ and the resampling family $\{F_\theta\}$. Chuang & Lai (2000) discuss these issues in general settings and give specific examples in group sequential trials with fixed group sizes and in possibly non-ergodic autoregressive models and branching processes.

Suppose we remove the assumption of normally distributed X_i and only assume that the X_i have mean μ and variance 1. We can estimate G by the empirical distribution $\widehat{G}_T$ of $(X_i - \bar{X}_T)/\widehat{\sigma}_T$ $(1 \le i \le T)$, where $\widehat{\sigma}^2_T = T^{-1}\Sigma(X_i - \bar{X}_T)^2$ and $\bar{X}_T = S_T/T$. Let $\epsilon_1, \epsilon_2, \ldots$ be independent with common distribution $\widehat{G}_T$ and let $X_i(\mu) = \mu + \epsilon_i$. Let T_μ be the stopping rule applied to $X_1(\mu), X_2(\mu), \ldots$ instead of to $X_1, X_2, \ldots$, and let $S_n(\mu) = X_1(\mu) + \cdots + X_n(\mu)$. Approximating $\mathrm{pr}\{(T, S_T) \ge (t_o, s_o)\}$ in (6.1) by $\mathrm{P}\{(T_\mu, S_{T_\mu}(\mu)) > (t_o, s_o) \mid \widehat{G}_T\}$, an approximate $1 - 2\alpha$ confidence interval for μ is

Table 6.2: Coverage errors in % for lower (L) and upper (U) confidence limits for normal mean for different values of $\sqrt{15}\mu$. Methods: HS, hybrid resampling with Siegmund's ordering; HL, hybrid resampling with likelihood ratio ordering.

	0		0.5		1		2.5		5	
Method	L	U	L	U	L	U	L	U	L	U
HS	5.15	5.20	5.40	4.80	4.45	4.55	5.35	5.15	5.05	4.60
HL	5.25	5.20	5.40	4.65	5.40	4.55	5.40	5.55	4.90	4.55

$\widehat{\mu}_\alpha < \mu < \widehat{\mu}_{1-\alpha}$, where $\widehat{\mu}_c$ is the value of μ for which

$$\mathrm{P}\{(T_\mu, S_{T_\mu}(\mu)) > (t_o, s_o) \mid \widehat{G}_T\} = c. \tag{6.3}$$

The probability in (6.3) can be computed by Monte Carlo. This method for constructing confidence intervals is called the "hybrid resampling method", and is shown by Chuang & Lai (1998) to yield second-order accurate confidence intervals for μ as $N \to \infty$ when the group sizes $n_1, \ldots, n_k$ are nonrandom. Table 6.2 taken from Chuang & Lai (1998) shows that hybrid resampling with Siegmund's ordering and likelihood ratio ordering give coverage errors close to the nominal values for the same normal mean example for which both naive method and bootstrap method give poor coverage errors.

By conditioning on $n_1, \ldots, n_k$ and noting that the probability in (6.3) only involves $n_1, \ldots, n_{j(t_o)}$, Lai & Li (2004) established the second-order accuracy of $\widehat{\mu}_\alpha < \mu < \widehat{\mu}_{1-\alpha}$ when the n_i are random variables independent of $X_1, X_2, \ldots, X_N$, where $N = n_1 + \cdots + n_k$ is the maximum allowable sample size, in the following theorem.

Theorem 1 *Suppose $N/\min\{n_1, \ldots, n_k\}$ is bounded in probability as $N \to \infty$ and the stopping rule T is of the form* (6.2)*. Let $\psi(t)$ be the characteristic function of X_1 and assume that*

$$\limsup_{|t|\to\infty} |\psi(t)| < 1 \quad \textit{and} \quad E|X_1 - \mu|^r \le C$$

for some $r > 18$ and $C > 0$. Then the confidence interval $\widehat{\mu}_\alpha < \mu < \widehat{\mu}_{1-\alpha}$ has coverage probability $1 - 2\alpha + O(N^{-1})$, where $\widehat{\mu}_c$ is the value of μ that satisfies (6.3)*.*

6.7 Some long-standing problems and recent developments

It has been a long-standing problem concerning how confidence intervals can be constructed for the treatment effect following a group sequential clinical trial, in which the study duration or the number of subjects is a random variable that depends on the data collected so far, instead of being fixed in advance.

6.7.1 Multivariate quantiles and a general ordering scheme

The ordering schemes for (T, S_T) in Section 6.6.1 lead to corresponding bivariate quantiles of (T, S_T). Under a total ordering $\leq$ of the sample space of (T, S_T), (t, s) is called a pth *quantile* if

$$\mathrm{P}\{(T, S_T) \leq (t, s)\} = p,$$

assuming the X_i have a strictly increasing continuous distribution. This is a natural generalization of the pth quantile of a univariate random variable. For the general setting where a stochastic process X_u (in which u denotes either discrete or continuous time) is observed up to a stopping time T, we can likewise define $\mathbf{x} = (x_u,\ u \leq t)$ to be a pth *quantile* if

$$\mathrm{P}\{\mathbf{X} \leq \mathbf{x}\} \geq p \quad \text{and} \quad \mathrm{P}\{\mathbf{X} \geq \mathbf{x}\} \geq 1 - p$$

after we define a total ordering $\leq$ for the sample space of $\mathbf{X} = (X_u,\ u \leq T)$.

For applications to confidence intervals of a real parameter θ, the choice of the total ordering should be targeted towards the objective of interval estimation. Let $U_r,\ r \leq T$, be real-valued statistics based on the observed process $X_s,\ s \leq T$. Lai & Li (2004) proposed the following total ordering on the sample space of $\mathbf{X}$ via $(U_r,\ r \leq T)$:

$$\mathbf{X} \geq \mathbf{x} \quad \text{if and only if} \quad U_{T \wedge t} \geq u_{T \wedge t}, \tag{6.4}$$

where $(u_r,\ r \leq t)$ is defined from $\mathbf{x} = (x_r,\ r \leq t)$ the same as $(U_r,\ r \leq T)$ is defined from $\mathbf{X}$.

In particular, suppose X_i are independent normal. Let U_n be the sample mean $\bar{X}_n$ of $X_1, \ldots, X_n$. In this case, (6.4) yields the following ordering:

$$(T, S_T) \geq (t, s_t) \quad \text{if and only if} \quad \bar{X}_{T \wedge t} \geq s_{T \wedge t}/(T \wedge t). \tag{6.5}$$

Note that (6.5) is equivalent to $S_{\tau \wedge t} \geq s_{\tau \wedge t}$, which is the same as Siegmund's ordering for stopping rules T of the type (6.2). Thus (6.4) can be considered as a generalization of Siegmund's ordering; moreover, it relates Siegmund's ordering to the intuitively appealing ordering via sample means advocated by Emerson & Fleming (1990).

Like Siegmund's ordering, (6.4) has the attractive feature that the probability mechanism generating X_t only needs to be specified up to the stopping time T to define the quantile $\mathbf{x}$.

Lai & Li (2004) recently applied this ordering to construct confidence intervals for the treatment effect following group sequential trials in the case of univariate covariates.

Another long-standing problem is the construction of confidence intervals for median survival as a function of the covariates in the Cox model. Burr & Doss (1993; 1994) proposed a bootstrap method, which is reviewed in Section 7.3.

6.8 Outline of remaining chapters

Monte Carlo simulations are generally used in the design stage to compute power and sample size in clinical trials. The classic simulation program of Halpern and Brown (1987) is reviewed in Section 7.1. A clear disadvantage to Monte Carlo simulations is that they are computationally intensive. Importance resampling techniques are developed in Section 7.2 to compute tail probabilities, which can be incorporated into the simulation program of Halpern and Brown (1987) to reduce the computing time substantially.

Burr & Doss's (1993; 1994) method for constructing confidence bands of median survival in the Cox model is reviewed in Section 7.3. While their method involves estimating a probability density function in the denominator that can be quite unstable, a stable test-based method for constructing confidence intervals for median survival is developed in Section 7.2.4 via bootstrap. Simulation studies show that the confidence intervals thus constructed have coverage probabilities close to the nominal values.

In sequentially designed experiments the sample size is not fixed in advance but is a random variable that depends on the data collected so far. This creates bias in parameter estimation and introduces substantial difficulties in constructing valid confidence intervals. We explore in Section 8.2 the Monte Carlo computation of hybrid resampling confidence intervals following time-sequential tests in the Cox model. Confidence intervals for the treatment effect following group sequential trials for multivariate covariates using ordering with partial likelihoods are developed in Section 8.3. This partial likelihood based method is applied to the β-blocker heart attack trial and some other hypothetical clinical trial examples in Section 8.4, and it yields accurate coverage probabilities within 1% of the nominal values. In Section 8.5, by combining the two test-based methods for constructing confidence intervals for treatment effect and median survival, we construct test-based confidence regions for treatment effect as the primary endpoint and median survival as the secondary endpoint.

Chapter 7

Monte Carlo Methods for Clinical Trials

Monte Carlo simulations are generally used in the design stage to compute power and sample size in clinical trials. A clear disadvantage to Monte Carlo simulations is that they are computationally intensive. Importance resampling techniques are developed to compute tail probabilities, which can be incorporated into the classic simulation program of Halpern and Brown (1987) to reduce the computing time substantially.

Burr & Doss's (1993; 1994) method for constructing confidence bands of median survival in the Cox model needs to estimate a probability density function in the denominator, which can be an unstable process. Test-based confidence intervals for median survival are constructed via bootstrap. Simulation studies show that the confidence intervals thus constructed have coverage probabilities close to the nominal values.

7.1 The simulation program of Halpern and Brown (1987)

Monte Carlo simulations are generally used in the design stage to compute power and sample size of clinical trials. Halpern and Brown (1987) developed a simulation program for the design of fixed-duration clinical trials using Monte Carlo simulations. The program allows arbitrary specifications of the null and alternative survival distributions and either the Gehan test or the logrank test of the null hypothesis. The goal of the design is to find the combination of accrual and follow-up times most attractive given the Type I error and the power. A clear disadvantage of Monte Carlo simulations is that they are computationally intensive.

7.2 Importance resampling techniques

Here we develop importance resampling techniques to compute tail probabilities, which are used to reduce substantially the number of simulations required to compute power and sample size in the design of clinical trials.

