Polynomial acceleration of Gibbs sampling for inverse problems

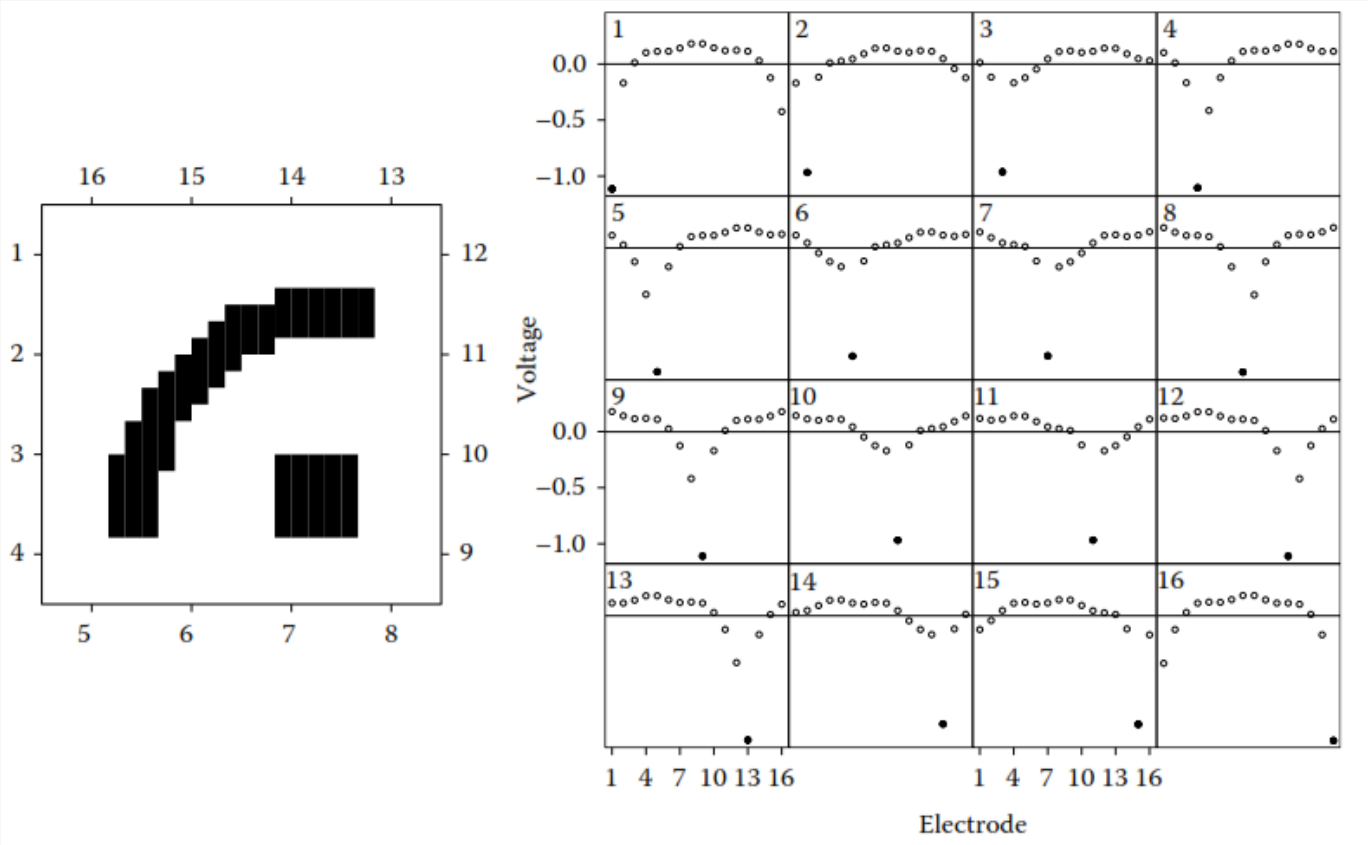
Colin Fox

Al Parker, Markus Neumayer

Two newish technologies

- Gibbs sampling for impedance tomography (EIT)
  - and other inverse problems
  - where Hamiltonian is quadratic in field and linear in material properties
  - (heat, sound, mechanics, electricity, ...)
- Polynomial acceleration of Gibbs sampling
  - optimal convergence of first and second moments
  - derived for Gaussians
  - learn covariance adaptively for Gaussian-like distributions (EIT)
- Some computational timings

Electrical impedance tomography (EIT)



Infer unknown spatial field $x(s)$ from observations y with Gaussian errors e

$$y = \eta(x) + e$$

$$L(y|x) \propto \exp\left\{-\frac{1}{2}(y-\eta(x))^{\mathsf{T}}\Sigma_e^{-1}(y-\eta(x))\right\}$$

Bayesian inference for the EIT inverse problem

Mathematical model for measurements $\eta : x \mapsto y$ is the Neumann boundary value problem

$$\nabla \cdot x(s) \nabla v(s) = 0, \qquad s \in \Omega$$

$$x(s) u \frac{\partial v(s)}{\partial n(s)} = j(s), \qquad s \in \partial\Omega$$

where $j(s)$ is the current at boundary location s. Voltages v at electrodes gives data y.

