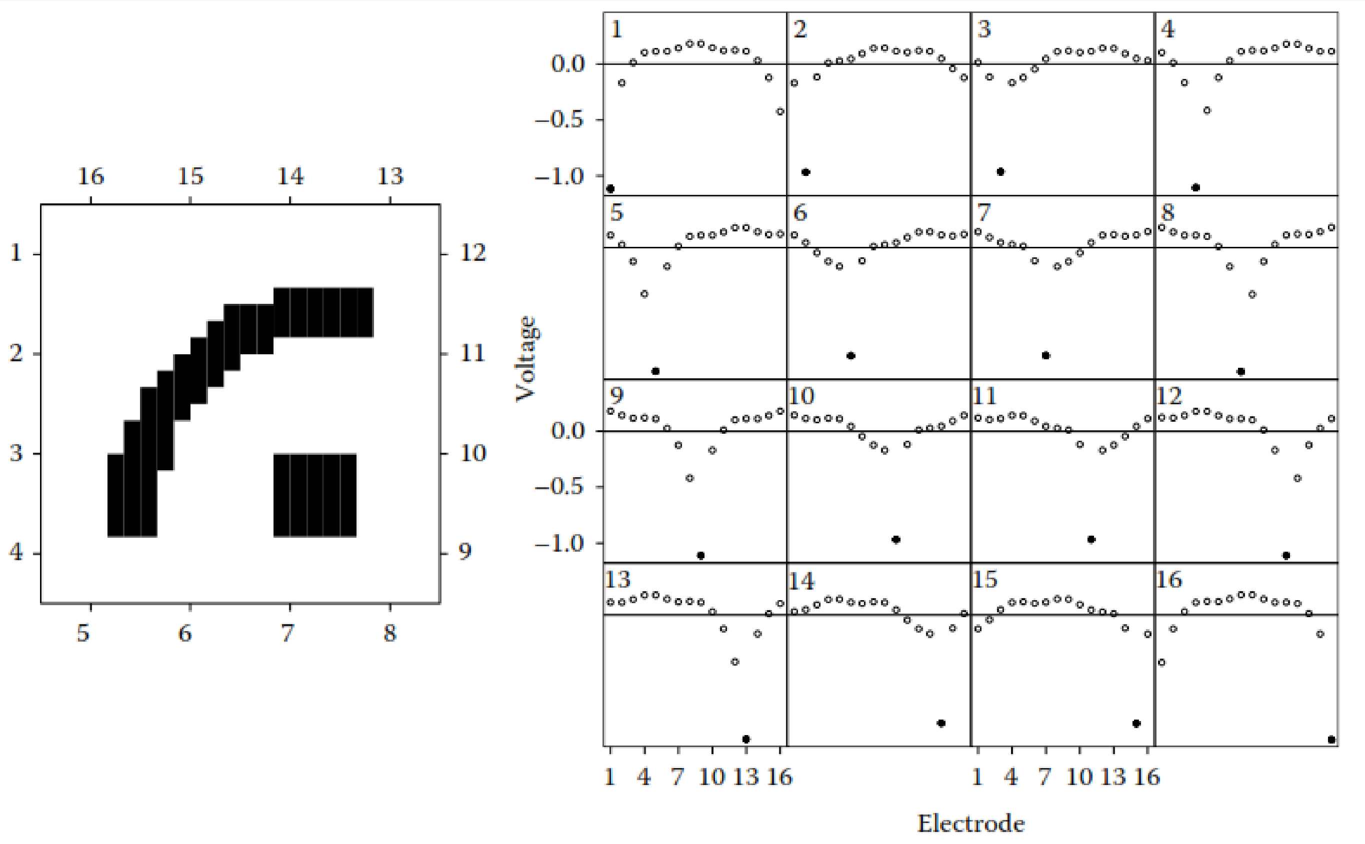
A Gibbs sampler for conductivity imaging and other inverse problems

Colin Fox

Two newish technologies

- Gibbs sampling for impedance tomography (EIT)
  - and other inverse problems
  - where PDE is 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 $\sigma(x)$ from observations d with Gaussian errors n

$$d = F(\sigma) + n$$

$$\pi(d|\sigma) \propto \exp\left\{-\frac{1}{2}(d - F(\sigma))^{\mathsf{T}}\Sigma_n^{-1}(d - F(\sigma))\right\}$$

Bayesian inference for the EIT inverse problem

Mathematical model for measurements $F : \sigma \mapsto d$ is the Neumann BVP

$$\nabla \cdot \sigma(x) \nabla \phi(x) = 0, \qquad x \in \Omega$$
$$\sigma(x) \frac{\partial \phi(x)}{\partial n(x)} = j(x), \qquad x \in \partial\Omega$$

$j(x)$ is the current at boundary location x. Voltages ϕ at electrodes gives data d.

