Approximate Solution of Large-Scale Linear Inverse Problems with Monte Carlo Simulation *

Nick Polydorides, Mengdi Wang, Dimitri P. Bertsekas[†]

November 19, 2009

Abstract

We consider the approximate solution of linear ill-posed inverse problems of high dimension with a simulation-based algorithm that approximates the solution within a low-dimensional subspace. The algorithm uses Tikhonov regularization, regression, and low-dimensional linear algebra calculations and storage. For sampling efficiency, we use variance reduction/importance sampling schemes, specially tailored to the structure of inverse problems. We demonstrate the implementation of our algorithm in a series of practical large-scale examples arising from Fredholm integral equations of the first kind.

1   Introduction

We consider linear inverse problems of the form

$$Ax = b, \tag{1}$$

where A is an $m \times n$ real matrix, b is a vector in $\Re^m$ and x is the unknown vector in $\Re^n$. Such problems typically arise from discretized Fredholm integral equations of the first kind, such as those encountered in image processing, geophysical prospecting, and astronomy [Gro07], [BB98]. The

system (1) may be either underdetermined or overdetermined. We consider a least-squares formulation

$$\min_{x \in \Re^n} \|Ax - b\|_\zeta^2, \tag{2}$$

where ζ is a known probability distribution vector with positive components. When n or m is very large, the exact optimal solution x^* of problem (2) becomes computationally formidable.

In this paper, we propose to approximate x^* within a low-dimensional subspace

$$S = \{\Phi r \mid r \in \Re^s\},$$

where Φ is an $n \times s$ matrix whose columns represent basis functions of S. Our methodology involves subspace approximation, Monte-Carlo simulation, regression, and most significantly, only low-dimensional vector operations (of order s, the number of basis functions).

We thus focus on the following approximation to problem (2):

$$\min_{r \in \Re^s} \|A\Phi r - b\|_\zeta^2.$$

The optimal solution is

$$r^* = G^{-1}c, \tag{3}$$

where

$$G = \Phi' A' Z A \Phi, \qquad c = \Phi' A' Z b, \tag{4}$$

Z is the $m \times m$ diagonal matrix with components of ζ along its diagonal (we assume throughout this paper that G is invertible).

Since the direct calculation of G and c may be prohibitively expensive, we propose to estimate their values by simulation, as suggested in [BY09]. For example, we may generate a sequence $\{(i_0, j_0, \bar{j}_0), \ldots, (i_t, j_t, \bar{j}_t)\}$ by sampling independently according to some distribution ξ from the set of index triples $(i, j, \bar{j}) \in \{1, \ldots, n\}^3$. Then we may estimate G and c with $\hat{G}$ and $\hat{c}$ given by

$$\hat{G} = \frac{1}{t+1} \sum_{k=0}^{t} \frac{\zeta_{i_k} a_{i_k j_k} a_{i_k \bar{j}_k}}{\xi_{i_k j_k \bar{j}_k}} \phi_{j_k} \phi'_{\bar{j}_k} \quad \hat{c} = \frac{1}{t+1} \sum_{k=0}^{t} \frac{\zeta_{i_k} a_{i_k j_k} b_{i_k}}{\xi_{i_k j_k}} \phi_{j_k}, \tag{5}$$

where we denote by a_{ij} the (i,j)th component of A, ξ_{ij} the marginal probability of (i,j), $\xi_{ij\bar{j}}$ the marginal probability of $(i,j,\bar{j})$, and ϕ'_j the jth row of Φ. One natural approximation of r^* is $\hat{r} = \hat{G}^{-1}\hat{c}$. When $t \to \infty$, as shown

in [BY09], we have $\hat{G} \to G$ and $\hat{c} \to c$ with probability 1, so $\hat{G}^{-1}\hat{c} \to r^*$ with probability 1.

There have been several proposals in the literature relating to the exact or approximate solution of large-scale inverse problems. One of the earliest attempts is by Lanczos [Lan58], which successively approximates the solution without explicit matrix inversion. Since then a number of iterative methods have been studied, such as the Landweber iteration, the conjugate gradient method, the LSQR algorithm (see the survey [HH93] for a comprehensive review of these methods). A projection-regularization approach, proposed by O'Leary and Simmons [OS81], approximates the solution within a subspace in which the projected coefficient matrix is bidiagonalizable. A related approach proposed by Calveti and Zhang [DCZ99] suggests the use of Lanczos bidiagonalization with Gauss quadrature. Later, a trust-region formulation was proposed by Rojas and Sorensen [RS02], which poses the regularized problem as an inequality constrained least-squares problem.

Our work differs from those mentioned above in that it involves both subspace approximation and simulation, and relies exclusively on low-dimensional vector operations. The origins of our approach can be traced to projected equation methods for approximate dynamic programming, which aim to solve forms of Bellman's equation of very large dimension by using simulation (see the books by Bertsekas and Tsitsiklis [BT96], Sutton and Barto [SB98], and Bertsekas [Ber07]). These methods were recently extended to apply to general square systems of linear equations and regression problems in the paper by Bertsekas and Yu [BY09], which was the starting point for the present paper. The companion paper [WPB09] emphasizes generic methodological aspects of regression and variance analysis for importance sampling schemes, and may serve as a theoretical basis for the present work, which emphasizes algorithmic approaches for the solution of practical inverse problems.

The paper is organized as follows. In Section 2, we present general aspects of subspace approximation, regression, and our simulation framework. In Section 3, we discuss alternative methods for designing importance sampling distributions for simulation, in the context of our algorithmic methodology. In Section 4, we apply our methodology to a number of practical inverse problems of large dimension, and we present the computational results.

2 Approximation Methodology Based on Simulation and Regression

2.1 Simulation Framework

We want to estimate the matrix G and vector c of Eq. (4), which define the optimal low-dimensional solution r^* [cf. Eq. (3)]. Equation (5) provides one such approach. In this section we will present a few alternative approaches.

One possibility, proposed in [WPB09], is to estimate each component of G and c using a separate sequence $\left\{(i_k, j_k, \bar{j}_k)\right\}_{k=0}^{t}$. Then we may estimate the (ℓ, q)th component of G or the ℓth component of c with

$$\hat{G}_{\ell q} = \frac{1}{t+1} \sum_{k=0}^{t} \frac{\zeta_{i_k} a_{i_k j_k} a_{i_k \bar{j}_k}}{\xi_{i_k j_k \bar{j}_k}} \phi_{j_k \ell} \, \phi'_{\bar{j}_k q}, \quad \hat{c}_{\ell} = \frac{1}{t+1} \sum_{k=0}^{t} \frac{\zeta_{i_k} a_{i_k j_k}}{\xi_{i_k j_k}} b_{i_k} \, \phi_{j_k \ell}, \ (6)$$

where we denote by $\phi_{j\ell}$ the (j, ℓ)th component of Φ. This component-by-component approach requires $(s^2 + 3s)/2$ separate sample sequences (since G is symmetric), which increases the time complexity of the computation. Nonetheless, as we will discuss later, this allows the customization of the sampling distribution to the particular component, according to principles of importance sampling, so fewer samples per component may be required for the same solution accuracy. Another possibility is to generate one sequence per column or row of G and one sequence for c, which requires $s+1$ separate sample sequences.

More generally, we may partition the set of components of G and c, and then generate one sample sequence per block. With a judicious partitioning strategy, the potential advantage of this strategy is twofold: first, grouping together components that can be estimated using similar distributions so as to improve the efficiency of the sampling process, and second, estimating "almost independent" components independently so as to reduce the bias induced by correlation among the components of the estimates.