Solve for 16 currents: injection at one electrode and uniform removal from all electrodes.

Numerically solve using FEM discretization, e.g. $m = 24 \times 24$ pixels.

Consider a low level pixel representation for $x(s)$ with MRF prior, giving posterior

$$\pi(x|y) \propto \exp\left\{-\frac{1}{2}(y - \eta(x))^{\mathsf{T}} \Sigma_e^{-1} (y - \eta(x))\right\} \exp\left\{\beta \sum_{i \sim j} u(x_i - x_j)\right\}$$

Not Gaussian, but can be evaluated (expensive) so is amenable to MH MCMC

F Nicholls 1997, Moulton F Svyatskiy 2007, Higdon Reese Moulton Vrugt F 2011

Gibbs sampling for EIT

EIT operator is a $\mathbf{WSW}^\mathsf{T}$ system: $\nabla \cdot x \nabla$

$\mathbf{W}$ is geometry, $\mathbf{S}$ is diagonal matrix of material properties

FEM discretization preserves (or creates) this

$$\mathbf{K}v = j \quad \text{where system matrix} \quad \mathbf{K} = \mathbf{W}\mathbf{S}_x\mathbf{W}^\mathsf{T}$$

Maintain Greens functions = columns of $\mathbf{K}^{-1}$ corresponding to electrodes

$$(\mathbf{K} + \Delta\mathbf{K})^{-1} = \mathbf{K}^{-1} - \mathbf{K}^{-1}\mathbf{W}\left(\mathbf{I} + \mathbf{S}_\Delta\mathbf{S}_x^{-1}\tilde{\mathbf{W}}^{-\mathsf{T}}\mathbf{W}\right)^{-1}\mathbf{S}_x\mathbf{W}^\mathsf{T}\mathbf{K}^{-1}$$

where $\tilde{\mathbf{W}}^{-\mathsf{T}}$ is a psuedo-inverse of $\mathbf{W}^\mathsf{T}$ that can be pre-evaluated

The matrix pencil

$$\left(\mathbf{I} + \gamma\mathbf{S}_\Delta\mathbf{S}_x^{-1}\tilde{\mathbf{W}}^{-\mathsf{T}}\mathbf{W}\right)\mathbf{u} = \mathbf{c}$$

solve for ≈free in co-ordinate directions, cheaply when $\lesssim 20$ components non-zero

Hence we can evaluate the likelihood cheaply in these directions, and perform Gibbs sampling (e.g. by ARS)

Strang *Intro. to App. Math.*, Meyer Cai Perron 2008, Neumayer PhD 2011

Accelerating MCMC (cartoon)

MH-MCMC repeatedly simulates fixed transition kernel $\mathcal{P}$ with

$$\pi\mathcal{P} = \pi$$

via detailed balance (self-adjoint in metric π)

n-step distribution: $\pi^{(n)} = \pi^{(n-1)}\mathcal{P} = \pi^{(0)}\mathcal{P}^n$

n-step distribution error: $\pi^{(n)} - \pi = \left(\pi^{(0)} - \pi\right)\mathcal{P}^n$

Accelerating MCMC (cartoon)

MH-MCMC repeatedly simulates fixed transition kernel $\mathcal{P}$ with

$$\pi\mathcal{P} = \pi$$

via detailed balance (self-adjoint in metric π)

n-step distribution: $\pi^{(n)} = \pi^{(n-1)}\mathcal{P} = \pi^{(0)}\mathcal{P}^n$

n-step distribution error: $\pi^{(n)} - \pi = \left(\pi^{(0)} - \pi\right)\mathcal{P}^n$

Initial error is multiplied by n-th order polynomial P_n of $(I - \mathcal{P})$ called the the error polynomial

Accelerating MCMC (cartoon)

MH-MCMC repeatedly simulates fixed transition kernel $\mathcal{P}$ with

$$\pi\mathcal{P} = \pi$$

via detailed balance (self-adjoint in metric π)

n-step distribution: $\pi^{(n)} = \pi^{(n-1)}\mathcal{P} = \pi^{(0)}\mathcal{P}^n$

n-step distribution error: $\pi^{(n)} - \pi = \left(\pi^{(0)} - \pi\right)\mathcal{P}^n$

error polynomial: $P_n(I - \mathcal{P}) = \mathcal{P}^n = (I - (I - \mathcal{P}))^n$ or $P_n(\lambda) = (1 - \lambda)^n$

Hence convergence is geometric for this iteration

Accelerating MCMC (cartoon)