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

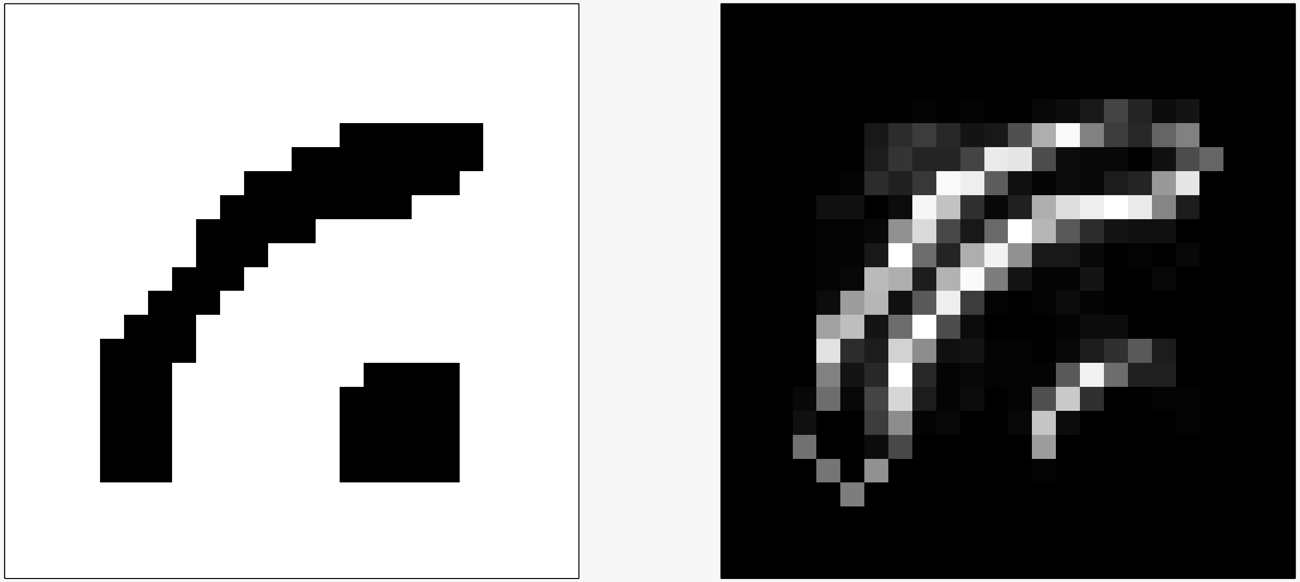
Consider a low level pixel representation for $\sigma(x)$ with MRF prior, giving posterior

$$\pi(\sigma|d) \propto \exp\left\{-\frac{1}{2}(d - F(\sigma))^{\mathsf{T}} \Sigma_n^{-1} (d - F(\sigma))\right\} \exp\left\{\beta \sum_{i \sim j} \rho(\sigma_i - \sigma_j)\right\}$$

Can be evaluated (expensive) so is amenable to MH MCMC

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

Results for EIT



(a) Marginal Posterior Mode (b) Conductivity Variance

Imaging from Strong Wave Scattering

quantity being imaged	governing PDE	PDE classification
electrical conductivity	$\nabla \cdot (\sigma \nabla \phi) = s$	elliptic
acoustic impedance	$\nabla \cdot (\sigma \nabla p) = \frac{\sigma}{c^2}\ddot{p}$	hyperbolic
thermal conductivity	$\nabla \cdot (\sigma \nabla u) = \dot{u}$	parabolic

$$\nabla \cdot \sigma \nabla = \mathbf{A}^{\mathsf{T}}\mathbf{C}\mathbf{A}$$

$\mathbf{A}$ is a 'matrix' that contains geometric information only

$\mathbf{C}$ is a diagonal 'matrix' of material properties

Strang *Intro to Applied Math* 1986

Gibbs sampling

Markov chain Monte Carlo (MCMC) methods draw (random) samples $\sigma^{(k)} \sim \pi(\sigma|d)$

Any function $g(\sigma)$ can be estimated by

$$\mathrm{E}(g|d) = \int g(\sigma)\, \pi(\sigma|d)\, \mathrm{d}\sigma \approx \frac{1}{N}\sum_{k=1}^{N} g\left(\sigma^{(k)}\right)$$

Variance $\propto N^{-1}$.