7.2.1 Importance resampling concept

Following Hall (1992), we give a brief description of the concept of importance resampling. The method of importance resampling is a standard technique for improving the efficiency of Monte Carlo approximations; see Hammersley and Handscomb (1964). It was first suggested in the context of bootstrap resampling by Johns (1988) and Davison (1988).

Let $\chi = \{X_1, \ldots, X_n\}$ denote the sample from which a resample will be drawn. Under importance resampling, each X_i is assigned a probability p_i of being selected on any given draw, where $\Sigma p_i = 1$. Sampling is conducted with replacement, so that the chance of drawing a resample of size n in which X_i appears just m_i times $(1 \le i \le n)$ is given by a multinomial formula,

$$\frac{n!}{m_1! \ldots m_n!} \prod_{i=1}^{n} p_i^{m_i}.$$

Of course, $\Sigma m_i = n$. Taking $p_i = n^{-1}$ for each i, we obtain the uniform resampling method. The name "importance" derives from the fact that resampling is designed to take place in a manner that ascribes more importance to some sample values than to others. The aim is to select the p_i's so that the value assumed by a bootstrap statistic is relatively likely to be close to the quantity whose value we wish to approximate.

There are two parts to the method of importance resampling: first, a technique for passing from a sequence of importance samples to an approximation of a quantity that would normally be defined in terms of a uniform resample; and second, a method for computing the appropriate values of p_i's so as to minimize the error, or variability, of the approximation. We know that there are $N = \binom{2n-1}{n}$ different possible resamples. Let these be $\chi_1, \ldots, \chi_N$, indexed in any order, and let m_{ji} denote the number of times X_i appears in χ_j. The probability of obtaining χ_j after n resampling operations, under uniform resampling or importance resampling, is

$$\pi_j = \frac{n!}{m_{j1}! \ldots m_{jn}!} n^{-n}$$

or

$$\pi_j' = \frac{n!}{m_{j1}! \ldots m_{jn}!} \prod_{i=1}^{n} p_i^{m_{ji}} = \pi_j \prod_{i=1}^{n} (np_i)^{m_{ji}},$$

respectively. Let U be the statistic of interest, a function of the original sample. We wish to

construct a Monte Carlo approximation to the bootstrap estimate $\hat{u}$ of the mean of U, $u = E(U)$. Let χ^* denote a resample drawn by uniform resampling, and write U^* for the value of U computed from χ^*. Of course, χ^* will be one of the χ_j's. Write u_j for the value of U^* when $\chi^* = \chi_j$. In this notation,

$$\hat{u} = E(U^* \mid \chi) = \sum_{j=1}^{N} u_j \pi_j = \sum_{j=1}^{N} u_j \pi_j' \prod_{i=1}^{n} (np_i)^{-m_{ji}}.$$

Let χ^+ denote a resample drawn by importance resampling, write U^+ for the value of U computed from χ^+, and let M_i^+ be the number of times X_i appears in χ^+. Then,

$$\hat{u} = E\{U^+ \prod_{i=1}^{n} (np_i)^{-M_i^+} \mid \chi\}.$$

Therefore, it is possible to approximate $\hat{u}$ by importance resampling. In particular, if χ_b^+, $1 \le b \le B$ denote independent resamples drawn by importance resampling, and if U_b^+ equals the value of U computed for χ_b^+, then the importance resampling approximant of $\hat{u}$ is given by

$$\hat{u}_B^+ = B^{-1} \sum_{b=1}^{B} U_b^+ \prod_{i=1}^{n} (np_i)^{-M_{bi}^+}.$$

This approximation is unbiased, in the sense that $E(\hat{u}_B^+ \mid \chi) = \hat{u}$. Note too that conditional on χ, $\hat{u}_B^+ \to \hat{u}$ with probability 1 as $B \to \infty$. If we take each $p_i = n^{-1}$ then $\hat{u}_B^+$ is just the usual uniform resampling approximant $\hat{u}_B^*$. We wish to choose $p_1, \ldots, p_n$ to optimize the performance of $\hat{u}_B^+$. Since $\hat{u}_B^+$ is unbiased, the performance of $\hat{u}_B^+$ may be described in terms of variance:

$$\begin{aligned} \mathrm{var}(\hat{u}_B^+ \mid \chi) &= B^{-1} \mathrm{var}\{U_b^+ \prod_{i=1}^{n} (np_i)^{-M_{bi}^+} \mid \chi\} \\ &= B^{-1}(\hat{v} - \hat{u}^2), \end{aligned}$$

where

$$\begin{aligned} \hat{v} = \hat{v}(p_1, \ldots, p_n) &= E[\{U_b^+ \prod_{i=1}^{n} (np_i)^{-M_{bi}^+}\}^2 \mid \chi] \\ &= \sum_{j=1}^{N} \pi_j' u_j^2 \prod_{i=1}^{n} (np_i)^{-2m_{ji}} \\ &= \sum_{j=1}^{N} \pi_j u_j^2 \prod_{i=1}^{n} (np_i)^{-m_{ji}} \\ &= E\{U^{*2} \prod_{i=1}^{n} (np_i)^{-M_i^*} \mid \chi\}. \end{aligned}$$

On the last line, M_i^* denotes the number of times X_i appears in the uniform resample χ^*. Ideally we would like to choose $p_1, \ldots, p_n$ so as to minimize $\hat{v}(p_1, \ldots, p_n)$ subject to $\sum p_i = 1$.

In the case of estimating a distribution function there can be a significant advantage in choosing nonidentical p_i's, with the amount of improvement depending on the argument of the distribution function.

7.2.2 Two-sample problem with complete observations

For a two-sample problem, denote the two samples by $\chi_1 = \{X_1, \ldots, X_m\}$ and $\chi_2 = \{Y_1, \ldots, Y_n\}$. The Mann-Whitney form of the Wilcoxon statistics is given by

$$U(X_i, Y_j) = U_{ij} = \begin{cases} +1 & \text{if } X_i > Y_j, \\ 0 & \text{if } X_i = Y_j, \\ -1 & \text{if } X_i < Y_j, \end{cases}$$

$$U = \sum_{i=1}^{m} \sum_{j=1}^{n} U_{ij}.$$

In this case, we fix the first sample and use importance resampling on the second sample:

$$\begin{aligned} \hat{v} = \hat{v}(p_1, \cdots, p_n) &= E\{I(U^* \le x) \prod_{i=1}^{n} (np_i)^{-M_i^*} \mid \chi_2\} \\ &= E\{I\big(\sum_{i=1}^{n} M_i^* u_i \le x\big) \prod_{i=1}^{n} (np_i)^{-M_i^*} \mid \chi_2\} \\ &= E\{I\big(\sum_{i=1}^{n} M_i^* (u_i - \bar{u}) \le x - n\bar{u}\big) \prod_{i=1}^{n} (np_i)^{-M_i^*} \mid \chi_2\} \\ &= E\{I\big(\sum_{i=1}^{n} M_i^* \frac{u_i - \bar{u}}{\sqrt{\sum (u_i - \bar{u})^2}} \le \frac{x - n\bar{u}}{\sqrt{\sum (u_i - \bar{u})^2}}\big) \prod_{i=1}^{n} (np_i)^{-M_i^*} \mid \chi_2\} \\ &= E\{I\big(\sum_{i=1}^{n} M_i^* \tilde{u}_i \le \tilde{x}\big) \prod_{i=1}^{n} (np_i)^{-M_i^*} \mid \chi_2\} \\ &\sim E\{I(N_1 \le \tilde{x}) e^{N_2} \mid \chi_2\}, \end{aligned}$$

where (N_1, N_2) is bivariate normal with means $(0, \frac{1}{2}s^2)$, variances $(1, s^2)$ and covariance $\sum \tilde{u}_i \delta_i$. Here

$$u_j = \sum_{i=1}^{m} U_{ij}, \quad \delta_i = -\log(np_i), \quad s^2 = \sum \delta_i^2,$$

$$\bar{u} = \frac{\sum_{i=1}^{n} u_i}{n}, \quad \tilde{x} = \frac{x - n\bar{u}}{\sqrt{\sum (u_i - \bar{u})^2}}, \quad \tilde{u}_i = \frac{u_i - \bar{u}}{\sqrt{\sum (u_i - \bar{u})^2}}.$$

Table 7.1: $\hat{p} \pm s.e.$ for 2-sample Wilcoxon statistic, $m = 30$, $n = 25$, generated from an exponential distribution with a median of 3.

$\Phi(\tilde{x})$	$A(x)$	uniform	importance
0.005	2.6561	0.0044 ± 0.0046	0.0045 ± 0.0006
0.01	2.5704	0.0096 ± 0.0078	0.0110 ± 0.0012
0.025	2.1787	0.0279 ± 0.0132	0.0279 ± 0.0030
0.05	1.8940	0.0472 ± 0.0139	0.0464 ± 0.0040
0.1	1.5751	0.1023 ± 0.0214	0.0994 ± 0.0078
0.5	0.6120	0.5084 ± 0.0382	0.5038 ± 0.0278
0.95	0.0602	0.9752 ± 0.0110	0.9753 ± 0.0105

We now have

$$E\{I(N_1 \leq \tilde{x})e^{N_2} \mid \chi_2\} \quad = \quad \Phi(\tilde{x} - s\rho)e^{s^2},$$

where $\rho = \sum u_i \delta_i / \sqrt{\sum u_j^2} s$.

The values of s and ρ that minimize $\Phi(\tilde{x} - s\rho)e^{s^2}$ are $(s, \rho) = \pm(A, 1)$, where $A = A(x) > 0$ is chosen to minimize $\Phi(\tilde{x} - A)e^{A^2}$. Taking $\delta_i = A\tilde{u}_i + C$, where C is chosen to ensure that $\sum p_i = n^{-1} \sum e^{-\delta_i} = 1$, we see that $s \to A$ and $\rho \to 1$. Therefore, the minimum asymptotic variance of the importance resampling approximant occurs when

$$p_i = \frac{e^{-A\tilde{u}_i}}{\sum_{j=1}^{n} e^{-A\tilde{u}_j}}, \ 1 \leq i \leq n.$$

Table 7.1 shows that this importance resampling approach is considerably more effective for tail probabilities. Uniform resampling fails when the probability is too small, while importance resampling can still give an accurate estimate. When we have high probabilities and $A(x)$ is close to 0, there is not much difference between importance resampling and uniform resampling; therefore not much variance reduction can be achieved. If we wish to approximate a high probability $P(U^* \leq x \mid \chi)$, it is advisable to work throughout with $-U^*$ rather than U^* and use importance resampling to calculate $P(-U^* \leq -x \mid \chi) = 1 - P(U^* \leq x \mid \chi)$.