MH-MCMC repeatedly simulates fixed transition kernel $\mathcal{P}$ with

$$\pi\mathcal{P} = \pi$$

via detailed balance (self-adjoint in metric π)

n-step distribution: $\pi^{(n)} = \pi^{(n-1)}\mathcal{P} = \pi^{(0)}\mathcal{P}^n$

n-step distribution error: $\pi^{(n)} - \pi = \left(\pi^{(0)} - \pi\right)\mathcal{P}^n$

error polynomial: $P_n(I - \mathcal{P}) = \mathcal{P}^n = (I - (I - \mathcal{P}))^n$ or $P_n(\lambda) = (1-\lambda)^n$

Hence convergence is geometric for this iteration

Polynomial acceleration: modify iteration to give better error polynomial e.g.

$$(\text{choose } x^{(i)} \text{ w.p. } \alpha_i) \sim \pi^{(0)} \sum_{i=1}^{n} \alpha_i \mathcal{P}^i \qquad \text{error poly: } Q_n = \sum_{i=1}^{n} \alpha_i (1-\lambda)^i$$

We choose α_i so can choose Q_n

Can build a *perfect* sampler when π over finite state space, or Gaussian(!)

Gibbs sampling from normal distributions

Gibbs sampling[a] repeatedly samples from (block) conditional distributions

Normal distributions

$$\pi(\mathbf{x}) = \sqrt{\frac{\det(\mathbf{A})}{2\pi^n}} \exp\left\{-\frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} + \mathbf{b}^\mathsf{T}\mathbf{x}\right\}$$

precision matrix $\mathbf{A}$, covariance matrix $\Sigma = \mathbf{A}^{-1}$ (both SPD)

Mean $\bar{\mathbf{x}}$ satisfies

$$\mathbf{A}\bar{\mathbf{x}} = \mathbf{b}$$

Particularly interested in case where $\mathbf{A}$ is sparse (GMRF) and n large

When is $\pi^{(0)}$ is also normal, then so is the n-step distribution

$$\mathbf{A}^{(n)} \to \mathbf{A} \qquad \Sigma^{(n)} \to \Sigma$$

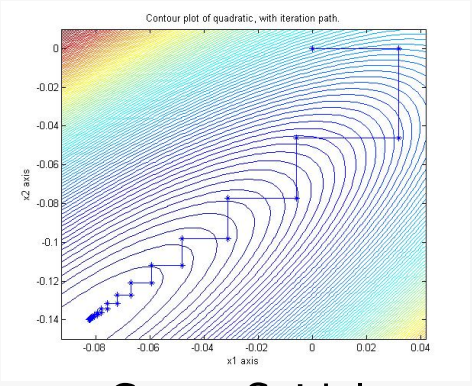
In what sense is "stochastic relaxation" related to "relaxation"?

What decomposition of $\mathbf{A}$ is this performing?

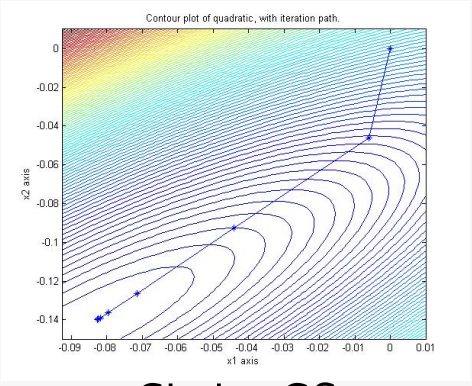
[a] Glauber 1963 (heat-bath algorithm), Turcin 1971, Geman and Geman 1984

Gibbs samplers and equivalent linear solvers

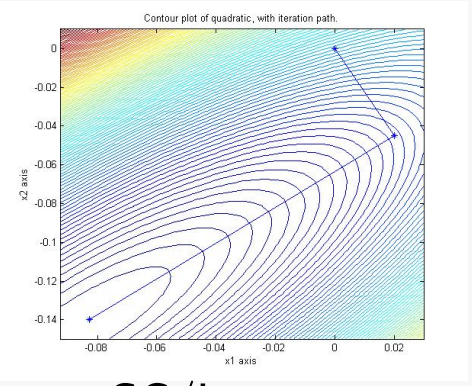
Optimization ...