Algorithm 1 (Gibbs sampler) *At state $\sigma^{(n)} = \sigma$, simulate $\sigma^{(n+1)}$ as follows:*

1. *Select one component σ_i .*
2. *Update σ_i by sampling from the conditional distribution for σ_i, i.e. $\pi(\sigma_i|\sigma_{-i}, d)$.*

Replace conditional sampling by conditional optimization to get Gauss-Seidel optimization

---

Glauber 1963, Turcin 1971, Geman Geman 1984, Parker F 2012

Gibbs sampling for EIT

EIT operator is a $\mathbf{A}^\mathsf{T}\mathbf{C}\mathbf{A}$ system

FEM discretization preserves (or creates) this

$$\mathbf{K}\phi = j \quad \text{where system matrix} \quad \mathbf{K} = \mathbf{A}^\mathsf{T}\mathbf{C}\mathbf{A}$$

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

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

where $\tilde{\mathbf{A}}^{-1}$ is a psuedo-inverse of $\mathbf{A}$ that can be pre-evaluated

The matrix pencil

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

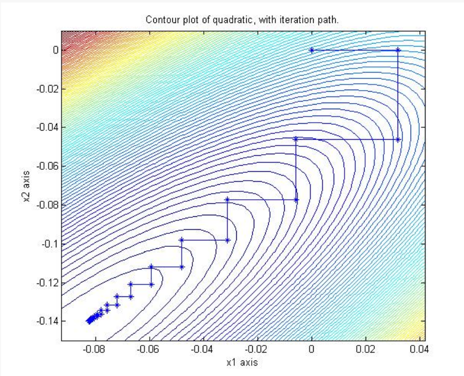
solve for ≈free in co-ordinate directions, cheaply when $\lesssim 20$ components non-zero

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

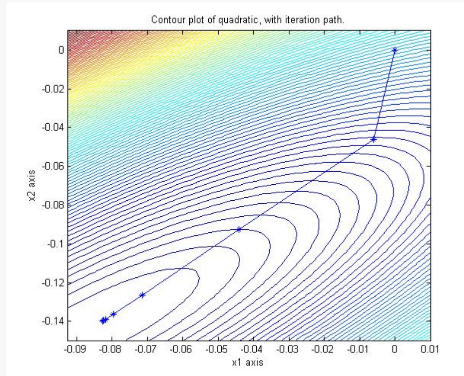
Meyer Cai Perron 2008, Neumayer PhD 2011

Gibbs samplers and equivalent linear solvers

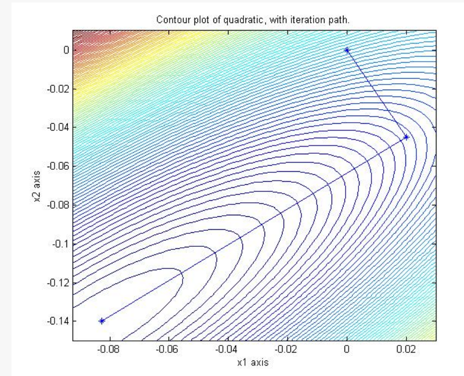
Optimization ...