The Gehan statistic is another widely used statistic. Ordering the combined sample, defining

$$Z_{(1)} < \cdots < Z_{(m+n)},$$

and letting R_{1i} = rank of X_i, we have $R_1 = \sum_{i=1}^{m} R_{1i}$. Since

$$R_1 = \frac{m(m+n+1)}{2} + \frac{1}{2}U,$$

the same approach can be applied to the Gehan statistic.

In general, this approach may be applied to any statistic whose value computed from a uniform resample can be written as a linear combination of multinomial random variables M_i^* so that normal approximation theory can be applied. The rank statistics reviewed in Section 6.1 differ only by weights; therefore their values computed from a uniform resample can all be written as a linear combination of multinomial random variables M_i^* in the same way as the Gehan statistic. As a result, they yield the same exponential tilting. Halpern and Brown's (1987) simulation program allows the user to choose either the Gehan or the logrank statistic, which are two examples of this class of rank statistics.

7.2.3 Two-sample problem with censored data

For a two-sample problem with censored data, let

$$\chi_1 = \{(X_1, \delta_1), \ldots, (X_m, \delta_m)\}, \qquad \chi_2 = \{(Y_1, \gamma_1), \ldots, (Y_n, \gamma_n)\}.$$

Gehan defines

$$U(X_i, Y_j) = U_{ij} = \begin{cases} +1 & \text{if } (X_i > Y_j, \gamma_j = 1) \\ \text{or} & (X_i = Y_j, \delta_i = 0, \gamma_j = 1), \\ 0 & \text{otherwise,} \\ -1 & \text{if } (X_i < Y_j, \delta_i = 1) \\ \text{or} & (X_i = Y_j, \delta_i = 1, \gamma_i = 0), \end{cases}$$

$$U = \sum_{i=1}^{m} \sum_{j=1}^{n} U_{ij},$$

where $X_i = X_i^o \wedge W_i$, X_i^o is the variable of interest, W_i is some independent censoring variable, and $\delta_i = 1$ or 0 as $X_i = X_i^o$ or W_i.

One way of doing bootstrap is to obtain $X_i^{o*} \sim \hat{S}^o$ and $W_i^* \sim \hat{R}$ independently, where $\hat{S}^o$ and $\hat{R}$ are the Kaplan-Meier estimators of the distribution functions of X_i^o and W_i, respectively. Define $X_i^* = X_i^{o*} \wedge W_i^*$ and $\delta_i^* = 1$ or 0 as $X_i^* = X_i^{o*}$ or W_i^*.

Efron (1981) showed that the procedure is the same as resampling the pairs uniformly. As a result, the approach for uncensored data can be applied to censored data directly.

7.2.4 Treating censoring as ancillary

Alternatively, we can treat the censoring variable δ_i as ancillary. Let $\hat{S}$ and $\hat{C}$ be the Kaplan-Meier estimators of the survival distribution and the censoring distribution of the pooled data, respectively. If X_i is uncensored, then W_i is censored by X_i^o, and we generate $W_i^* \geq X_i$ from $\hat{R}$ by rejection sampling and then generate $X_i^* \leq W_i^*$ from $\hat{S}$ by rejection sampling. If X_i is censored, there is no need to generate X_i^*. The above simulation method preserves the censoring indicator.

7.2.5 More general statistics

Let $X_1, \ldots, X_n$ be i.i.d. p-dimensional random vectors. For any symmetric statistic $S = S(X_1, \ldots, X_n)$, we have the following i.i.d. approximation to S (Lai and Wang, 1993):

$$S = n\mu + \sum_{i=1}^{n} \beta(X_i) + R_n,$$

where β is a nonrandom Borel function that is invariant under permutation of the arguments, and R_n is of smaller order of magnitude. Thus,

$$S \approx n\mu + \sum_{i=1}^{n} \beta(X_i)$$

and the weights for these statistics are

$$p_i = \frac{e^{-A\beta(X_i)}}{\sum_{j=1}^{n} e^{-A\beta(X_j)}}.$$

Censoring is also addressed by this approximation because the censoring random variable can be incorporated into the X_i, i.e., $X_i = (Y_i, \delta_i)$, where Y_i is the observed lifetime.

7.2.6 Improvements of Halpern & Brown's simulation program

The importance resampling techniques can be incorporated in the simulation program of Halpern and Brown (1987) to reduce the computational cost substantially for the design of fixed-duration clinical trials. When computing the power of a test, Halpern & Brown's program needs to run a large number of simulations to make sure that the variance of the Monte Carlo estimate of the power is small. For each simulation a sample of survival time needs to be generated from the survival distributions. By incorporating the importance resampling techniques in the program, we only need to generate one sample and then do importance resampling on top of it. Since the importance resampling techniques reduce the variance of the Monte Carlo estimate of the power greatly, far fewer simulations are needed and thus the computation time is greatly reduced.

7.3 The Burr-Doss confidence bands of median survival

Burr & Doss (1993; 1994) constructed confidence bands for median survival as a function of the covariates in the Cox model.

Let ξ_p be the pth quantile of the distribution of the life length of an individual with covariate x in the Cox model. Burr & Doss introduced an estimator $\tilde{\xi}_p$ of ξ_p and developed confidence bands. First, they showed that as $n \to \infty$, where n is the number of individuals in the study,

$\sqrt{n}(\hat{\xi}_p(x) - \xi_p(x))$ converges weakly to a Gaussian process $W(x)$ with a complicated covariance structure. They then estimated this covariance structure from the data, and simulated many Gaussian processes with this estimated covariance structure. The critical constants required for the construction of the confidence bands are obtained from the simulated processes.

When estimating the covariance structure of $W(x)$, the method needs to estimate a probability density function in the denominator. This process can be quite unstable.

7.4 Improved test-based bootstrap confidence intervals

In this section we propose a test-based stable method to construct confidence intervals for median survival as a function of the covariates in the Cox model using bootstrap.

For survival data, Brookmeyer and Crowley (1982) proposed the test-based interval estimator

$$\left\{ t : |\hat{S}(t) - \frac{1}{2}| \leq \hat{\xi}_S(t) z_{1-\alpha/2} \right\}$$

for the median survival time, where $\hat{S}_t$ is the Kaplan-Meier estimator of the survival function, $\hat{\xi}_S^2(t)$ is the Greenwood's estimator for the variance of $\hat{S}(t)$, and $z_{1-\alpha/2}$ is the $(1-\alpha/2)$th quantile of standard normal.

We can generalize this idea to construct confidence intervals for median survival as a function of the covariates in the Cox model.

A $1-2\alpha$ confidence region is the set of all parameters not rejected by

$$\left\{ t : c_\alpha(t) \leq \frac{\hat{S}_z(t) - \frac{1}{2}}{\hat{\xi}_z(t)} \leq c_{1-\alpha}(t) \right\},$$

where z is the covariate, $\hat{S}_z(t)$ is an estimate of $S_z(t)$, the survival function with covariate z, $\hat{\xi}_z^2(t)$ is an estimate of $\xi_z^2(t)$, the variance of $S_z(t)$, and c_p is the pth quantile of the distribution of $\frac{\hat{S}_z(t)-\frac{1}{2}}{\hat{\xi}_z(t)}$.

For every t, instead of using normal quantiles, we use $\hat{c}_\alpha$ and $\hat{c}_{1-\alpha}$, the αth and $(1-\alpha)$th bootstrap quantiles of

$$\frac{\hat{S}_z^*(t) - \hat{S}_z(t)}{\hat{\xi}_z^*(t)}$$

to estimate c_α and $c_{1-\alpha}$. Here, $\hat{S}_z^*(t)$ is the value of $\hat{S}_z(t)$ computed from a bootstrap sample.

Let $\hat{\Lambda}(t)$ be Breslow's (1974) estimator of the cumulative hazard function. The estimate of the baseline survival function is given by $\hat{S}_0(t) = \exp(-\hat{\Lambda}(t))$ and the estimate of the survival function with covariate z is given by $\hat{S}_z(t) = \hat{S}_0(t)^{\exp(\hat{\beta}z)}$, where $\hat{\beta}$ maximizes the partial likelihood at time t. Under mild regularity conditions, $\hat{S}_z(t)$ has an asymptotic normal distribution with mean $S_z(t)$ and a variance that can be estimated by $\hat{\xi}_z^2(t)$; see Klein & Moeschberger (2003).

Table 7.2: *Example* 7.1: Coverage errors in % for lower (L) and upper (U) confidence limits and coverage probabilities (P) of confidence intervals for median survival of the control group and the treatment group.

	$\beta = 0$			$\beta = \log \frac{2}{3}$			$\beta = \log \frac{1}{2}$		
	L	U	P	L	U	P	L	U	P
Control	4.75	5.95	89.30	4.35	4.70	90.95	4.95	5.80	89.25
Treatment	4.95	5.90	89.15	4.55	6.10	89.35	4.95	5.95	89.10

We now have

$$\hat{\xi}_z^2(t) = \hat{S}_z^2(t)[Q_1(t) + Q_2(t; z)],$$

where

$$Q_1(t) = \sum_{i=1}^{n} \frac{\delta_i(t)}{W^{(0)}(Y_i(t), \hat{\beta})^2},$$

$$Q_2(t; z) = \left(\sum_{i=1}^{n} \left[\frac{W^{(1)}(Y_i(t), \hat{\beta})}{W^{(0)}(Y_i(t), \hat{\beta})} - z \right] \left[\frac{\delta_i(t)}{W^{(0)}(Y_i(t), \hat{\beta})} \right] \right)^2 / \hat{V}(\hat{\beta}),$$

$$\hat{V}(\hat{\beta}) = \delta_i(t) \left[\sum_{i=1}^{n} \frac{W^{(2)}(Y_i(t), \hat{\beta})}{W^{(0)}(Y_i(t), \hat{\beta})} - \sum_{i=1}^{n} \left(\frac{W^{(1)}(Y_i(t), \hat{\beta})}{W^{(0)}(Y_i(t), \hat{\beta})} \right)^2 \right].$$

Here, $W^{(k)}(Y_i(t), \hat{\beta}) = \sum_{j \in R_i(t)} z_j^k \exp(\hat{\beta} z_j)$ for $k = 0, 1, 2$. It is straightforward to apply this method to construct confidence bands instead of confidence intervals.