We now briefly discuss alternative mechanisms to generate sample triples $(i, j, \bar{j})$. The simplest scheme is to sample i_k, j_k, and $\bar{j}_k$ independently from one another, according to distributions μ_1, μ_2, and μ_3, respectively. Then the marginal probabilities for pairs (i_k, j_k) and triples $(i_k, j_k, \bar{j}_k)$ are

$$\xi_{i_k j_k} = \mu_{1, i_k} \mu_{2, j_k}, \qquad \xi_{i_k j_k \bar{j}_k} = \mu_{1, i_k} \mu_{2, j_k} \mu_{3, \bar{j}_k}.$$

An alternative is to generate an independent sequence of indices $\left\{i_0, i_1, \ldots\right\}$ according to a distribution μ, and then generate j_k and $\bar{j}_k$ conditioned on

each i_k, according to transition probabilities $q_{i_k j_k}$ and $\tilde{q}_{i_k \bar{j}_k}$. In this case, the marginal probabilities are

$$\xi_{i_k j_k} = \mu_{i_k} q_{i_k j_k}, \qquad \xi_{i_k j_k \bar{j}_k} = \mu_{i_k} q_{i_k j_k} \tilde{q}_{i_k \bar{j}_k}.$$

A somewhat more complex scheme is to generate a sequence of state transitions $\{i_0, i_1, \ldots\}$ using an irreducible Markov chain with transition probability matrix P and initial distribution ξ_0. Sampling j_k and $\bar{j}_k$ according to some transition probabilities $q_{i_k j_k}$ and $\tilde{q}_{i_k \bar{j}_k}$ yields marginal probabilities for pairs (i_k, j_k) and triples $(i_k, j_k, \bar{j}_k)$:

$$\xi_{i_k j_k} = (\xi_0' P^k)_{i_k} q_{i_k j_k}, \qquad\qquad \xi_{i_k j_k \bar{j}_k} = (\xi_0' P^k)_{i_k} q_{i_k j_k} \tilde{q}_{i_k \bar{j}_k}.$$

Here the choice of P should ensure that all row indices are sampled infinitely often, so that $\hat{G} \to G$ and $\hat{c} \to c$ (and hence also $\hat{G}^{-1}\hat{c} \to r^*$) as $t \to \infty$, with probability 1. In particular, if P is an irreducible Markov chain, we can use as ξ the distribution of long-term frequencies of state visits corresponding to P.

2.2   Regression Methods

Given $\hat{G}$ and $\hat{c}$, we may estimate r^* [cf. Eq. (3)] with $\hat{G}^{-1}\hat{c}$, but this estimate may be highly susceptible to simulation noise, particularly if G is nearly singular. As a more reliable alternative, we consider the estimation of r^* using a form of regression and the model

$$\hat{c} = \hat{G}r + e,$$

where e is the vector representing simulation error,

$$e = (G - \hat{G})r + \hat{c} - c.$$

The standard least squares/regression/Tikhonov regularization approach yields the estimate

$$\hat{r} = \arg\min_{r \in \Re^s} \left\{ (\hat{G}r - \hat{c})'\Sigma^{-1}(\hat{G}r - \hat{c}) + (r - \bar{r})'\Gamma^{-1}(r - \bar{r}) \right\},$$

where $\bar{r}$ is an *a priori* estimate (for example $\bar{r} = \hat{G}^{-1}\hat{c}$ or a singular valued-based estimate of $\hat{G}^{-1}\hat{c}$), and Σ and Γ are some positive definite symmetric matrices. Equivalently,

$$\hat{r} = (\hat{G}'\Sigma^{-1}\hat{G} + \Gamma^{-1})^{-1}(\hat{G}'\Sigma^{-1}\hat{c} + \Gamma^{-1}\bar{r}). \tag{7}$$

An effective choice is to use as Σ an estimate of the covariance of the error $e = \hat{c} - \hat{G}r^*$. Such an estimate can be obtained from the simulation using a nominal guess $\tilde{r}$ of r^*, i.e., the matrix

$$\Sigma(\tilde{r}) = \frac{1}{t+1}\sum_{k=0}^{t} e_k e_k' = \frac{1}{t+1}\sum_{k=0}^{t}\left((G_k-\hat{G})\tilde{r}+(\hat{c}-c_k)\right)\left((G_k-\hat{G})\tilde{r}+(\hat{c}-c_k)\right)',$$

(8)

where each e_k can be viewed as a sample of e, and G_k and c_k are corresponding sample terms that are averaged to form the estimates $\hat{G}$ and $\hat{c}$. We refer to the companion paper [WPB09] for further discussion and a derivation of an associated confidence interval for the estimate $\hat{r}$ of Eq. (7).

For the experimental results reported in this paper, we have used the preceding regression procedure with $\Sigma(\tilde{r})$ as given above, and with a nominal guess $\tilde{r}$ based on $\bar{r}$. Another possibility is to use a form of iterative regression, whereby we estimate r^* by repeatedly using Eq. (7) with intermediate correction of the matrix Σ using Eq. (8), which is the iteration

$$r_{k+1} = \left(\hat{G}'\Sigma(r_k)^{-1}\hat{G}+\Gamma^{-1}\right)^{-1}\left(\hat{G}'\Sigma(r_k)^{-1}\hat{c}+\Gamma^{-1}\bar{r}\right). \qquad (9)$$

This iteration has been shown to converge locally [WPB09] to a fixed point of Eq. (9), provided that the covariance of e is sufficiently small.

Under certain circumstances, we may have prior knowledge about the high-dimensional solution x^* of problem (2), which may suggest a natural type of regression. For example, in some inverse problems arising in physical science, it is desired that the regularization term be proportional to $(x - \bar{x})'L'L(x - \bar{x})$ for some $l \times n$ matrix L and an arbitrary prior guess $\bar{x}$. Thus, we may take

$$\Gamma^{-1} = \beta\Phi'L'L\Phi,$$

for some $\beta > 0$. If the matrix $\Phi'L'L\Phi$ is not available in analytical form, we may estimate $\Phi'L'L\Phi$ based on simulation and take Γ^{-1} to be

$$\Gamma^{-1} = \frac{\beta}{t+1}\sum_{k=0}^{t}\frac{l_{i_k j_k}l_{i_k \bar{j}_k}}{\xi_{i_k j_k \bar{j}_k}}\phi_{j_k}\phi_{\bar{j}_k}',$$

where we denote by l_{ij} the (i,j)th component of L.

2.3   Extensions

2.3.1   Special Case 1: Underdetermined Problems

In dealing with severely underdetermined problems (see [KS04] for examples of inverse problems of this type), we can estimate the components of

the high-dimensional solution x^* of problem (2) directly, without subspace approximation. Assuming that m is reasonably small, we may take $\Phi = I$ and adapt the preceding methodology as follows.