Gauss-Seidel

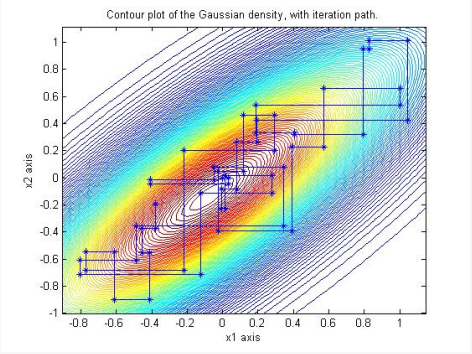


Cheby-GS

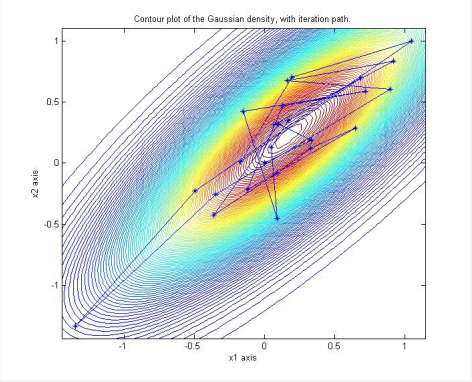


CG/Lanczos

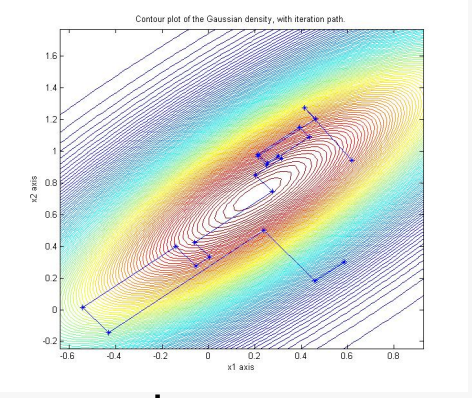
Sampling ...



Gibbs



Cheby-Gibbs



Lanczos

Matrix splitting form of stationary iterative methods

Want to solve

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

The *splitting* $\mathbf{A} = \mathbf{M} - \mathbf{N}$ converts $\mathbf{A}\mathbf{x} = \mathbf{b}$ to $\mathbf{M}\mathbf{x} = \mathbf{N}\mathbf{x} + \mathbf{b}$

If $\mathbf{M}$ is nonsingular

$$\mathbf{x} = \mathbf{M}^{-1}\mathbf{N}\mathbf{x} + \mathbf{M}^{-1}\mathbf{b}$$

Iterative methods compute successively better approximations by

$$\begin{aligned}\mathbf{x}^{(k+1)} &= \mathbf{M}^{-1}\mathbf{N}\mathbf{x}^{(k)} + \mathbf{M}^{-1}\mathbf{b} \\ &= \mathbf{G}\mathbf{x}^{(k)} + \mathbf{g}\end{aligned}$$

Many splittings use terms in $\mathbf{A} = \mathbf{L} + \mathbf{D} + \mathbf{U}$. Gauss-Seidel sets $\mathbf{M} = \mathbf{L} + \mathbf{D}$

$$\mathbf{x}^{(k+1)} = -\mathbf{D}^{-1}\mathbf{L}\mathbf{x}^{(k+1)} - \mathbf{D}^{-1}\mathbf{L}^{\mathsf{T}}\mathbf{x}^{(k)} + \mathbf{D}^{-1}\mathbf{b}$$

Matrix formulation of Gibbs sampling from $\mathrm{N}(0, \mathbf{A}^{-1})$

Let $\mathbf{y} = (y_1, y_2, ..., y_n)^T$

Component-wise Gibbs updates each component in sequence from the (normal) conditional distributions

One 'sweep' over all n components can be written

$$\mathbf{y}^{(k+1)} = -\mathbf{D}^{-1}\mathbf{L}\mathbf{y}^{(k+1)} - \mathbf{D}^{-1}\mathbf{L}^T\mathbf{y}^{(k)} + \mathbf{D}^{-1/2}\mathbf{z}^{(k)}$$

where: $\mathbf{D} = \mathrm{diag}(\mathbf{A})$, $\mathbf{L}$ is the strictly lower triangular part of $\mathbf{A}$, $\mathbf{z}^{(k-1)} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$

$$\mathbf{y}^{(k+1)} = \mathbf{G}\mathbf{y}^{(k)} + \mathbf{c}^{(k)}$$

$\mathbf{c}^{(k)}$ is iid 'noise' with zero mean, finite covariance

---

Goodman & Sokal 1989

Matrix formulation of Gibbs sampling from $\mathrm{N}(0, \mathbf{A}^{-1})$

Let $\mathbf{y} = (y_1, y_2, ..., y_n)^T$

Component-wise Gibbs updates each component in sequence from the (normal) conditional distributions

One 'sweep' over all n components can be written

$$\mathbf{y}^{(k+1)} = -\mathbf{D}^{-1}\mathbf{L}\mathbf{y}^{(k+1)} - \mathbf{D}^{-1}\mathbf{L}^T\mathbf{y}^{(k)} + \mathbf{D}^{-1/2}\mathbf{z}^{(k)}$$

where: $\mathbf{D} = \mathrm{diag}(\mathbf{A})$, $\mathbf{L}$ is the strictly lower triangular part of $\mathbf{A}$, $\mathbf{z}^{(k-1)} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$

$$\mathbf{y}^{(k+1)} = \mathbf{G}\mathbf{y}^{(k)} + \mathbf{c}^{(k)}$$

$\mathbf{c}^{(k)}$ is iid 'noise' with zero mean, finite covariance

Spot the similarity to Gauss-Seidel iteration for solving $\mathbf{A}\mathbf{x} = \mathbf{b}$

$$\mathbf{x}^{(k+1)} = -\mathbf{D}^{-1}\mathbf{L}\mathbf{x}^{(k+1)} - \mathbf{D}^{-1}\mathbf{L}^{\mathsf{T}}\mathbf{x}^{(k)} + \mathbf{D}^{-1}\mathbf{b}$$