Burr & Doss (1993; 1994) did not do coverage studies with their method, which makes it hard to compare our method with theirs numerically. We demonstrate the accuracy of our method through the following example.

Example 7.1. Consider a trial in which $n = 100$ patients enter the trial uniformly during a 3-year recruitment period and are randomized to treatment or control with probability $\frac{1}{2}$. The trial is designed to last for 5.5 years.

Table 7.2 reports a simulation study of the coverage errors of upper and lower confidence limits for median survival of treatment and control, with nominal error $\alpha = 0.05$. It also gives the coverage probabilities of the two-sided confidence intervals with 90% nominal coverage probability. This simulation study assumes that the lifetimes of the control group have an exponential distribution with mean 3 years, and that those of the treatment group have an exponential distribution with mean $3e^{-\beta}$ years, with $e^{\beta} = 1, \frac{2}{3}, \frac{1}{2}$. Table 7.2 shows that the method yields quite accurate confidence intervals, with all probabilities within 1% of their nominal values.

7.4.1 Computation of the confidence limits

We now describe an algorithm, based on the method of successive secant approximations, to find the limits of the confidence interval. To find the upper limit, we let

$$f(t) = \frac{\hat{S}_z(t) - \frac{1}{2}}{\hat{\xi}_z(t)} - \hat{c}_\alpha(t)$$

and solve the equation $f(t) = 0$. First we find $a_1 < b_1$ such that $f(a_1) > 0$ and $f(b_1) < 0$. Let $f_1(t)$ be linear in $t \in [a_1, b_1]$ with $f_1(a_1) = f(a_1)$ and $f_1(b_1) = f(b_1)$, and let t_1 be the root of $f_1(t) = 0$. If $f(t_1) > 0$, set $a_2 = t_1$ and $b_2 = b_1$. If $f(t_1) < 0$, set $b_2 = t_1$ and $a_2 = a_1$. We can proceed inductively in this manner, letting $f_k(t)$ linearly interpolate $f(a_k)$ and $f(b_k)$ for $a_k \leq t \leq b_k$, and letting t_k be the root of $f_k(t) = 0$. This procedure terminates if t_k differs little from t_{k-1} or if k reaches some upper bound, and the terminal value of t_k is taken to be the upper limit. The same induction process can be applied to

$$f(t) = \frac{\hat{S}_z(t) - \frac{1}{2}}{\hat{\xi}_z(t)} - \hat{c}_{1-\alpha}(t)$$

to find the lower limit of the confidence interval. Chuang & Lai (2000) used a similar approach to compute limits of confidence intervals.

Chapter 8

Confidence Intervals in Time-Sequential Trials

8.1 Introduction and background

8.1.1 Time-sequential clinical trials

Suppose that n patients enter a trial serially. Let $T_i \geq 0$ denote the entry time and $Y_i > 0$ the survival time after entry of the ith patient. Interim analyses of the trial are scheduled at calendar times t_j $(1 \leq j \leq k)$, $0 < t_1 < \cdots < t_k$, where $t_k = t^*$ is the pre-scheduled duration of the trial. The data at calendar time t consist of $(Y_i(t), \delta_i(t), z_i),\ i = 1, \ldots, n$, where

$$Y_i(t) = \min\{Y_i, \xi_i, (t - T_i)^+\},\ \delta_i(t) = I_{\{Y_i(t)=Y_i\}},$$

z_i is the covariate, and ξ_i is the withdrawal time of the ith patient. This is illustrated schematically in Figure 8.1.

Since a time-sequential trial is typically monitored at prescribed calendar times, there are two time-scales in the problem. One is calendar time, while the other is "information time", which is related to how much information has been accrued at the calendar time of the interim analysis. These two time-scales create substantial difficulties in the analysis of group sequential clinical trials with time-to-event endpoints because there is no simple relationship between them. As a result, one does not know at interim analysis the information time that corresponds to the calendar time when the trial ends.

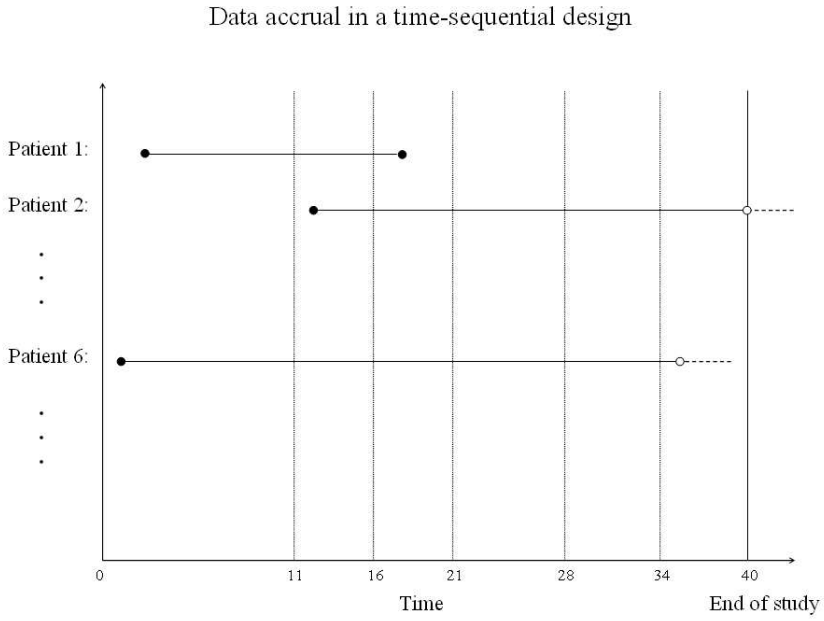

Figure 8.1: *Data accrual in a time-sequential design. Patient 1 entered the trial 40 days after the trial began and died 72 days after the 2nd interim analysis; patient 2 entered the trial 15 days after the 1st interim analysis and was still alive at the end of the trial; patient 6 entered the trial 12 days after the trial began and was lost to follow-up 28 days after the 5th interim analysis.*

8.1.2 Review of Lai & Li (2004)

We review in this section the hybrid resampling method Lai & Li (2004) proposed to overcome the difficulties due to the two different time-scales in constructing valid confidence intervals, following a time-sequential test, for the regression parameter in a Cox model with univariate covariates.

Assume that T_i is independent of (Y_i, ξ_i, z_i) and ξ_i is independent of (z_i, Y_i). Also assume that the hazard function of Y_i is given by Cox's (1972) proportional hazards model

$$\mathrm{P}\{y \le Y_i \le y + dy \mid Y_i \ge y, z_i\} = e^{\beta z_i} d\Lambda(y),$$

where β is the regression parameter and Λ is the baseline cumulative hazard function. To test the null hypothesis $H_0 : \beta = 0$, we can differentiate the log partial likelihood for β at $\beta = 0$ and calendar time t to get Cox's score statistic,

$$S_n(t) = \sum_{i=1}^{n} \delta_i(t) \left\{ z_i - \left(\sum_{j \in R_i(t)} z_j \right) \Big/ |R_i(t)| \right\}, \tag{8.1}$$

where $R_i(t) = \{j : Y_j(t) \geq Y_i(t)\}$ and $|R_i(t)|$ denotes the size of $R_i(t)$. The observed Fisher information at calendar time t is

$$V_n(t) = \sum_{i=1}^{n} \delta_i(t) \left[\sum_{j \in R_i(t)} z_j^2/|R_i(t)| - \left\{ \sum_{j \in R_i(t)} z_j/|R_i(t)| \right\}^2 \right], \tag{8.2}$$

which provides an estimate of the null variance of $S_n(t)$. Asymptotic theory suggests use of a repeated significance test that rejects H_0 at the jth interim analysis ($1 \leq j \leq k$) if

$$S_n(t_j)/V_n^{1/2}(t_j) \geq b_j \quad \text{or} \quad S_n(t_j)/V_n^{1/2}(t_j) \leq a_j, \tag{8.3}$$

and stops the trial as soon as (8.3) occurs, where $a_j < 0 < b_j$.

Let τ be the stopping time of the trial. Denote $S_n(t)$ by $S(t)$ and $V_n(t)$ by $V(t)$ for notation simplicity. A standard approach in the literature is to use the space-time Brownian motion approximation of $(S(t), V(t))$ (see Jones & Whitehead (1979) and Siegmund (1985)), to which Siegmund's ordering can be applied because stopping rule (8.3) has the form (6.2) under this approximation.

Letting $\Psi_t = S(t)/V(t)$, Lai & Li (2004) proposed a general ordering to construct confidence intervals for treatment effect assuming univariate covariates, which orders the sample space of (τ, Ψ_τ) by

$$(\tau_1, \Psi_{\tau_1}^{(1)}) \leq (\tau_2, \Psi_{\tau_2}^{(2)}) \quad \text{if and only if} \quad \Psi_{\tau_1 \wedge \tau_2}^{(1)} \leq \Psi_{\tau_1 \wedge \tau_2}^{(2)}.$$

Similar to the normal mean case, let $p(\beta) = \text{pr}_\beta\{(\tau, \Psi_\tau) > (\tau, \Psi_\tau)_{\text{obs}}\}$, where $(\tau, \Psi_\tau)_{\text{obs}}$ denotes the observed value of (τ, Ψ_τ). Then $\{\beta : \alpha < p(\beta) < 1 - \alpha\}$ is a $1 - 2\alpha$ confidence set for β. The probability $p(\beta)$ has to be evaluated by simulation. Lai & Li (2004) replaced G by $\widehat{G} = 1 - e^{-\hat{\Lambda}}$, where $\widehat{\Lambda}$ is Breslow's (1974) estimator of the cumulative hazard function from all the data at the end of the trial:

$$\widehat{\Lambda}(s) = \sum_{i: Y_i(\tau) \leq s} \left\{ \delta_i(\tau) \Big/ \left(\sum_{j \in R_i(\tau)} e^{\hat{\beta} z_j} \right) \right\},$$

in which $\widehat{\beta}$ is Cox's (1972) estimate of β that maximizes the partial likelihood at time τ. They also replaced C by the Kaplan-Meier estimator $\widehat{C}$. Thus, $p(\beta)$ was replaced by

$$\widehat{p}(\beta) = \text{P}\{(\tau^{(\beta)}, \Psi_{\tau^{(\beta)}}^{(\beta)}) > (\tau, \Psi_\tau)_{\text{obs}}\}, \tag{8.4}$$

where the superscript (β) means that the observations are generated with regression parameter β. Usually $\widehat{p}(\beta)$ is monotone in β, so the confidence set $\{\beta : \alpha < \widehat{p}(\beta) < 1 - \alpha\}$ with approximate coverage probability $1 - 2\alpha$ can be expressed as an interval, whose endpoints $\underline{\beta} < \bar{\beta}$ are defined by $\widehat{p}(\underline{\beta}) = \alpha$, $\widehat{p}(\bar{\beta}) = 1 - \alpha$. The following example and Table 8.1 taken from Lai & Li (2004) show that their hybrid resampling method gives coverage probabilities close to the nominal values while the Brownian motion approximation does not give accurate coverage probabilities.