Let $\Sigma^{-1} = I$, $\Gamma^{-1} = \beta I$, and $Z = I$ in Eq. (7) for some $\beta > 0$, and let $\hat{x} = \Phi \hat{r} = \hat{r}$ and $\bar{x} = \Phi \bar{r} = \bar{r}$. Equation (7) can be rewritten as

$$\hat{x} = \bar{x} + (A'A + \beta I)^{-1} A'(b - A\bar{x}). \tag{10}$$

We now note that

$$A'(AA' + \beta I) = A'AA' + \beta A' = (A'A + \beta I)A',$$

and that both matrices $(AA' + \beta I)$ and $(A'A + \beta I)$ are positive definite and thus invertible. Hence we have

$$(A'A + \beta I)^{-1} A' = A'(AA' + \beta I)^{-1}.$$

Thus Eq. (10) is equivalent with

$$\hat{x} = \bar{x} + A'(F + \beta I)^{-1} d, \tag{11}$$

where we define the $m \times m$ matrix F and the m-dimensional vector d by

$$F = AA', \qquad d = b - A\bar{x}.$$

In analogy with the estimation of G and c by using Eq. (5), we may generate one sample sequence $\{(i_0, j_0), \ldots, (i_t, j_t)\}$ according to a distribution ξ, and estimate F and d with

$$\hat{F} = \frac{1}{t+1} \sum_{k=0}^{t} \frac{1}{\xi_{i_k j_k}} a_{i_k} a'_{j_k}, \quad \hat{d} = b - \frac{1}{t+1} \sum_{k=0}^{t} \frac{\bar{x}_{i_k}}{\xi_{i_k}} a_{i_k},$$

where we denote by a_{i_k} the i_kth column of A. Alternatively, in analogy with Eq. (6), we may use one sample sequence per component of F and d, and estimate $F_{\ell q}$ and d_ℓ respectively with

$$\hat{F}_{\ell q} = \frac{1}{t+1} \sum_{k=0}^{t} \frac{a_{\ell i_k} a_{q j_k}}{\xi_{i_k j_k}}, \qquad \hat{d}_\ell = b - \frac{1}{t+1} \sum_{k=0}^{t} \frac{a_{\ell i_k} \bar{x}_{i_k}}{\xi_{i_k}}.$$

We now obtain the approximate solution $\hat{x}$ whose ith entry is computed as

$$\hat{x}_i = \bar{x}_i + a'_i (\hat{F} + \beta I)^{-1} \hat{d}.$$

In this way, we are able to estimate components of x^* directly, using only low-dimensional vector operations.

2.3.2 Special Case 2: Equality Constrained Problems

As a variation of problem (2), consider the following equality constrained least-squares problem

$$\min_{x \in \Re^n} \|Ax - b\|_\zeta^2$$
$$\text{s.t. } Lx = 0,$$

$$(12)$$

where L is an $l \times n$ matrix. Following a similar approximation approach, we restrict problem (12) within the subspace S. Now the constraint $Lx = 0$ becomes $L\Phi r = 0$ or equivalently $r'\Phi'L'L\Phi r = 0$, which is also equivalent with $\Phi'L'L\Phi r = 0$. Thus, we may write the approximate problem as

$$\min_{r \in \Re^s} \|A\Phi r - b\|_\zeta^2,$$
$$\text{s.t. } \Phi'L'L\Phi r = 0.$$

$$(13)$$

We assume that there exists at least one feasible solution for this problem. Introducing a Lagrange multiplier vector $\lambda \in \Re^l$ and using standard duality arguments, we obtain the following necessary and sufficient condition for (r^*, λ^*) to be an optimal solution-Lagrange multiplier pair for problem (13):

$$H \begin{pmatrix} r^* \\ \lambda^* \end{pmatrix} = f,$$

$$(14)$$

where we define the $2s \times 2s$ matrix H and $2s$-vector f as

$$H = \begin{pmatrix} \Phi'A'ZA\Phi & \Phi'L'L\Phi' \\ \Phi'L'L\Phi & 0 \end{pmatrix}, \qquad f = \begin{pmatrix} \Phi'A'Zb \\ 0 \end{pmatrix}.$$

We may now apply our simulation-based approximation approach of the preceding section to the system (14) (which is always low-dimensional, even if L has a large row dimension). In particular, similar to Eq. (5), we may generate a sample sequence

$$\{(i_0, j_0, \bar{j}_0), \dots, (i_t, j_t, \bar{j}_t)\}$$

according to distribution ξ, and estimate H and f with $\hat{H}$ and $\hat{f}$ given by

$$\hat{H} = \frac{1}{t+1} \sum_{k=0}^{t} \frac{1}{\xi_{i_k j_k \bar{j}_k}} \begin{pmatrix} \zeta_{i_k} a_{i_k j_k} a_{i_k \bar{j}_k} \phi_{j_k} \phi'_{\bar{j}_k} & l_{i_k j_k} l_{i_k \bar{j}_k} \phi_{j_k} \phi'_{\bar{j}_k} \\ l_{i_k j_k} l_{i_k \bar{j}_k} \phi_{j_k} \phi'_{\bar{j}_k} & 0 \end{pmatrix},$$

and

$$\hat{f} = \frac{1}{t+1} \sum_{k=0}^{t} \frac{\zeta_{i_k} a_{i_k j_k} b_{i_k}}{\xi_{i_k}} \begin{pmatrix} \phi_{j_k} \\ 0 \end{pmatrix}.$$

Alternatively, we may generate one sample sequence per component of H and f, and estimate the components with formulas that are similar to Eq. (6).

2.3.3 Special Case 3: Inequality Constrained Problems

Another variation of problem (2) is the inequality constrained least-squares problem

$$\min_{x \in \Re^n} \|Ax - b\|_\zeta^2$$
$$\text{s.t. } Lx \leq g, \tag{15}$$

where L is an $l \times n$ matrix and *the row dimension l is assumed to be small.* We consider a restriction of this problem within the subspace S, given by

$$\min_{r \in \Re^s} \|A\Phi r - b\|_\zeta^2$$
$$\text{s.t. } L\Phi r \leq g,$$

or equivalently

$$\min_{r \in \Re^s} r'Gr - 2c'r,$$
$$\text{s.t. } Mr \leq g,$$

where G and c are defined in Eq. (4), and $M = L\Phi$. We may now apply the simulation approach of Section 2.1. For example, we may generate one single sample sequence, then estimate G and c with $\hat{G}$ and $\hat{c}$ using Eq. (5), and estimate M with $\hat{M}$ given by

$$\hat{M} = \frac{1}{t+1} \sum_{k=0}^{t} \frac{1}{\xi_{i_k}} l_{i_k} \phi'_{i_k},$$

where we denote by l_i the ith column of L. Alternatively, we may generate one sample sequence per component of M, and estimate $M_{\ell q}$ with

$$\hat{M}_{\ell q} = \frac{1}{t+1} \sum_{k=0}^{t} \frac{1}{\xi_{i_k}} l_{\ell i_k} \phi_{i_k q}.$$

The resulting approximate problem takes the form of

$$\min_{r \in \Re^s} r'\hat{G}r - 2\hat{c}'r,$$
$$\text{s.t. } \hat{M}r \leq g,$$

which is low-dimensional in both the cost function and the inequality constraints. Now we can apply standard quadratic programming techniques to solve this problem. Note that it is essential to assume that L has a small row dimension, so that $\hat{M}$ has low dimension.

3 Variance Reduction by Importance Sampling

3.1 Variance Analysis

The central idea of our simulation method is to evaluate G and c of Eq. (4) with a weighted average of samples generated by some probabilistic mechanism [cf. Eq. (5) and Eq. (6)]. A critical issue is the reduction of the variances of the simulation errors $\hat{G} - G$ and $\hat{c} - c$. To this end, we consider the use of importance sampling, which aims at variance reduction in estimating large sums and integrals by choosing an "appropriate" probability distribution for generating samples.