Goodman & Sokal 1989; Amit & Grenander 1991

Gibbs converges $\iff$ solver converges

Theorem 1 *Let* $\mathbf{A} = \mathbf{M} - \mathbf{N}$, $\mathbf{M}$ *invertible. The stationary linear solver*

$$\begin{aligned}\mathbf{x}^{(k+1)} &= \mathbf{M}^{-1}\mathbf{N}\mathbf{x}^{(k)} + \mathbf{M}^{-1}\mathbf{b} \\ &= \mathbf{G}\mathbf{x}^{(k)} + \mathbf{M}^{-1}\mathbf{b}\end{aligned}$$

converges, if and only if the random iteration

$$\begin{aligned}\mathbf{y}^{(k+1)} &= \mathbf{M}^{-1}\mathbf{N}\mathbf{y}^{(k)} + \mathbf{M}^{-1}\mathbf{c}^{(k)} \\ &= \mathbf{G}\mathbf{y}^{(k)} + \mathbf{M}^{-1}\mathbf{c}^{(k)}\end{aligned}$$

converges in distribution. Here $\mathbf{c}^{(k)} \overset{iid}{\sim} \pi_n$ *has zero mean and finite variance*

Proof. Both converge iff $\varrho(\mathbf{G}) < 1$ $\square$

Convergent splittings generate convergent (generalized) Gibbs samplers

Mean converges with asymptotic convergence factor $\varrho(\mathbf{G})$, covariance with $\varrho(\mathbf{G})^2$

Young 1971 Thm 3-5.1, Duflo 1997 Thm 2.3.18-4, Goodman & Sokal, 1989, Galli & Gao 2001
F Parker 2012

Average contractive iterated random functions

Diaconis Freedman SIAM Review 1999, Stenflo JDEA $\geq$2012

Some not so common Gibbs samplers for $\mathrm{N}(0, \mathbf{A}^{-1})$

splitting/sampler	$\mathbf{M}$	$\mathbf{Var}\left(\mathbf{c}^{(k)}\right) = \mathbf{M}^T + \mathbf{N}$	converge if
Richardson	$\frac{1}{\omega}\mathbf{I}$	$\frac{2}{\omega}\mathbf{I} - \mathbf{A}$	$0 < \omega < \frac{2}{\varrho(\mathbf{A})}$
Jacobi	$\mathbf{D}$	$2\mathbf{D} - \mathbf{A}$	$\mathbf{A}$ SDD
GS/Gibbs	$\mathbf{D} + \mathbf{L}$	$\mathbf{D}$	always
SOR/B&F	$\frac{1}{\omega}\mathbf{D} + \mathbf{L}$	$\frac{2-\omega}{\omega}\mathbf{D}$	$0 < \omega < 2$
SSOR/REGS	$\frac{\omega}{2-\omega}\mathbf{M}_{\mathsf{SOR}}\mathbf{D}^{-1}\mathbf{M}_{\mathsf{SOR}}^{\mathsf{T}}$	$\frac{\omega}{2-\omega}\left(\mathbf{M}_{\mathsf{SOR}}\mathbf{D}^{-1}\mathbf{M}_{\mathsf{SOR}}^{T} + \mathbf{N}_{\mathsf{SOR}}^{T}\mathbf{D}^{-1}\mathbf{N}_{\mathsf{SOR}}\right)$	$0 < \omega < 2$

Good choice has: convenient to solve $\mathbf{Mu} = \mathbf{r}$ and sample from $\mathrm{N}(0, \mathbf{M}^T + \mathbf{N})$

Relaxation parameter ω can accelerate Gibbs

SOR: Adler 1981; Barone & Frigessi 1990, Amit & Grenander 1991, SSOR: Roberts & Sahu 1997

Some not so common Gibbs samplers for $\mathrm{N}(0, \mathbf{A}^{-1})$

splitting/sampler	$\mathbf{M}$	$\mathbf{Var}\left(\mathbf{c}^{(k)}\right) = \mathbf{M}^T + \mathbf{N}$	converge if
Richardson	$\frac{1}{\omega}\mathbf{I}$	$\frac{2}{\omega}\mathbf{I} - \mathbf{A}$	$0 < \omega < \frac{2}{\varrho(\mathbf{A})}$
Jacobi	$\mathbf{D}$	$2\mathbf{D} - \mathbf{A}$	$\mathbf{A}$ SDD
GS/Gibbs	$\mathbf{D} + \mathbf{L}$	$\mathbf{D}$	always
SOR/B&F	$\frac{1}{\omega}\mathbf{D} + \mathbf{L}$	$\frac{2-\omega}{\omega}\mathbf{D}$	$0 < \omega < 2$
SSOR/REGS	$\frac{\omega}{2-\omega}\mathbf{M}_{\mathsf{SOR}}\mathbf{D}^{-1}\mathbf{M}_{\mathsf{SOR}}^{\mathsf{T}}$	$\frac{\omega}{2-\omega}\left(\mathbf{M}_{\mathsf{SOR}}\mathbf{D}^{-1}\mathbf{M}_{\mathsf{SOR}}^{T} + \mathbf{N}_{\mathsf{SOR}}^{T}\mathbf{D}^{-1}\mathbf{N}_{\mathsf{SOR}}\right)$	$0 < \omega < 2$