Table 8.1: *Example* 8.1: Coverage errors in % for lower (L) and upper (U) confidence limits and coverage probabilities (P) of confidence intervals for β. Methods: H, hybrid resampling; S, Siegmund; N, naive normal.

	$\beta = 0$			$\beta = \log \frac{2}{3}$			$\beta = \log \frac{1}{2}$		
Method	L	U	P	L	U	P	L	U	P
H	4.45	4.55	91.00	5.25	5.35	89.40	5.05	4.05	90.90
S	4.45	5.05	90.50	4.65	0.35	95.00	5.75	0.00	94.25
N	4.15	5.05	90.80	5.80	7.75	86.45	3.75	3.15	93.10

Example 8.1. Consider a time-sequential trial in which $n = 350$ patients enter the trial uniformly during a 3-year recruitment period and are randomized to treatment or control with probability $\frac{1}{2}$. The trial is designed to last for a maximum of $t^* = 5.5$ years, with interim analyses after 1 year and every 6 months thereafter. The logrank statistic is used to test $H_0 : \beta = 0$ at each data monitoring time t_j $(j = 1, \ldots, 10)$ and the test is stopped at the smallest t_j such that

$$V_n(t_j) \geq 55, \quad \text{or} \quad V_n(t_j) \geq 11 \quad \text{and} \quad |S_n(t_j)|/V_n^{\frac{1}{2}}(t_j) \geq 2.85, \tag{8.5}$$

or at $t_{10}(= t^*)$ when (8.5) does not occur, where $V_n(t)$ is defined by (8.2). If the test stops with $V_n(t_j) \geq 55$ or at t^*, reject H_0 if $|S_n(t^*)|/V_n^{\frac{1}{2}}(t^*) \geq 2.05$. Also reject H_0 if the second event in (8.5) occurs for some $j < 10$. The lifetimes of the control group have an exponential distribution with mean 3 years and those of the treatment group have an exponential distribution with mean $3e^{-\beta}$ years, with $e^{\beta} = 1, \frac{2}{3}, \frac{1}{2}$.

8.1.3 Extensions to multivariate covariates and multiple endpoints

Lai & Li (2004) assumed univariate covariates when constructing confidence intervals for the treatment effect. The theory for constructing confidence intervals for the treatment effect given multivariate covariates is still lacking. In Section 8.3 we develop a method to construct confidence intervals for the treatment effect given general covariates using Wilk's statistic and ordering with partial likelihoods. Section 8.4 then generalizes our method to construct confidence regions for multiple endpoints. Specifically, confidence regions for the treatment effect (primary endpoint) and median survival (secondary endpoint) are constructed. As noted by Lai & Li (2004), their method is computationally intensive, and it takes about 8 hours of 3.2GHz Pentium 4 CPU time to generate a table like Table 8.1. In Section 8.2 we develop importance resampling techniques to reduce substantially the computing time for both Lai & Li's (2004) method and our likelihood-based method.

8.2 Monte Carlo computation of confidence intervals

8.2.1 Importance resampling

Lai & Li (2004) used a straightforward Monte Carlo implementation to compute $\widehat{p}(\beta)$ defined in (8.4) as follows. The observed entry times T_i and covariates z_i are taken as fixed constants in $\widehat{p}(\beta)$, for which only the survival times Y_i^* and censoring times ξ_i^* need to be generated. Since $\widehat{G}$ (or $\widehat{C}$) can only be estimated up to the longest observed survival (or censoring) time, denoted by t' (or t''), only $Y_i^* \wedge t'$ and $\xi_i^* \wedge t''$ can be generated. However, this suffices for the time-sequential score statistic (8.1) and its estimated null variance (8.2) for $t \leq \tau$. To generate $Y_i^* \wedge t'$, note that if U is uniformly distributed on $[0,1]$, then $(1-\widehat{G})^{-1}(\max\{U^{\exp(-\beta z_i)},\ 1-\widehat{G}(t')\})$ has the same distribution as $Y_i^* \wedge t'$.

The above Monte Carlo simulation procedure is computationally intensive. The reason is that a large number of simulations are needed to compute $\widehat{p}(\beta)$ for each β, and a sample of survival time needs to be generated for each simulation. We propose an importance sampling technique to reduce substantially the variance of the Monte Carlo estimate of the probability and thus the number of simulations required.

Specifically, when computing

$$\widehat{p}(\beta) = \mathrm{pr}\{(\tau^{(\beta)}, \Psi^{(\beta)}_{\tau^{(\beta)}}) > (\tau, \Psi_\tau)_{\mathrm{obs}}\},$$

the Monte Carlo method computes the average of N_1 realizations of

$$I\big((\tau^{(\beta)}, \Psi^{(\beta)}_{\tau^{(\beta)}}) > (\tau, \Psi_\tau)_{\mathrm{obs}}\big),$$

where $I(\cdot)$ is the indicator function. Our importance sampling method computes the average of N_2 realizations of

$$I\big((\tau^{(\widehat{\beta})}, \Psi^{(\widehat{\beta})}_{\tau^{(\widehat{\beta})}}) > (\tau, \Psi_\tau)_{\mathrm{obs}}\big)\frac{L(\beta)}{L(\widehat{\beta})},$$

where $\hat{\beta}$ is Cox's (1972) estimate of β that maximizes the partial likelihood at time τ and $L(\cdot)$ is the full likelihood at time τ; see Siegmund (1985, p. 122). This importance sampling technique reduces the variance of the Monte Carlo estimate of the probability. As a result, N_2 can be much smaller than N_1 and computing time is greatly reduced. Another important advantage to the importance sampling method is that it is a one-pass algorithm. Instead of generating data for each β in the Monte Carlo case, we only need to generate data once under $\widehat{\beta}$. Since every β is tilted to $\widehat{\beta}$, we can do resampling from the data set generated under $\widehat{\beta}$ for each β, and it greatly reduces the computing time required. $\widehat{\beta}$ is a good choice for tilting because $\Pr\{(\tau^{(\widehat{\beta})}, \Psi^{(\widehat{\beta})}_{\tau^{(\widehat{\beta})}}) > (\tau, \Psi_\tau)_{\mathrm{obs}}\}$ is around $\frac{1}{2}$.

8.2.2 Treating censoring indicators as ancillary

When the control group ($z_i = 0$) and the treatment group ($z_i = 1$) have different censoring distributions C_1 and C_2, a straightforward extension is to use separate Kaplan-Meier estimators $\widehat{C}_1$ and $\widehat{C}_2$. An alternative approach is to treat the censoring variable ξ_i as ancillary like z_i, thereby allowing possible dependence between z_i and ξ_i. For Monte Carlo simulation of $\hat{p}(\beta)$ in (8.4), let $\widehat{G}_i$ denote the distribution whose cumulative hazard function is $e^{\beta z_i}\hat{\Lambda}$, and $\widehat{C}$ denote the Kaplan-Meier estimator of the combined data. If Y_i is uncensored, then ξ_i is censored by Y_i, and we generate $\xi_i^* \geq Y_i$ from $\widehat{C}$ by rejection sampling and then generate $Y_i^* \leq \xi_i^*$ from $\widehat{G}_i$ by rejection sampling. If Y_i is censored, there is no need to generate Y_i^*. The above simulation method preserves the censoring indicators.

8.2.3 Computation of the confidence limits

We can use the method of successive secant approximations to find the limits of the confidence intervals. To find the lower limit, we define

$$f(\beta) = \widehat{p}(\beta) - \alpha$$

and solve the equation $f(\beta) = 0$ as in section 7.4.1. The same induction process can be applied to

$$f(\beta) = \widehat{p}(\beta) - (1 - \alpha)$$

to find the upper limit of the confidence interval. Chuang & Lai (2000) used a similar approach to compute limits of confidence intervals.

8.3 Multivariate covariates and ordering with partial likelihoods

Methods for constructing confidence intervals for the treatment effect given univariate covariates have been developed by Lai & Li (2004). The theory for constructing confidence intervals for the treatment effect given general covariates is still lacking. Whereas it is hard to generalize the methodology from the univariate covariates case to the general covariates case when Cox's score statistic is used, we use Wilks statistic and ordering with partial likelihoods to construct confidence intervals for the treatment effect given general covariates.

For univariate covariates, the log partial likelihood at calendar time t is

$$l_t(\beta) = \sum_{i=1}^{n} \delta_i(t) \left\{ \beta z_i - \log \left(\sum_{j \in R_i(t)} e^{\beta z_j} \right) \right\}.$$

Letting $\Psi_t = \sqrt{l_t(\hat{\beta}) - l_t(\beta)}$, where $\hat{\beta}$ is Cox's (1972) estimate of β that maximizes the partial likelihood at time t, we order the sample space of (τ, Ψ_τ) by

$$(\tau_1, \Psi^{(1)}_{\tau_1}) \le (\tau_2, \Psi^{(2)}_{\tau_2}) \quad \text{if and only if} \quad \Psi^{(1)}_{\tau_1 \wedge \tau_2} \le \Psi^{(2)}_{\tau_1 \wedge \tau_2}.$$

In the normal mean case,

$$\Psi_t = \frac{S_t - \mu t}{\sqrt{t}},$$

which is equivalent to Siegmund's ordering because

$$\Psi^{(1)}_{\tau_1 \wedge \tau_2} \le \Psi^{(2)}_{\tau_1 \wedge \tau_2} \Leftrightarrow S^{(1)}_{\tau_1 \wedge \tau_2} \le S^{(2)}_{\tau_1 \wedge \tau_2}.$$

In the case of general covariates, suppose $\beta = (\beta_1, \ldots, \beta_K)^T$ and β_1 corresponds to the treatment effect, which is the primary endpoint. Defining

$$l_t(\beta_1) = \sup_{\beta_2, \ldots, \beta_K} \sum_{i=1}^{n} \delta_i(t) \left\{ \beta^T z_i - \log \left(\sum_{j \in R_i(t)} e^{\beta^T z_j} \right) \right\}$$

and letting $\Psi_t = \sqrt{l_t(\hat{\beta}_1) - l_t(\beta_1)}$, we can use the same ordering with partial likelihoods. The importance sampling method developed in Section 8.2.1, the resampling method developed in Section 8.2.2 (which keeps the censoring variables as ancillary), and the method to compute endpoints of confidence intervals developed in Section 8.2.3, can all be incorporated to speed up the computations. By doing Edgeworth expansions, we can show that our method is first-order accurate, which is as good as the method proposed by Lai & Li (2004) but with the advantage that our method can be generalized to the multivariate covariates case naturally.