Let Ω be the sample space, $\nu : \Omega \mapsto \Re^d$ be a function, and $\{\omega_0, \ldots, \omega_t\}$ be samples generated from Ω according to some process with invariant distribution ξ. We may estimate the large sum $z = \sum_{\omega \in \Omega} \nu_\omega$ with

$$\hat{z} = \frac{1}{t+1} \sum_{k=0}^{t} \frac{\nu_{\omega_k}}{\xi_{\omega_k}},$$

and we would like to choose ξ so that $\hat{z}$ has a small variance. If $d = 1$ and ν is a nonnegative function,[1] this variance is

$$\mathrm{var}\left\{\hat{z}\right\} = \frac{z^2}{t+1} \left(\sum_{k=0}^{t} \frac{(\nu_{\omega_k}/z)^2}{\xi_{\omega_k}} - 1 \right),$$

and is minimized when the sampling distribution is $\xi^* = \nu/z$. Calculating ξ^* is impossible because it requires knowledge of z, but by designing the distribution ξ to be close to ν/z, we can reduce the variance of $\hat{z}$ (see the companion paper [WPB09] for further analysis). In what follows in this section, we discuss a few schemes for designing importance sampling distributions, which are tailored to the data of the problem.

---

[1]The nonnegativity of ν may be assumed without essential loss of generality. If ν takes negative values, we may decompose ν as

$$\nu = \nu^+ - \nu^-,$$

so that both ν^+ and ν^- are positive functions, and then estimate separately $z_1 = \sum_{\omega \in \Omega} \nu_\omega$ and $z_2 = \sum_{\omega \in \Omega} \nu_\omega^-$.

3.2 Designing Importance Sampling Distributions

3.2.1 An Importance Sampling Scheme for Estimating $G_{\ell q}$

We focus on estimating the component $G_{\ell q}$ by generating a sequence of index triples and using Eq. (6). In this case the sample space Ω and the function ν are

$$\Omega = \{1, \ldots, n\}^3, \qquad \nu(i, j, \bar{j}) = \zeta_i a_{ij} a_{i\bar{j}} \phi_{j\ell} \phi_{\bar{j}q}.$$

We want to design the sampling distribution ξ so that it is close to ξ^* and belongs to some family of relatively simple distribution functions.

We have used a scheme that generates the indices i, j, and $\bar{j}$ sequentially. The optimal distribution satisfies

$$\xi^*_{ij\bar{j}} \propto (\phi_{j\ell} \|a_j\|_1) (\phi_{\bar{j}q} \|a_{\bar{j}}\|_1) \frac{\zeta_i a_{ij} a_{i\bar{j}}}{\|a_j\|_1 \|a_{\bar{j}}\|_1},$$

where we denote by a_j the jth column of A, and denote by $\|a_j\|_1$ the L_1 norm of a_j (i.e., $\|a_j\|_1 = \sum_{i=1}^n |a_{ij}|$). We approximate ξ^* by approximating the functions

$$\phi_{j\ell} \|a_j\|_1, \qquad \phi_{\bar{j}q} \|a_{\bar{j}}\|_1, \qquad \frac{\zeta_i a_{ij} a_{i\bar{j}}}{\|a_j\|_1 \|a_{\bar{j}}\|_1}$$

with distributions

$$\xi_j, \qquad \xi_{\bar{j}}, \qquad \xi(i \mid j, \bar{j}),$$

respectively, so $\xi^*_{ij\bar{j}}$ is approximated with $\xi_{ij\bar{j}} = \xi_j \, \xi_{\bar{j}} \, \xi(i \mid j, \bar{j})$.

Let us denote by T the approximation operator that maps a set of trial values of ν to an approximation of the entire function ν. For instance, we can take T to be a *piecewise constant approximation*: for any $y \in \Re^n$ and $I = \{i_1, \ldots, i_K\} \subset \{1, \ldots, n\}$,

$$T\left([y_i]_{i \in I}\right) = \sum_{k=1}^{K} y_{i_k} \mathbf{1}\left([(i_{k-1} + i_k)/2, (i_k + i_{k+1})/2]\right),$$

where $\mathbf{1}$ denotes the function that is equal to 1 within the corresponding interval, and 0 otherwise, and we define $i_0 = -i_1$ and $i_{K+1} = 2n - i_K$. For another example, we may take T to be the *piecewise linear approximation*,

$$T\left([y_i]_{i \in I}\right) = \sum_{k=1}^{K-1} \mathbf{L}\left((i_k, y_{i_k}), (i_{k+1}, y_{i_{k+1}})\right) \mathbf{1}\left([i_k, i_{k+1}]\right),$$

where we denote by $\mathbf{L}$ the linear function that takes value y_{i_k} at i_k and value $y_{i_{k+1}}$ at i_{k+1}, and assume without loss of generality that $i_1 = 0$ and

An Importance Sampling Scheme for Estimating $G_{\ell q}$:

1. Select a small set $[a_{ij}]_{i,j \in I}$ of components of A, and a corresponding small set of rows $[\phi'_j]_{j \in I}$ of Φ.

2. Generate the sample triple $(i_k, j_k, \bar{j}_k)$ by

    (a) sampling j_k according to $\xi_{j_k} \propto \mathrm{T}_{j_k}\left(\left[\phi_{j\ell} \sum_{i \in I} a_{ij}\right]_{j \in I}\right)$,

    (b) sampling $\bar{j}_k$ according to $\xi_{\bar{j}_k} \propto \mathrm{T}_{\bar{j}_k}\left(\left[\phi_{jq} \sum_{i \in I} a_{ij}\right]_{j \in I}\right)$,

    (c) sampling i_k conditioned on j_k and $\bar{j}_k$ according to

$$\xi(i_k \mid j_k, \bar{j}_k) \propto \mathrm{T}_{i_k}\left(\left[\zeta_i a_{ij_k} a_{i\bar{j}_k}\right]_{i \in I}\right)$$

    where we denote by $T_j(\cdot)$ the jth component of $T(\cdot)$.

$i_K = n$. The resulting importance sampling algorithm is summarized in what follows.

Figure 1 illustrates the last step of this importance sampling scheme, where A is an 1000×1000 matrix, $Z = I$ and T is taken to be the operator of piecewise constant/linear approximation. We start with a low-dimensional representation of A, namely $[a_{ij}]_{i,j \in I}$, which can be implemented using a uniformly spaced discretization as illustrated in Fig. 1(a). The resulting distributions $\xi(i_k \mid j_k, \bar{j}_k)$ are plotted in Fig. 1(b)-(c), and compared with the exact optimal conditional distribution.

3.2.2 Variations of the Importance Sampling Scheme

The importance sampling scheme given in the preceding section is only one possibility for generating samples to estimate $G_{\ell q}$. An alternative is to replace the distributions in steps 2(a) and 2(b) with

$$\xi_{j_k} \propto \phi_{j_k \ell}, \qquad \xi_{\bar{j}_k} \propto \phi_{j_k q},$$

or with approximations of the above functions. This simplified version is easier to implement, and may reduce the computational complexity greatly if Φ is known to have a simple analytical form.