Good choice has: convenient to solve $\mathbf{Mu} = \mathbf{r}$ and sample from $\mathrm{N}(0, \mathbf{M}^T + \mathbf{N})$

Relaxation parameter ω can accelerate Gibbs

SSOR is a forwards and backwards sweep of SOR to give a *symmetric* splitting

SOR: Adler 1981; Barone & Frigessi 1990, Amit & Grenander 1991, SSOR: Roberts & Sahu 1997

A closer look at convergence

To sample from $\mathrm{N}(\mu, \mathbf{A}^{-1})$ where $\mathbf{A}\mu = \mathbf{b}$

Split $\mathbf{A} = \mathbf{M} - \mathbf{N}$, $\mathbf{M}$ invertible. $\mathbf{G} = \mathbf{M}^{-1}\mathbf{N}$, and $\mathbf{c}^{(k)} \overset{\text{iid}}{\sim} \mathrm{N}(0, \mathbf{M}^T + \mathbf{N})$

The iteration

$$\mathbf{y}^{(k+1)} = \mathbf{G}\mathbf{y}^{(k)} + \mathbf{M}^{-1}\left((\mathbf{c}^{(k)} + \mathbf{b}\right)$$

implies

$$\mathrm{E}\left(\mathbf{y}^{(n)}\right) - \mu = \mathbf{G}^n\left[\mathrm{E}\left(\mathbf{y}^{(0)}\right) - \mu\right]$$

and

$$\mathrm{Var}\left(\mathbf{y}^{(n)}\right) - \mathbf{A}^{-1} = \mathbf{G}^n\left[\mathrm{Var}\left(\mathbf{y}^{(0)}\right) - \mathbf{A}^{-1}\right]\mathbf{G}^n$$

(Hence asymptotic average convergence factors $\varrho(\mathbf{G})$ and $\varrho(\mathbf{G})^2$)

Errors go down as the polynomial

$$P_n(\lambda) = (1 - \lambda)^n$$

A closer look at convergence

To sample from $\mathrm{N}(\mu, \mathbf{A}^{-1})$ where $\mathbf{A}\mu = \mathbf{b}$

Split $\mathbf{A} = \mathbf{M} - \mathbf{N}$, $\mathbf{M}$ invertible. $\mathbf{G} = \mathbf{M}^{-1}\mathbf{N}$, and $\mathbf{c}^{(k)} \overset{\text{iid}}{\sim} \mathrm{N}(0, \mathbf{M}^T + \mathbf{N})$

The iteration

$$\mathbf{y}^{(k+1)} = \mathbf{G}\mathbf{y}^{(k)} + \mathbf{M}^{-1}\left((\mathbf{c}^{(k)} + \mathbf{b}\right)$$

implies

$$\mathrm{E}\left(\mathbf{y}^{(n)}\right) - \mu = \mathbf{G}^n\left[\mathrm{E}\left(\mathbf{y}^{(0)}\right) - \mu\right]$$

and

$$\mathrm{Var}\left(\mathbf{y}^{(n)}\right) - \mathbf{A}^{-1} = \mathbf{G}^n\left[\mathrm{Var}\left(\mathbf{y}^{(0)}\right) - \mathbf{A}^{-1}\right]\mathbf{G}^n$$

(Hence asymptotic average convergence factors $\varrho(\mathbf{G})$ and $\varrho(\mathbf{G})^2$)

Errors go down as the polynomial

$$P_n(\lambda) = (1 - \lambda)^n$$

solver and sampler have exactly the same error polynomial

Controlling the error polynomial

The splitting

$$\mathbf{A} = \frac{1}{\tau}\mathbf{M} + \left(1 - \frac{1}{\tau}\right)\mathbf{M} - \mathbf{N}$$

gives the iteration operator

$$\mathbf{G}_\tau = \left(\mathbf{I} - \tau\mathbf{M}^{-1}\mathbf{A}\right)$$

and error polynomial $Q_n(\lambda) = (1 - \tau\lambda)^n$

The *sequence* of parameters $\tau_1, \tau_2, \ldots, \tau_n$ gives the error polynomial