The following hypothetical clinical trial example shows that our likelihood-based method gives accurate coverage probabilities.

Example 8.2. Consider a time-sequential trial in which $n = 350$ subjects enter the trial uniformly during a 3-year recruitment period and are randomized to treatment or control with probability $\frac{1}{2}$. The trial is designed to last for a maximum of $t^* = 5.5$ years, with interim analyses after 1 year and every 6 months thereafter. The Wilks statistic is used to test $H_0 : \beta = 0$ at each data monitoring time t_j $(j = 1, \ldots, 10)$ and the test is stopped at the smallest t_j such that

$$V_n(t_j) \ge 55, \quad \text{or} \quad V_n(t_j) \ge 11 \quad \text{and} \quad \sqrt{l_t(\hat{\beta}) - l_t(0)} \ge 2.85, \tag{8.6}$$

or at $t_{10} (= t^*)$ when (8.6) does not occur, where $V_n(t)$ is defined by (8.2). If the test stops with $V_n(t_j) \ge 55$ or at t^*, reject H_0 if $\sqrt{l_{t^*}(\hat{\beta}) - l_{t^*}(0)} \ge 2.05$. Also reject H_0 if the second event in (8.6) occurs for some $j < 10$. For the control group, each year there is a 7% chance of being lost

Table 8.2: *Example* 8.2: Coverage errors in % for lower (L) and upper (U) confidence limits and coverage probabilities (P) of confidence intervals for β using Wilk's statistic with ordering with partial likelihoods. Methods: H, hybrid resampling with different Kaplan-Meier estimators $\widehat{C}_1$ and $\widehat{C}_2$; H_a, bybrid resampling treating ξ_i as ancillary.

	$\beta = 0$			$\beta = \log \frac{2}{3}$			$\beta = \log \frac{1}{2}$		
Method	L	U	P	L	U	P	L	U	P
H	4.85	4.65	90.50	5.10	5.70	89.20	5.15	3.95	90.90
H_a	4.55	4.60	90.85	4.95	5.95	89.10	5.10	4.05	90.85

to follow-up. For the treatment group, there is a 12% chance of loss to follow-up during the first year, 8% during the second year, and 6% per year starting from the third year. The lifetimes of the control group have an exponential distribution with mean 3 years, and those of the treatment group have an exponential distribution with mean $3e^{-\beta}$ years, with $e^{\beta} = 1, \frac{2}{3}, \frac{1}{2}$.

Table 8.2 shows that hybrid resampling in conjunction with ordering with partial likelihoods give accurate coverage probabilities within 1% of their nominal values.

8.4 The β-blocker heart attack trial

Time-sequential clinical trials received much attention from the biomedical community following early termination of the *Beta-Blocker Heart Attack Trial* (BHAT) in the early 1980s. The primary objective of BHAT was to determine whether regular, chronic administration of propranolol, a beta-blocker, to patients who had at least one documented *myocardial infarction* (MI) would result in significant reduction in mortality from all causes during the follow-up period. It was designed as a multicenter, double-blind, randomized placebo-controlled trial with a projected total of 4200 eligible patients recruited within 21 days of the onset of hospitalization for MI. The trial was planned to last 4 years, beginning in June 1978 and ending in June 1982, with patient accrual completed within the first 2 years so that all patients could be followed for a period of 2 to 4 years. The sample size calculation was based on a 3-year mortality rate of 18% in the placebo group and a 28% reduction of this rate in the treatment group, with a significance level of 0.05 and 0.9 power using a two-sided logrank test. In addition, periodic reviews of the data were planned to be conducted by a Data and Safety Monitoring Board roughly once every 6 months beginning at the end of the first year, whose functions were to monitor safety and adverse events and to advise the Steering and Executive Committees on policy issues related to the progress of the trial.

The actual recruitment period was 27 months, within which 3837 patients were accrued from 136 coronary care units in 31 clinical centers, with 1916 patients randomized into the propranolol group and 1921 into the placebo group. Although the recruitment goal of 4200 patients had not been met, the projected power was only slightly reduced to 0.89, as accrual was approximately uniform during the recruitment period.

The Data and Safety Monitoring Board arranged meetings at 11, 16, 21, 28, 34, and 40 months to review the data collected so far, before the scheduled end of the trial at 48 months. Besides monitoring safety and adverse events, the board also examined the standardized logrank statistics to examine if propranolol was indeed effective. Instead of continuing the trial to its scheduled 48 months, the Data and Safety Monitoring Board recommended terminating it in their last meeting because of conclusive evidence in favor of propranolol. Their recommendation was adopted and the trial was terminated on October 2, 1980.

We demonstrate the accuracy of our method by constructing confidence intervals for the β-blocker heart attack trial.

Example 8.3: Confidence intervals for the β-blocker heart attack trial. Applying our hybrid resampling method in conjunction with ordering with partial likelihoods, we found the 90% confidence interval for the treatment effect to be $-0.51 \le \beta \le -0.09$, and the 80% confidence interval to be $-0.43 \le \beta \le -0.12$. Lai & Li (2004) found the 90% confidence interval for the treatment effect to be $-0.50 \le \beta \le -0.08$, and the 80% confidence interval to be $-0.43 \le \beta \le -0.12$, which are quite close to our results. Using the Wiener process approximation to time-sequential logrank statistics under the proportional hazards model together with his own ordering scheme for normal data, Siegmund (1985, p. 134) found the 80% confidence interval to be $-0.42 \le \beta \le -0.11$, which is also close to ours. Siegmund considered $-\beta$ and assumed the Pocock-Haybittle stopping boundary instead of the O'Brien-Fleming stopping boundary. He also noted the asymmetry of the confidence interval about $\widehat{\beta} = -0.32$, in contrast with the naive 80% confidence interval $-0.32 \pm 0.14 = [-0.46, -0.18]$.

8.5 Bivariate confidence regions

It has been a long-standing problem how confidence regions can be constructed for multiple endpoints following group sequential clinical trials. We review in Section 8.5.1 Chuang & Lai's (2000) method for constructing bivariate confidence regions following group sequential tests for two population means, and generalize it in Section 8.5.2 to construct bivariate confidence regions for treatment effect and median survival following group sequential tests when treatment effect is the primary endpoint and median survival is the secondary endpoint.

8.5.1 Review of Chuang & Lai (2000)

We review in this section Chuang & Lai's (2000) hybrid resampling method for constructing bivariate confidence regions for two population means.

Let $X_1, X_2, \ldots$ be i.i.d. random variables with unknown mean θ, and suppose the stopping rule τ of a group sequential test depends on the sample sum $S_n = \sum_{i=1}^{n} X_i$ up to the stopping time. Suppose that one is also interested in estimating the common mean μ of i.i.d. random

variables $Y_1, Y_2, \ldots$ that are observed up to the stopping time τ. Let $\bar{Y}_n = n^{-1}\sum_{i=1}^n Y_i$. Although $\sqrt{n}(\bar{Y}_n - \mu)$ is an asymptotic pivot having a limiting standard normal distribution, $\sqrt{\tau}(\bar{Y}_\tau - \mu)$ is no longer an asymptotic pivot because its limiting distribution depends on θ, which determines the distribution of τ.

First, suppose that (X_i, Y_i) is bivariate normal with known correlation coefficient ρ, and X_i and Y_i have common known variance 1. Let V denote the covariance matrix of (X_i, Y_i), and let $\mathbf{X} = (X_1, \ldots, X_\tau; Y_1, \ldots, Y_\tau; \tau)$. An exact $1-2\alpha$ confidence region for (θ, μ) is

$$\{(\theta, \mu) : R(\mathbf{X}, \theta, \mu) \le u_{1-2\alpha}(\theta)\},$$

where $R(\mathbf{X}, \theta, \mu) = \tau(\bar{X}_\tau - \theta, \bar{Y}_\tau - \mu)V^{-1}(\bar{X}_\tau - \theta, \bar{Y}_\tau - \mu)^T$, and $u_{1-2\alpha}(\theta)$ is the $(1-2\alpha)$th quantile of $R(\mathbf{X}, \theta, \mu)$.