We may also change the order of generating i_k, j_k, and $\bar{j}_k$. For instance, we can generate i_k first, and then j_k and $\bar{j}_k$ conditioned on i_k, according to
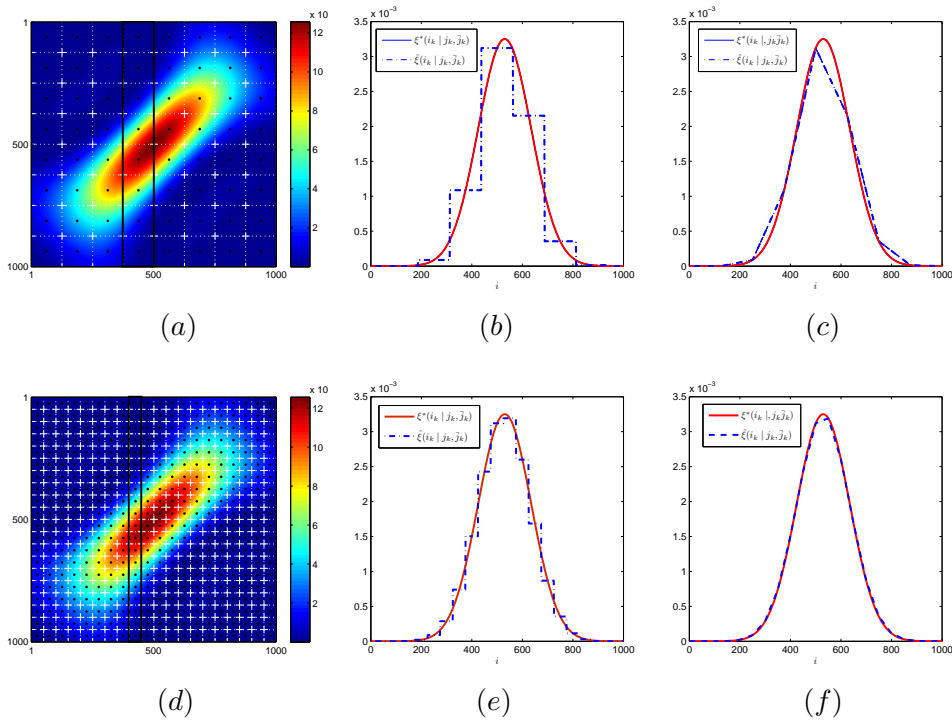
12

Figure 1: Illustration of step 2(c) of the proposed importance sampling scheme. In (a), the color field represents the 1000×1000 matrix A; the two vertical lines represent the columns a_{j_k} and $a_{\bar{j}_k}$; and the grid represents $[a_{ij}]_{i,j\in I}$, which is an 8×8 discretization of A. In (b)/(c) the conditional distribution $\xi(i_k \mid j_k, \bar{j}_k)$ (obtained by piecewise constant/linear approximation using $[a_{ij}]_{i,j\in I}$) is plotted against the optimal distribution $\xi^*(i_k \mid j_k, \bar{j}_k)$. In (d)-(f) the same process is repeated with a finer 20×20 discretization of A.

the distributions

$$\xi_{i_k} \propto \|a_{i_k}\|_1, \qquad \xi(j_k \mid i_k) \propto \phi_{j_k\ell} a_{i_k j_k}, \qquad \xi(\bar{j}_k \mid i_k) \propto \phi_{j_k q} a_{i_k j_k}.$$

If A and Φ have complicated forms, we may first replace them with coarse approximations, and then introduce a step of function approximation when computing the distributions. When A has relatively sparse rows, by sampling the row index first, we may greatly improve the efficiency of sampling.

The most straightforward scheme is to approximate the three-dimensional function $\nu = \zeta_i a_{ij} a_{i\bar{j}} \phi_{j\ell} \phi_{\bar{j}q}$ directly: first take trial samples from the sample space $\Omega = \{1, \ldots, n\}^3$ and approximate ν by fitting some function (e.g., a

13

piecewise constant/linear function) based on the trial samples. More specifically, we may take $I \subset \{1, \ldots, n\}$ and obtain $[\nu(i, j, \bar{j})]_{i,j,\bar{j} \in I}$. Then we can compute the approximate function by

$$\hat{\nu} = T\left([\nu(i, j, \bar{j})]_{i,j,\bar{j} \in I}\right),$$

where we maintain the notation T for the operator of function approximation, and finally normalize $\hat{\nu}$ to obtain a distribution function. However, this scheme may be computationally expensive, because it involves selecting trial samples from a three-dimensional space and then sampling according to a three-dimensional distribution. A critical choice in all the schemes mentioned above is the function approximation operator T, and a good choice may depend on the characteristics of the problem at hand, i.e., A and Φ.

Also, as an alternative to the piecewise constant/linear approximation used in Figure 1, we may consider an approximation approach based on a least-squares fit from some family of parameterized functions. In particular, we may approximate the function $\nu : \Omega \mapsto \Re$ by introducing a parametric family, which we denote by

$$\mathfrak{F} = \left\{ f_\eta \ \middle| \ \eta \in \Re^d \right\},$$

where $f_\eta : \Omega \mapsto \Re$ is a function parameterized by η; or we may consider the family of a finite sum of parameterized functions, in which case

$$\mathfrak{F} = \left\{ \sum_{k=1}^{M} f_{\eta_k} \ \middle| \ \eta_k \in \Re^d, k = 1, \ldots, M \right\},$$

where M is a positive integer. Given the trial samples and corresponding function values $\{\nu(i, j, \bar{j})\}_{i,j,\bar{j} \in I}$, we can approximate ν with $\hat{\nu}$ by minimizing the total squared error corresponding to the trial samples, i.e.,

$$T\left([\nu(i, j, \bar{j})]_{i,j,\bar{j} \in I}\right) = \mathrm{argmin}_{\hat{\nu} \in \mathfrak{F}} \sum_{i,j,\bar{j} \in I} \|\nu(i, j, \bar{j}) - \hat{\nu}(i, j, \bar{j})\|^2.$$

This method may be preferable in circumstances where ν is known to have some special structure, in which case we may choose $\mathfrak{F}$ accordingly and improve the approximation accuracy.

We have focused so far on estimating the components $G_{\ell q}$. It is straightforward to extend the preceding methodology to estimating components of c, and also components of F and d for underdetermined problems, cf. the special cases of Section 2.3.

14

4 Inverse Problem Applications

In this section we apply the proposed algorithmic methodology on a number of practical inverse problems, involving both underdetermined systems (see Sections 4.1 and 4.2) and square systems (see Sections 4.3-4.6). These problems take the form of Fredholm integral equations of the first kind, and are discretized into linear systems $Ax = b$, where A is $m \times n$ or $n \times n$, and n, the dimension of the solution space, is taken to be $n = 10^9$. The matrix A is typically ill-conditioned and dense. The components of A and b are accessible, and can be computed analytically.

We aim for the solution x^* of the discretized system $Ax = b$. For square systems, we consider its approximate solution within a subspace spanned by $s = 50$ or $s = 100$ multi-resolution basis functions, which are piecewise constant functions with disjoint local support [KS04]. For underdetermined systems, we use the approach introduced in Section 2.3.1 and estimate specific components of x^* directly. Note that the computational complexity is completely determined by s (or m for underdetermined systems). Our experiments are run on a dual processor personal computer with 4GB RAM running Matlab. The estimates $\hat{G}$ and $\hat{c}$ (or $\hat{F}$ and $\hat{d}$ for underdetermined systems) are obtained component-by-component based on separate sample sequences using Eq. (6). Each sequence is generated by using the importance sampling scheme given in Section 3.2, where we discretize the n-vectors involved (i.e., a_j and ϕ_j) into vectors of dimension 100, and use piecewise linear approximation to compute the sampling distributions. We have estimated each component of G and c (or F and d) with 10^4 samples and each sample takes $50\mu s$ on average.

The computational results are presented by comparing the high-dimensional approximate solution $\hat{x} = \Phi \hat{r}$ with the exact solution x^*, and Πx^*, the projection of x^* on the subspace S. The performances of our importance sampling schemes are assessed with the total sample covariances of estimated components of $\hat{G}$ and $\hat{c}$ (or $\hat{F}$ and $\hat{d}$ for underdetermined systems).