$$Q_n(\lambda) = \prod_{l=1}^{m} (1 - \tau_l \lambda)$$

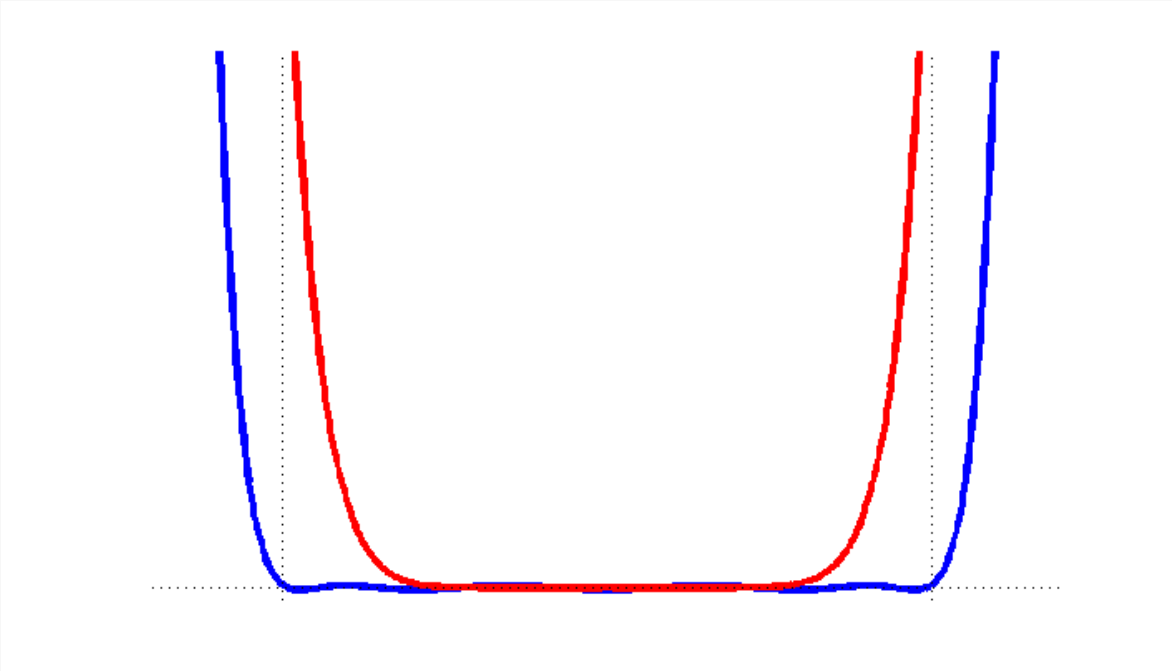
... so we can choose the zeros of Q_n

This gives a non-stationary solver $\equiv$ non-homogeneous Markov chain

---

Golub & Varga 1961, Golub & van Loan 1989, Axelsson 1996, Saad 2003, F & Parker 2012

The best (Chebyshev) polynomial



10 iterations, factor of 300 improvement

Choose

$$\frac{1}{\tau_l} = \frac{\lambda_n + \lambda_1}{2} + \frac{\lambda_n - \lambda_1}{2} \cos\left(\pi \frac{2l+1}{2p}\right) \quad l = 0, 1, 2, \ldots, p-1$$

where λ_1 λ_n are extreme eigenvalues of $\mathbf{M}^{-1}\mathbf{A}$

Second-order accelerated sampler

First-order accelerated iteration turns out to be unstable

Numerical stability, and optimality at each step, is given by the second-order iteration

$$\mathbf{y}^{(k+1)} = (1-\alpha_k)\mathbf{y}^{(k-1)} + \alpha_k \mathbf{y}^{(k)} + \alpha_k \tau_k \mathbf{M}^{-1}(\mathbf{c}^{(k)} - \mathbf{A}\mathbf{y}^{(k)})$$

with α_k and τ_k chosen so error polynomial satisfies Chebyshev recursion.

Theorem 2 *2^{nd}-order solver converges $\Rightarrow$ 2^{nd}-order sampler converges (given correct noise distribution)*
Error polynomial is optimal, at each step, for both mean and covariance

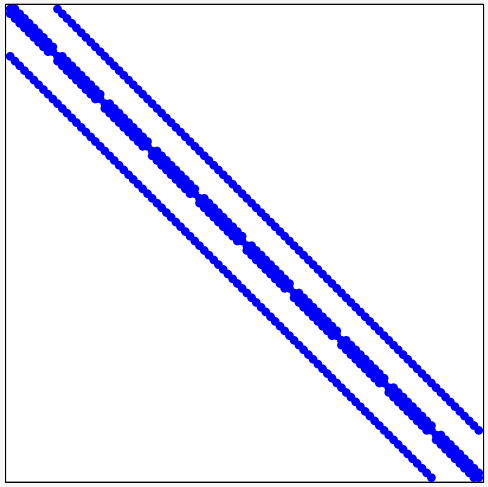
Asymptotic average reduction factor (Axelsson 1996) is