Without assuming (X_i, Y_i) to be standard normal and ρ to be known, we can replace ρ by the sample correlation $\hat{\rho}_\tau$. Letting $\hat{V}$ denote the matrix with 1 on the diagonal and $\hat{\rho}_\tau$ elsewhere, we have

$$\hat{R}(\mathbf{X}, \theta, \mu) = \tau(\bar{X}_\tau - \theta, \bar{Y}_\tau - \mu)\hat{V}^{-1}(\bar{X}_\tau - \theta, \bar{Y}_\tau - \mu)^T. \tag{8.7}$$

Let $\hat{G}$ be the empirical distribution of $((X_i - \bar{X}_\tau)/\hat{\sigma}_{x,\tau}, (Y_i - \bar{Y}_\tau)/\hat{\sigma}_{y,\tau})$, where $\hat{\sigma}_{x,\tau}$ and $\hat{\sigma}_{y,\tau}$ are the sample variances. Let $(\epsilon_1, \eta_1), \ldots$ be i.i.d. with common distribution $\hat{G}$ and let $X_i(\theta) = \theta + \epsilon_i$. Let $\tau(\theta)$ be the stopping rule applied to $X_1(\theta), \ldots$. Using $\tilde{\rho}_{\tau(\theta)}$ to denote the sample correlation coefficient of the (ϵ_i, η_i), $1 \le i \le \tau(\theta)$, we let $\hat{V}_{\tau(\theta)}$ denote the matrix with 1 on the diagonal and $\tilde{\rho}_{\tau(\theta)}$ elsewhere. Defining $\hat{u}_{1-2\alpha}(\theta)$ as the $(1-2\alpha)$th quantile of

$$(\tau(\theta))^{-1}\Big(\sum_{i=1}^{\tau(\theta)}\epsilon_i, \sum_{i=1}^{\tau(\theta)}\eta_i\Big)\hat{V}_{\tau(\theta)}^{-1}\Big(\sum_{i=1}^{\tau(\theta)}\epsilon_i, \sum_{i=1}^{\tau(\theta)}\eta_i\Big)^T, \tag{8.8}$$

we obtain the hybrid confidence region for (μ, θ) with nominal coverage error 2α as

$$\{(\theta, \mu) : \hat{R}(\mathbf{X}, \theta, \mu) \le \hat{u}_{1-2\alpha}(\theta)\}.$$

Without assuming unit variance of Y_i, we can replace $\hat{V}$ and $\hat{V}_{\tau(\theta)}$ by

$$\tilde{V} = \begin{pmatrix} 1 & \hat{\rho}_\tau\hat{\sigma}_{y,\tau} \\ \hat{\rho}_\tau\hat{\sigma}_{y,\tau} & \hat{\sigma}_{y,\tau}^2 \end{pmatrix}, \qquad \tilde{V}_{\tau(\theta)} = \begin{pmatrix} 1 & \hat{\rho}_{\tau(\theta)}\hat{\sigma}_{y,\tau(\theta)} \\ \hat{\rho}_{\tau(\theta)}\hat{\sigma}_{y,\tau(\theta)} & \hat{\sigma}_{y,\tau(\theta)}^2 \end{pmatrix},$$

where $\tilde{\sigma}_{\eta,m}^2 = m^{-1}\sum_{i=1}^m(\eta_i - \bar{\eta}_m)^2$. Let $\tilde{u}_{1-2\alpha}(\theta)$ be the $(1-2\alpha)$th quantile of (8.8) with $\hat{V}_{\tau(\theta)}$ replaced by $\tilde{V}_{\tau(\theta)}$. The hybrid confidence region is

$$\{(\theta, \mu) : \tilde{R}(\mathbf{X}, \theta, \mu) \le \tilde{u}_{1-2\alpha}(\theta)\},$$

where $\tilde{R}(\mathbf{X}, \theta, \mu)$ is defined by (8.7) with $\hat{V}$ replaced by $\tilde{V}$. Chuang & Lai (2000, p. 19) gave an algorithm to compute the hybrid confidence regions explicitly.

8.5.2 Confidence regions for treatment effect and median survival

Motivated by Chuang & Lai (2000) and our methods for constructing confidence intervals for treatment effect and for median survival, we can combine these two methods for constructing confidence intervals to develop a method for constructing bivariate confidence regions for treatment effect and median survival following group sequential trials, where treatment effect is the primary endpoint and median survival is the secondary endpoint.

For Chuang & Lai's (2000) method for constructing bivariate confidence regions for two population means, taking the root $R(\mathbf{X}, \theta, \mu) = \tau(\bar{X}_\tau - \theta, \bar{Y}_\tau - \mu)V^{-1}(\bar{X}_\tau - \theta, \bar{Y}_\tau - \mu)^T$ is straightforward. In the case of constructing confidence regions for treatment effect and median survival, the root needs to be chosen carefully.

Motivated by standard multiple hypotheses testing approaches, we can take the maximum of the two self-normalized statistics

$$l_\tau(\hat{\beta}) - l_\tau(\beta), \qquad \left(\frac{\hat{S}_z(m) - \frac{1}{2}}{\hat{\xi}_z(m)} \right)^2$$

to form the new statistic U_τ that will be used to order the sample space:

$$U_\tau = \max \left\{ l_\tau(\hat{\beta}) - l_\tau(\beta), \ \left(\frac{\hat{S}_z(m) - \frac{1}{2}}{\hat{\xi}_z(m)} \right)^2 \right\}.$$

A $1 - 2\alpha$ confidence region is the set of all parameters (β, m) not rejected by

$$\{(\beta, m) : \alpha < \hat{P}_\beta\{(\tau, U_\tau) > (\tau, U_\tau)_{obs}\} < 1 - \alpha\},$$

where

$$(T, U_T) \geq (t, u_t) \quad \text{if and only if} \quad U_{T \wedge t} \geq u_{T \wedge t}.$$

Similar to the case where confidence intervals for treatment effect β are constructed, $\hat{P}_\beta\{(\tau, U_\tau) > (\tau, U_\tau)_{obs}\}$ needs to be calculated via simulations. Again, Breslow's (1974) estimator of the cumulative hazard function from all the data at the end of the trial is used for simulations.

Chuang & Lai's (2000) algorithm for computing the confidence regions for two population means explicitly can be incorporated in our method to compute the confidence regions explicitly for treatment effect and median survival, and the importance sampling techniques developed in Section 8.2 can be incorporated to speed up the simulations and thus reduce the computing time substantially.

The following hypothetical clinical trial example shows that our method gives accurate coverage probabilities.

Example 8.4. Consider a time-sequential trial in which $n = 350$ subjects enter the trial uniformly during a 3-year recruitment period and are randomized to treatment or control with probability $\frac{1}{2}$. The trial is designed to last for a maximum of $t^* = 5.5$ years, with interim analyses after 1 year and every 6 months thereafter. The Wilks statistic is used to test $H_0 : \beta = 0$ at each data monitoring time t_j $(j = 1, \ldots, 10)$, and the test is stopped at the smallest t_j such that

$$V_n(t_j) \geq 55, \quad \text{or} \quad V_n(t_j) \geq 11 \quad \text{and} \quad \sqrt{l_t(\hat{\beta}) - l_t(0)} \geq 2.85, \tag{8.9}$$

or at $t_{10}(= t^*)$ when (8.9) does not occur, where $V_n(t)$ is defined by (8.2). If the test stops with $V_n(t_j) \geq 55$ or at t^*, reject H_0 if $\sqrt{l_{t^*}(\hat{\beta}) - l_{t^*}(0)} \geq 2.05$. Also reject H_0 if the second event in (8.9) occurs for some $j < 10$. The distribution for loss to follow-up is exponential with a median of 12 years. The lifetimes of the control group have an exponential distribution with mean 3 years, and those of the treatment group have an exponential distribution with mean $3e^{-\beta}$ years, with $e^{\beta} = 1, \frac{2}{3}, \frac{1}{2}$.

Table 8.3 shows that our method for constructing bivariate confidence regions gives accurate coverage probabilities that are within 1.2% of their nominal values.

Table 8.3: *Example* 8.4: Coverage errors in % for lower (L) and upper (U) confidence limits and coverage probabilities (P) of confidence regions for treatment effect as the primary endpoint and median survival as the secondary endpoint. Two bivariate confidence regions are constructed: the confidence region for treatment effect and the median survival of the control group, and the confidence region for treatment effect and the median survival of the treatment group.

	$\beta = 0$			$\beta = \log \frac{2}{3}$			$\beta = \log \frac{1}{2}$		
Group	L	U	P	L	U	P	L	U	P
Control	4.25	4.90	90.85	3.95	4.90	91.15	4.10	4.70	91.20
Treatment	4.55	4.80	90.65	4.50	4.90	90.60	5.05	4.00	90.95

Chapter 9

Concluding remarks

9.1 Summary of methods and results

Monte Carlo simulations are often used in the design stage to compute power and sample size in clinical trials. A clear disadvantage of Monte Carlo simulations is that they are computationally intensive. Importance resampling techniques are developed in Section 7.2 to compute tail probabilities, in order to reduce the number of simulations required to compute power and sample size in the design of fixed-duration clinical trials. They can be incorporated in the simulation program of Halpern and Brown (1987) to reduce computing time substantially. A test-based method for constructing confidence intervals for median survival in a proportional hazards model via bootstrap is given in Section 7.2.4 and shown to give coverage probabilities close to the nominal values.

Lai & Li (2004) developed a hybrid resampling method to construct test-based confidence intervals for treatment effect given univariate covariates. Importance sampling techniques are developed in Section 8.2 to speed up Lai & Li's (2004) computations of the confidence intervals substantially. Section 8.3 gives a method for constructing test-based confidence intervals for treatment effect given multivariate covariates, which is shown to give coverage probabilities close to the nominal values. In Section 8.5, by combining the two test-based methods for constructing confidence intervals for treatment effect and median survival, we construct test-based confidence regions for treatment effect as the primary endpoint and median survival as the secondary endpoint.

9.2 Goal of the study

Clinical trial design and analysis software is crucial to the success of clinical trials. Our ultimate goal is to provide clinicians with powerful clinical trial design and analysis software to give them the flexibility to design and analyze fixed-duration and time-sequential clinical trials with failure-time endpoints. This software will have two modules: the design module and the analysis module.

Importance sampling and resampling techniques will be incorporated in both modules to speed up the Monte Carlo simulations and thus reduce computational cost. Software packages currently available serve as a benchmark for the development of our new software; they are reviewed in Section 9.2.1. The on-going work towards the realization of our goal is described in Section 9.2.2.

9.2.1 Review of available software

- Halpern and Brown (1987) developed a simulation program for the design of fixed-duration clinical trials using Monte Carlo simulations, which we have improved by using importance resampling techniques. The program allows arbitrary specifications of the null and alternative survival distributions for the logrank or Gehan test, which are assumed to be either piecewise exponential or cure-rate models. Halpern & Brown (1993) subsequently developed another simulation program for these test statistics incorporating group sequential testing using either the Pocock or the O'Brien-Fleming boundary.