4.1 The inverse contamination release history problem

This is an underdetermined problem, whereby we seek to recover the release history of an underground contamination source based on measurements of plume concentration. Let $u(w, \tau)$ be the contaminant concentration at time τ and distance w away from the source, and let $x(\tau)$ be the source release at time τ. The transport of contaminant in the ground is governed by the

15

advection-diffusion model [WU96]

$$\frac{\partial u}{\partial \tau} = D \frac{\partial^2 u}{\partial w^2} - V \frac{\partial u}{\partial w}, \qquad w \geq 0, \ \tau \in [0, \tau_f],$$

subject to Cauchy initial and boundary conditions

$$u(0, \tau) = x(\tau), \quad u(w, 0) = 0, \quad \lim_{w \to \infty} u(w, \tau) = 0,$$

where D and V are coefficients for the diffusion and velocity respectively. At a time $T \geq \tau_f$ the plume concentration is distributed as

$$u(w, T) = \int_0^T \mathrm{d}\tau \, \mathbf{A}(w, T - \tau) \, x(\tau),$$

where $\mathbf{A}$ is the transport kernel

$$\mathbf{A}(w, T - \tau) = \frac{w}{\sqrt{4\pi D(T-\tau)^3}} \exp\left\{ -\frac{\left(w - V(T - \tau)\right)^2}{4D(T - \tau)} \right\}.$$

In our experiment we take $D = 0.8$, $V = 1$, $T = 300$, and $\tau_f = 250$, and we assume the unknown release history to be

$$x(\tau) = \sum_{i=1}^5 \kappa_i \exp\left\{ -\frac{(\tau - \mu_i)^2}{2\sigma_i^2} \right\},$$

where

$$\kappa = \{0.5, 0.4, 0.3, 0.5, 0.5\}, \ \mu = \{60, 75, 150, 190, 225\}, \ \sigma = \{35, 12, 10, 7, 3\},$$

and we discretize it into a vector of length 10^9, which is used as the vector x^*. Then we compute m borehole concentration measurements at locations $\{w_i\}_{i=1}^m$, as a discretization of $u(w, T)$ and form the vector b.

In accordance with Section 2.3.1, we formulate the problem into Eq. (11) and estimate F and d using simulation. Then we compute 1000 entries of $\hat{x}$ using the estimates $\hat{F}$ and $\hat{d}$, the regularization matrix $\Gamma^{-1} = 10^{-11}I$ and the initial guess $\bar{r} = 0$. In Fig. 2, we compare the resulted entries $\hat{x}_i$ against those of the exact solution x^*.

To analyze the effect of importance sampling, we evaluate the simulation error in terms of the total sample variances for components of $\hat{F}$ and $\hat{d}$. In Fig. 3 we compare the reduction of simulation error for alternative importance sampling schemes and alternative ways of function approximation. It can be seen that the proposed importance sampling scheme substantially reduces the simulation error and improves the simulation efficiency. Similar results have been observed in the subsequent problems.
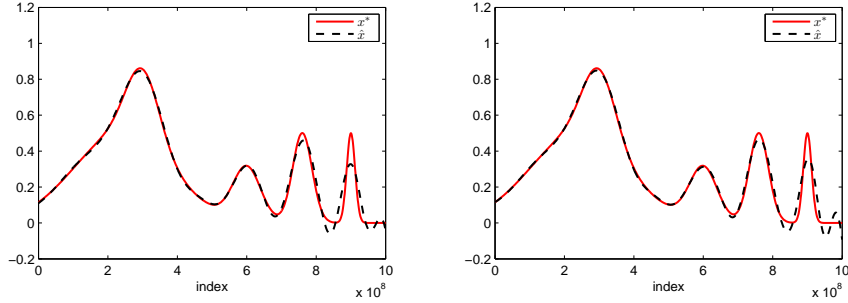
Figure 2: The simulation-based approximate solution $\hat{x}$ for the contamination release history reconstruction problem, compared with the exact solution x^*, with $m = 50$ (left) and $m = 100$ (right).
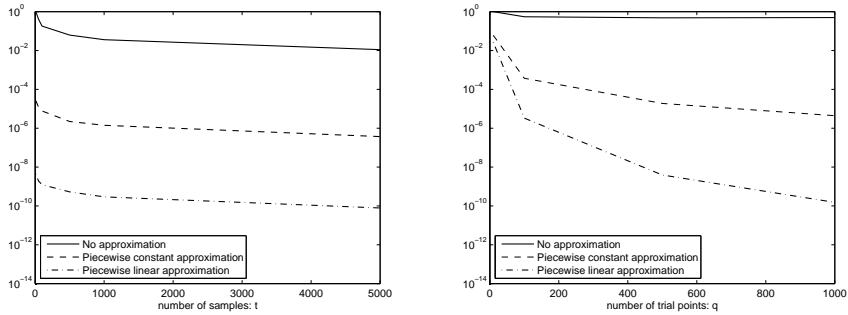

Figure 3: The reduction of simulation error for alternative importance sampling schemes. The simulation error is measured in terms of the sum of sample covariances for components of $\hat{F}$ and $\hat{d}$. The solid lines represent the case where no approximation is implemented and a uniform sampling distribution is used; the dotted lines represent the cases where importance sampling is used, with distributions obtained by piecewise constant/linear approximations. The left-side figure illustrates the reduction of simulation error as the number of samples t varies, while the number of trial points (i.e., the cardinality of I, which is introduced for the purpose of function approximation; see Section 3.2.1) is fixed at $q = 500$; the right-side figure plots the results when q varies, with the number of samples fixed at $t = 1000$.
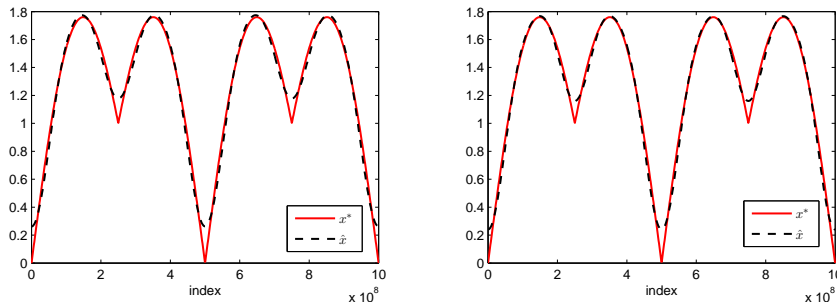
17

Figure 4: The simulation-based approximate solution $\hat{x} = \Phi\hat{r}$ for the gravitational prospecting problem, compared with the exact solution x^*, with $m = 50$ (left) and $m = 100$ (right).

4.2 Gravitational prospecting

This is an inverse problem encountered in searching for oil and natural gas resources. We want to estimate the earth density distribution based on measurements of gravitational force at some distance away from the surface.

Here we consider a simplified version of this problem as posed in [Gro07], where the spatial variation of the density is confined within the interior of a ring-shaped domain, and the measurements b are take on a circular trajectory positioned at the same plane but outside the ring. When the unknown density function x and the data are defined on concentric trajectories, we express the problem in polar coordinates as

$$b(\varphi) = \int_0^{2\pi} d\theta \mathbf{A}(\varphi, \theta) x(\theta), \quad 0 \le \varphi \le 2\pi.$$

where

$$\mathbf{A}(\varphi, \theta) = \frac{2 - \cos(\varphi - \theta)}{(5 - 4\cos(\varphi - \theta))^{3/2}}.$$

In the experiment, we take the unknown density function to be

$$x(\theta) = |\sin\theta| + |\sin 2\theta|, \quad 0 \le \theta \le 2\pi,$$

so the measurement function b can be computed accordingly. We discretize the problem into a system of $m = 50$ and $m = 100$ equations, corresponding to m measurements, with $n = 10^9$ unknowns. For regularization we use $\Gamma^{-1} = 10^{-13}I$ and $\bar{r} = 0$. The approximate solution $\hat{x}$ is illustrated in Fig. 4, compared with x^* the exact solution.