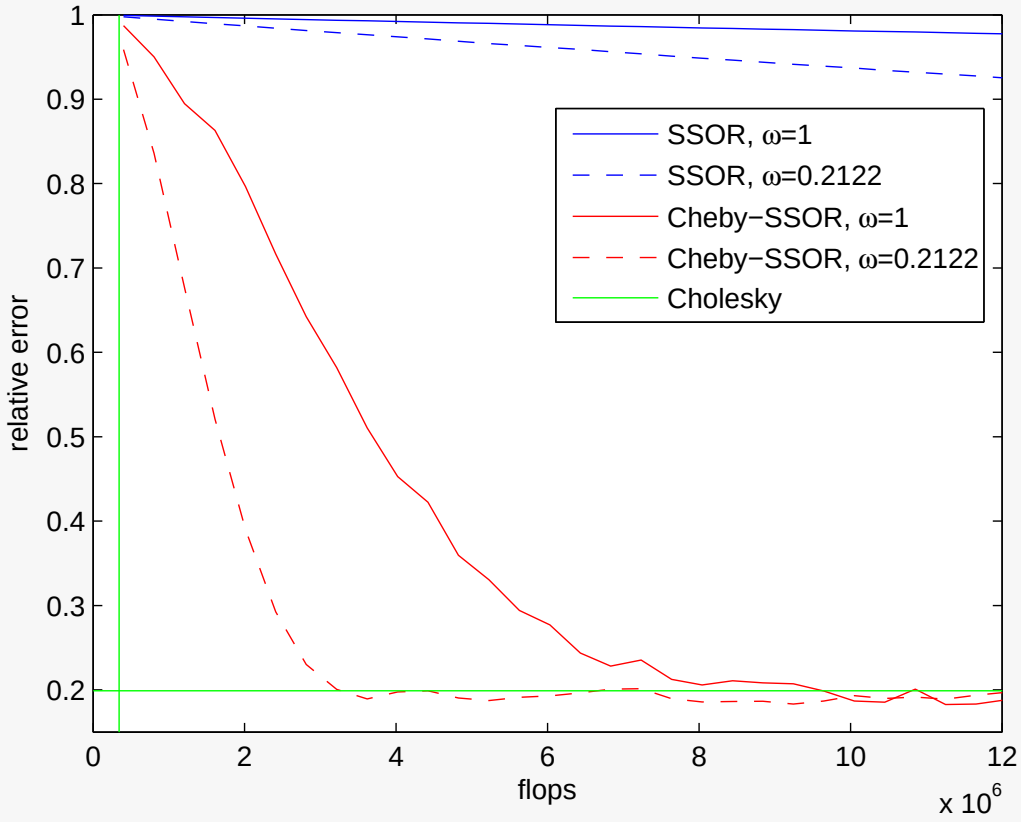
$$\sigma = \frac{1 - \sqrt{\lambda_1 / \lambda_n}}{1 + \sqrt{\lambda_1 / \lambda_n}}$$

Axelsson 1996, F & Parker 2012

10×10 lattice ($d = 100$) sparse precision matrix

$$[\mathbf{A}]_{ij} = 10^{-4}\delta_{ij} + \begin{cases} n_i & \text{if} \quad i = j \\ -1 & \text{if} \quad i \neq j \quad \text{and} \quad ||s_i - s_j||_2 \leq 1 \\ 0 & \text{otherwise} \end{cases} .$$





$\approx 10^4$ times faster

Polynomial acceleration of parameter estmation in EIT

Second-order Chebyshev acceleration of Gibbs give optimal convergence of first and second moments – given mean and inverse of covariance matrix $\mathbf{A} = \Sigma^{-1}$ where $\Sigma = \mathrm{cov}(\pi(x|y))$
We don't have $\mathbf{A}$, so we adapt to it.

Initialize $\mu = x_{\mathsf{MAP}}$ and $\mathbf{A} =$ Hessian of $-\log \pi$ at x_{MAP}

Algorithm 1 *At state x^l with values for τ and α:*

1. *Simulate x' via generalised scaled Gibbs sweep with parameter τ from x^l*
2. *Set $x^{l+1} = \alpha x' + (1-\alpha)x^{l-1}$*
3. *Evaluate recursion on α and τ*
4. *Update μ and $\mathbf{A}$ using empirical estimates (as AM)*

IACT for Gibbs was ≈ 3 sweeps bit slower than optimization
'IACT' after acceleration is ~ 1 sweep bit faster than optimization
passes all numerical tests, but no proof of convergence

Some observations

In the Gaussian setting

- GS $\equiv$ GS
- *stochastic relaxation* is fundamentally equivalent to classical *relaxation*
- existing fast solvers make fast samplers (mutligrid, fast multipole, parallel tools)

More generally

- acceleration of convergence in mean and covariance not limited to Gaussian targets
- ... can be applied to inverse problems
- A sequence of non-optimal kernels can do *much* better than repeated application of the 'optimal' kernel