- Gu & Lai (1999) developed a simulation program for the design of group sequential clinical trials. This broadened the scope of the Halpern-Brown programs in several major ways to provide the investigator with a flexible tool to plan clinical trials with failure-time endpoints and interim analyses. First, instead of the Pocock and O'Brien-Fleming boundaries, the program provides four options to the user for choosing a stopping boundary: Slud-Wei, Lan-DeMets, Haybittle-type boundaries, plus any other boundary specified by the user. Second, the program allows the user to incorporate withdrawal and noncompliance by specifying the censoring distributions and crossover rates of the two treatment groups. Third, the program gives the user an additional way to specify the survival distribution of the new treatment by specifying its time-dependent hazard ratio relative to the baseline survival distribution, which is assumed to be piecewise constant. Finally, the program allows the user to choose any test statistic in the beta family of statistics, in which the logrank statistic and the Gehan statistic are two examples.

- EAST is commercial clinical trial design and analysis software; it is developed by Cytel company and can be purchased from the company's website. EAST has a design module and an analysis module. The design module calculates sample size and stopping boundaries for superiority, futility, and non-inferiority studies. The design module provides different choices to the user for choosing a stopping boundary: Wang-Tsiatis, Lan-Demets, Haybittle-Peto families, plus stopping boundaries derived from a series of published error-spending functions. EAST provides a user-friendly Excel interface and well-prepared documentation, which are advantages over the free programs of Halpern & Brown (1987; 1993) and Gu & Lai (1999). Unlike the free programs, EAST has an analysis module. However, the information that the analysis module can provide is limited. It gives point estimates of the regression parameter

at the end of the study, and gives confidence intervals for the treatment effect following a group sequential test that is based on Brownian motion approximations. We have shown in Chapter 8 that confidence intervals based on Brownian motion approximations give poor coverage probabilities. The analysis module does not provide confidence intervals for median survival.

- PEST is another commercial clinical trial design and analysis package; it is developed by MPS Research Unit and can be purchased from its website. The most recent version is PEST4, which also has both design and analysis modules. PEST4 provides a user-friendly interface based on SAS/AF, and has well-prepared documentation. PEST4 provides direct reading of SAS permanent datasets. However, the information that the analysis module can provide is also limited. Similar to EAST, PEST4 gives point estimates of the regression parameter at the end of the study, and gives confidence intervals for the treatment effect following a group sequential test based on Brownian motion approximations. The analysis module does not provide confidence intervals for median survival.

9.2.2 On-going work towards the goal

Since confidence intervals for treatment effect and median survival are important for clinicians to draw conclusions about a new treatment, we are currently developing an analysis module that can provide confidence intervals for treatment effect and median survival following group sequential tests. Bearing in mind the goal of providing clinicians with a powerful software package that has both design and analysis modules for fixed-duration and time-sequential trials, we are extending our importance sampling and resampling techniques for fixed-duration tests to group sequential tests so that the simulation time for both the design module and the analysis module of group sequential trials can also be reduced substantially. Although our current analysis module is already capable of giving confidence intervals for treatment effect and median survival based on the methodology we developed in Chapter 7 and Chapter 8, we are working on extensions of our methods to other (non-proportional hazards) survival models and other test statistics in order to make our analysis module more versatile.

Bibliography

Armitage, P., McPherson, C.K. & Rowe, B.C. (1969). Repeated significance tests on accumulating data. *J. Roy. Statist. Soc.* A, **132**, 235–244.

Beta-Blocker Heart Attack Trial Research Group (1982). A randomized trial of propranolol in patients with acute myocardial infarction. *J. Amer. Med. Assoc.*, **147**, 1707–1714.

Beta-Blocker Heart Attack Trial Research Group (1984). Beta-Blocker Heart Attack Trial: Design, methods and baseline results. *Contr. Clin. Trials*, **5**, 382–437.

Bhattacharya, R.N. & Ghosh, J.K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.*, **6**, 434–51.

Bilias, Y., Gu, M.G. & Ying, Z. (1997). Towards a general asymptotic theory for Cox model with staggered entry. *Ann. Statist.*, **25**, 662–82.

Brookmeyer, R. & Crowley, J. (1982). A confidence interval for the median survival time. *Biometrics*, **38**, 29–41.

Chang, M.N. (1989). Confidence intervals for a normal mean following a group sequential test. *Biometrics*, **45**, 247–54.

Chuang, C.S. & Lai, T.L. (1998). Resampling methods for confidence intervals in group sequential trials. *Biometrika*, **85**, 317–52.

Chuang, C.S. & Lai, T.L. (2000). Hybrid resampling methods for confidence intervals (with discussion). *Statistica Sinica*, **10**, 1–50.

Cox, D.R. (1982). Regression models and life-tables (with discussion). *J. Roy. Statist. Soc.* B, **34**, 187–220.

Davison, A.C. (1988). Discussion of papers by D.V. Hinkley and by T.J. DiCiccio and J.P. Romano. *J. Roy. Statist. Soc.* B, **50**, 356–357.

DeMets, D.L., Hardy, R., Freedman, L.M. & Lan, G.K.K. (1984). Statistical aspects of early termination in the Beta-Blocker Heart Attack Trial. *Contr. Clin. Trials*, **5**, 362–72.

Efron, B. (1981). Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika*, **68**, 589–599.

Efron, B. (1987). Better bootstrap confidence intervals (with discussion). *J. Am. Statist. Assoc.*, **82**, 171–200.

Emerson, S.S. & Fleming, T.R. (1990). Parameter estimation following group sequential hypothesis testing. *Biometrika*, **77**, 875–92.

Greenwood, M. (1926). The Natural Curation of Cancer. *Reports on Public Health and Medical Subjects*, **33**, 1–26.

Gu, M. (1992). On the Edgeworth expansion and bootstrap approximation for the Cox regression model under random censorship. *Can. J. Statist.*, **20**, 399–414.

Gu, M.G. & Lai, T.L. (1990). Functional laws of the iterated logarithm for product-limit estimator of a distribution function under random censorship or truncation. *Ann. Probab.*, **18**, 160–89.

Gu, M.G. & Lai, T.L. (1991). Weak convergence of time-sequential rank statistics with applications to sequential testing in clinical trials. *Ann. Statist.*, **19**, 1403–1433.

Gu, M.G. & Lai, T.L. (1998). Repeated significance testing with censored rank statistics in interim analysis of clinical trials. *Statist. Sinica*, **8**, 411–423.

Gu, M.G. & Lai, T.L. (1999). Determination of power and sample size in the design of clinical trials with failure-time endpoints and interim analyses. *Contr. Clin. Trials*, **20**, 423–438.

Gu, M.G. & Ying, Z. (1995). Group sequential methods for survival data using partial likelihood score processes with covariate adjustment. *Statist. Sinica*, **5**, 793–804.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansions*. New York: Springer-Verlag.

Hammerskey, J.M. & Handscomb, D.C. (1964). *Monte Carlo Methods*. Methuen, London.

Haybittle, J.L. (1971). Repeated assessments of results in clinical trials of cancer treatment. *Br. J. Radiol.*, **44**, 793–97.

Helmers, R. (1991). On the Edgeworth expansion and the bootstrap approximation for a Studentized U-statistic. *Ann. Statist.*, **19**, 470–84.

Jennison, C. & Turnbull, B.W. (1997). Group sequential analysis incorporating covariate information. *J. Amer. Statist. Assoc.*, **92**, 1330–1341.

Jennison, C. & Turnbull, B.W. (2000). *Group Sequential Methods with Applications to Clinical Trials*. New York: Chapman & Hall/CRC.

Johns, M.V. Jr (1988). Importance sampling for bootstrap confidence intervals. *J. Amer. Statist. Assoc.*, **83**, 709–714.

Jones, D. & Whitehead, J. (1979). Sequential forms of logrank and modified logrank tests for censored data. *Biometrika*, **66**, 105–113.

Klein, J.P. & Moeschberger, M.L. (2004). *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer-Verlag.

Lai, T.L. & Li, W.Z. (2004). Confidence intervals in group sequential trials with random group sizes and applications to survival analysis. *Technical Report, Statist. Dept., Stanford Univ.*.

Lai, T.L. & Wang, J.Q. (1993). Edgeworth expansions for symmetric statistics with applications to Bootstrap methods. *Statist. Sinica*, **3**, 517–542.

Lan, K.K.G. & DeMets, D.L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika*, **70**, 659–663.

Lan, K.K.G. & DeMets, D.L. (1989). Group sequential procedures: calendar versus information time. *Statist. Med.*, **8**, 1191–1198.

Lin, D.Y. (1992). Sequential logrank tests adjusting for covariates with accelerated life model. *Biometrika*, **79**, 523–529.

Li, W. (2001). Confidence Intervals following group sequential trials with random group sizes and applications to survival analysis. *PhD thesis, Dept. of Stat., Stanford Univ.*.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in their consideration. *Canc. Chemo. Rep.*, **94**, 510–521.

O'Brien, P.C. & Fleming, T.R. (1979). A multiple testing procedure for clinical trials. *Biometrika*, **35**, 549–556.

Pocock, S.J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika*, **64**, 191–99.

Prentice, R.L. (1978). Linear rank statistics with right censored data. *Biometrics*, **65**, 167–179.

Rosner, G.L. & Tsiatis, A.A. (1988). Exact confidence intervals following a group sequential trial: a comparison of methods. *Biometrika*, **75**, 723–29.

Scharfstein, D.O., Tsiatis, A.A. & Robins, J.M. (1997). Semiparametric efficiency and its implication on the design and analysis of group sequential studies. *J. Amer. Statist. Assoc.*, **92**, 1342–1350.

Scharfstein, D.O. & Tsiatis, A.A. (1998). The use of simulation and bootstrap in information-based group sequential studies. *Statist. Med.*, **17**, 75–87.

Sellke, T. & Siegmund, D. (1983). Sequential analysis of the proportional hazards model. *Biometrika*, **70**, 315–26.

Siegmund, D. (1978). Estimation following sequential tests. *Biometrika*, **65**, 341–49.

Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. New York: Springer-Verlag.

Slud, E.V. & Wei, L.J. (1982). Two-sample repeated significance tests based on the modified Wilcoxon statistic. *J. Amer. Statist. Assoc.*, **77**, 862–868.

Wallis, A.W. (1980). The Statistical Research Group. *J. Amer. Statist. Assoc.*, **75**, 320–334.

Tsiatis, A.A. (1982). Repeated significance testing for a general class of statistics used in censored survival analysis. *J. Amer. Statist. Assoc.*, **77**, 855–861.

Tsiatis, A.A., Rosner, G.L. & Tritchler, D.L. (1985). Group sequential tests with censored survival data adjusting for covariates. *Biometrika*, **72**, 365–373.