18

4.3 The second derivative problem

This problem refers to differentiating noisy signals that are usually obtained from experimental measurements. This problem has been extensively studied and the solution has been shown to exhibit instability with increasing level of noise [Cul71]. We denote by b the noisy function to be differentiated and denote by x its second derivative. It is given by

$$b(w) = \int_0^1 d\tau\, \mathbf{A}(w, \tau) x(\tau) \qquad 0 \leq \tau, w \leq 1,$$

where $\mathbf{A}(w, \tau)$ is the Green's function of the second derivative operator

$$\mathbf{A}(w, \tau) = \begin{cases} w(\tau - 1) & w < \tau, \\ \tau(w - 1) & w \geq \tau. \end{cases}$$

In our experiment, we have used

$$x(\tau) = \cos(2\pi\tau) - \sin(6\pi\tau).$$

Following the approach of [Han94] we discretize the integral using the Galerkin method, and obtain a system of n linear equations with n unknowns where $n = 10^9$.

   We consider the approximate solution of the system using the preceding methodology, with the initial guess $\bar{r} = 0$ and the regularization matrix $\Gamma = 10^{-5} L_3' L_3$, where L_3 is the $(s-3) \times s$ third-order difference operator. The obtained approximate solution $\Phi\hat{r}$ is presented in Fig. 5, and is compared with the exact solution x^* and the projected solution Πx^*.

4.4 The Fox & Goodwin problem

This problem, introduced by Fox and Goodwin in [FG53], considers the solution of the integral equation

$$b(w) = \int_0^1 d\tau \sqrt{w^2 + \tau^2}\, x(\tau), \quad 0 \leq w \leq 1.$$

As shown in [FG53], this is a severely ill-posed problem and the condition number of its discretized integral operator increases exponentially with n. In our experiment, we assume the unknown solution to be
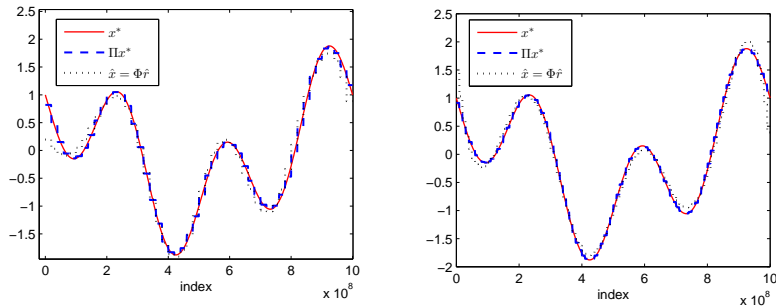
$$x(\tau) = \tau, \quad 0 \leq \tau \leq 1,$$

Figure 5: The simulation-based approximate solution $\hat{x} = \Phi\hat{r}$ for the second derivative problem, compared with the exact solution x^* and the projected solution Πx^*. The subspace S has dimension $s = 50$ for the left-hand plot and dimension $s = 100$ for the right-hand plot.

compute b accordingly, and discretize the system into a square linear system of dimension $n = 10^9$.

We consider its approximate solution in the subspace spanned by $s = 50$ or $s = 100$ multi-resolution basis functions, and introduce the regularization matrix $\Gamma^{-1} = 10^{-3} I$ and the initial guess $\bar{r} = 0$. The obtained approximate solution $\Phi\hat{r}$ is presented in Fig. 6, plotted against the exact solution x^* and the projected solution Πx^*.

4.5 The inverse heat conduction problem

This problem seeks to reconstruct the time profile of a heat source by monitoring the temperature at a fixed distance away [Car82]. The one-dimensional heat transfer in a homogeneous quarter plane medium, with known heat conductivity α, is expressed by the elliptic partial differential (heat) equation

$$\frac{\partial u}{\partial \tau} = \alpha \frac{\partial^2 u}{\partial w^2}, \qquad w \geq 0, \ \tau \geq 0,$$

$$u(w, 0) = 0, \quad u(0, \tau) = x(\tau),$$

where $u(w, \tau)$ denotes the temperature at location w and time τ. Let b be the temperature at a fixed location $\bar{w}$ away from the source, and it satisfies

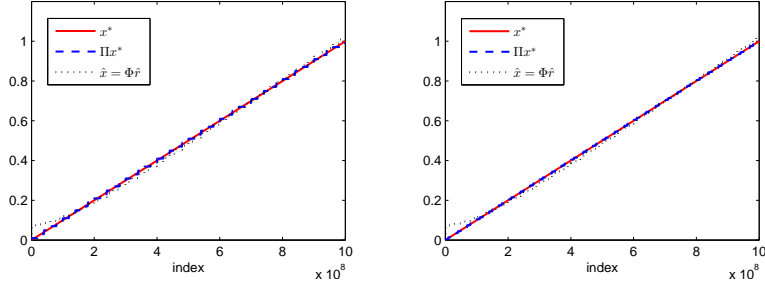$$b(\tau) = \int_0^T \mathrm{d}v \mathbf{A}(v, \tau) x(v),$$

20

Figure 6: The simulation-based approximate solution $\hat{x} = \Phi\hat{r}$ for the Fox-Goodwin problem, compared with the exact solution x^* and the projected solution Πx^*. The subspace S has dimension $s = 50$ for the left-hand plot and dimension $s = 100$ for the right-hand plot.

where $\mathbf{A}$ is a lower-triangular kernel given by

$$\mathbf{A}(\upsilon, \tau) = \begin{cases} \frac{\bar{w}/\alpha}{\sqrt{4\pi(\upsilon-\tau)^3}} \exp\left\{-\frac{(\bar{w}/\alpha)^2}{4(\upsilon-\tau)}\right\}, & 0 \leq \tau < \upsilon \leq T, \\ 0, & 0 \leq \upsilon \leq \tau \leq T. \end{cases}$$

In the experiment we take $T = 1$ and take the unknown target temperature function to be

$$x(\tau) = \sum_{i=1}^{3} \kappa_i \exp\left\{-\frac{(\tau-\mu_i)^2}{2\sigma_i^2}\right\}, \quad 0 \leq \upsilon \leq 1,$$

with $\kappa = \{4, 3, 6\} \times 10^{-4}$, $\mu = \{0.3, 0.6, 0.8\}$ and $\sigma = \{0.1, 0.1, 0.05\}$, so b can be obtained accordingly.

We discretize the integral equation into a linear square system of dimension $n = 10^9$ and consider its approximate solution within the subspace spanned by $s = 50$ or $s = 100$ multi-resolution basis functions. Also we assume an initial guess $\bar{r} = 0$ and the regularization matrix $\Gamma^{-1} = \beta L_1' L_1$, where L_1 is the $(s-1) \times s$ discrete first-order difference operator and $\beta = 10^{-5}$. The computational results are illustrated in Fig. 7.

4.6 A problem in optical imaging

Consider light passing through a thin slit, where the intensity of the diffracted light is a function of the outgoing angle and can be measured by some instrument. We wish to reconstruct the light intensity at the incoming side of the slit based on these measurements. Let x be the incoming light intensity
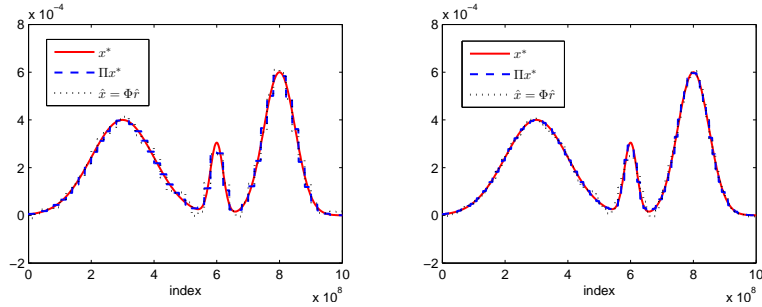
Figure 7: The simulation-based approximate solution $\hat{x} = \Phi\hat{r}$ for the inverse heat conduction problem, compared with the exact solution x^* and the projected solution Πx^*. The subspace S has dimension $s = 50$ for the left-hand plot and dimension $s = 100$ for the right-hand plot.

as a function of the incoming angle, and let b be the outgoing light intensity as a function of the outgoing angle, so that

$$b(\varphi) = \int_{-\pi/2}^{\pi/2} d\theta\, \mathbf{A}(\varphi, \theta)\, x(\theta), \quad \varphi, \theta \in [-\pi/2, \pi/2],$$

where

$$\mathbf{A}(\varphi, \theta) = \left(\cos\varphi + \cos\theta\right)^2 \left(\frac{\sin(\pi(\sin\varphi + \sin\theta))}{\pi(\sin\varphi + \sin\theta)}\right)^2$$

(we refer to [Jr.72] for further explanation of the physical aspects of this application). We discretize this integral equation into a square system of dimension $n = 10^9$, and consider its approximation within the subspace spanned by $s = 50$ and $s = 100$ multi-resolution functions. The regularization matrix is taken to be $\Gamma^{-1} = \beta L_3' L_3$ and $\bar{r} = 0$, where L_3 is the third-order difference operator and $\beta = 10^{-5}$. The corresponding computational results are plotted in Fig. 8.

5 Conclusions

In this paper, we have considered the approximate solution of linear inverse problems within a low-dimensional subspace spanned by an arbitrary given set of basis functions. We have proposed a simulation-based regularized regression approach, which can also be applied to large-scale problems with equality or inequality constraints. The algorithm uses importance sampling
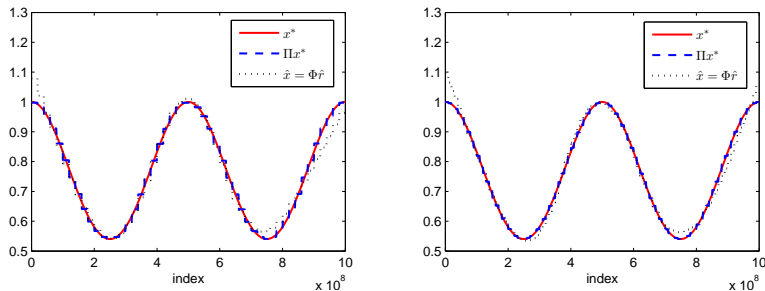
Figure 8: The simulation-based approximate solution $\hat{x} = \Phi\hat{r}$ for the optics problem, compared with the exact solution x^* and the projected solution Πx^*. The subspace S has dimension $s = 50$ for the left-hand plot and dimension $s = 100$ for the right-hand plot.

and low-dimensional computations, and relies on designing sampling distributions involving the model matrices and basis functions spanning the subspace. We have elaborated on a few approaches for designing near-optimal distributions, which exploit the continuity of the underlying models. The performance of our method has been numerically evaluated on a number of classical problems. The computation experiments demonstrate an adequate reduction in simulation noise after a relatively small number of samples and an attendant improvement in quality of the resulted approximate solution.

A central characteristic of our methodology is the use of low-dimensional calculations in solving high-dimensional problems. Two important approximation issues arise within this context: first the solution of the problem should admit a reasonably accurate representation in terms of a relatively small number of basis functions, and second, the problem should possess a reasonably continuous/smooth structure so that effective importance sampling distributions can be designed with relatively small effort. In our computational experiments, simple piecewise polynomial approximations have proved adequate, but other more efficient alternatives may be possible. We finally note that the use of regularized regression based on a sample covariance obtained as a byproduct of the simulation was another critical element for the success of our methodology with nearly singular problems.

References

[BB98]   M. Bertero and P. Boccacci. *Introduction to Inverse Problems in Imaging.* Institute of Physics, Bristol, 1998.

23

[Ber07]   D.P. Bertsekas. Solution of large systems of equations using approximate dynamic programming methods. *Lab. for Information and Decision Systems Report 2754, MIT*, 2007.

[BT96]   D. P. Bertsekas and J. Tsitsiklis. *Neuro-Dynamic Programming.* Athena Scientific, Nashua, USA, 1996.

[BY09]   D.P. Bertsekas and H. Yu. Projected equation methods for approximate solution of large linear systems. *Journal of Computational and Applied Mathematics*, 227:27–50, 2009.

[Car82]   A. Carasso. Determining surface temperatures from interior observations. *SIAM Journal of Applied Mathematics*, 42(3):558–574, 1982.

[Cul71]   J. Cullum. Numerical differentiation and regularization. *SIAM Journal of Numerical Analysis*, 8(2):254–265, 1971.

[DCZ99]   L. Reichel D. Calvetti and Q. Zhang. *Applied and Computational Control, Signals and Systems*, volume 1, chapter Iterative solution methods for large linear discrete ill-posed problems, pages 313–367. Birkhauser, 1999.

[FG53]   L. Fox and E. T. Goodwin. The numerical solution of non-singular linear integral equations. *Phil. Trans. R. Soc. Lond. A*, 245:501–534, 1953.

[Gro07]   C.W. Groetsch. Integral equations of the first kind, inverse problems and regularization: a crash course. *Journal of Physics: Conference Series: Inverse Problems in Applied Sciences, towards breakthrough*, 73:1–32, 2007.

[Han94]   P.C. Hansen. Regularization tools: A matlab package for analysis and solution of discrete ill-posed problems. *Numerical Algorithms*, 6:1–35, 1994.

[HH93]   M. Hanke and P.C. Hansen. Regularization methods for large scale problems. *Surveys on Mathematics for Industry*, 3:253–315, 1993.

[Jr.72]   C. B. Shaw Jr. Imrovement of the resolution of an instrument by numerical solution of an integral equation. *Journal of Mathematical Analysis and Applications*, 37:83–112, 1972.

[KS04]   J. Kaipio and E. Somersalo. *Statistical and computational inverse problems*. Springer, New York, 2004.

[Lan58]   C. Lanczos. Iterative solution of large-scale linear systems. *J. Soc. Indust. and Appl. Math.*, 6(91):91–109, 1958.

[OS81]   D. P. O'Leary and J. A. Simmons. A bidiagonalization- regularization procedure for large scale discretizations of ill-posed problems. *SIAM J. on Scientific and Statistical Computing*, 2(4):474–489, 1981.

[RS02]   M. Rojas and D. C. Sorensen. A trust region approach to the regularization of large-scale discrete forms of ill-posed problems. *SIAM J. on Scientific Computing*, 23(6):1843–1861, 2002.

[SB98]   R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, Cambridge, 1998.

[WPB09] M. Wang, N. Polydorides, and D. P. Bertsekas. Approximate simulation-based solution of large-scale least squares problems. *LIDS Report, MIT*, 2009.

[WU96]   A.D. Woodbury and T.J. Ulrych. Minimum relative entropy inversion: Theory and application to recovering the release history of a groundwater contaminant. *Water Resoures Research*, 9(32):2671–2681, 1996.