Gauss-Seidel

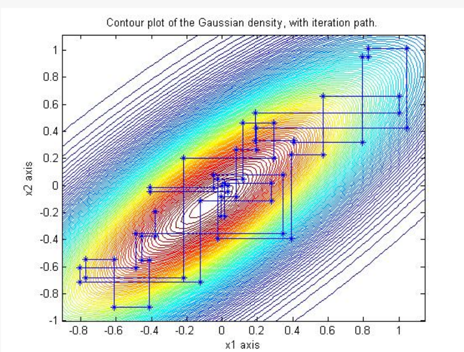


Cheby-GS

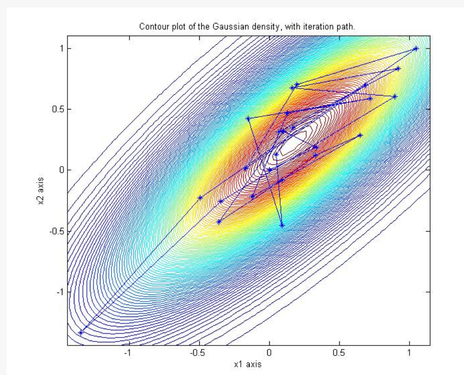


CG/Lanczos

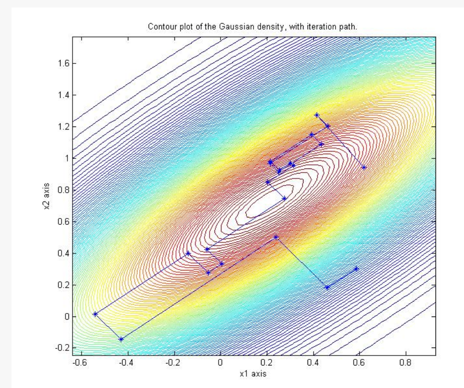
Sampling ...



Gibbs



Cheby-Gibbs



Lanczos

Parker F SISC 2012

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

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	$\frac{1}{\omega}\mathbf{I}$	$\frac{2}{\omega}\mathbf{I} - \mathbf{A}$	$0 < \omega < \dfrac{2}{\varrho(\mathbf{A})}$
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	$\frac{1}{\omega}\mathbf{I}$	$\frac{2}{\omega}\mathbf{I} - \mathbf{A}$	$0 < \omega < \dfrac{2}{\varrho(\mathbf{A})}$
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 $\mathbf{M}\mathbf{u} = \mathbf{r}$ 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

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

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

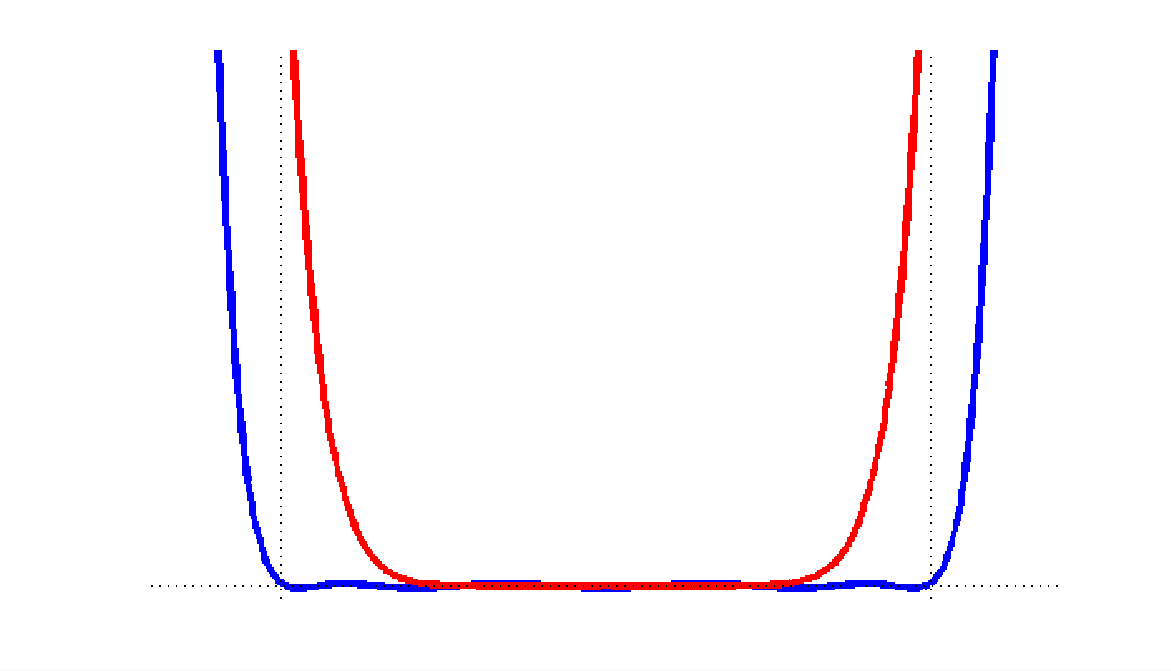
... so we can choose the zeros of Q_n

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

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

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

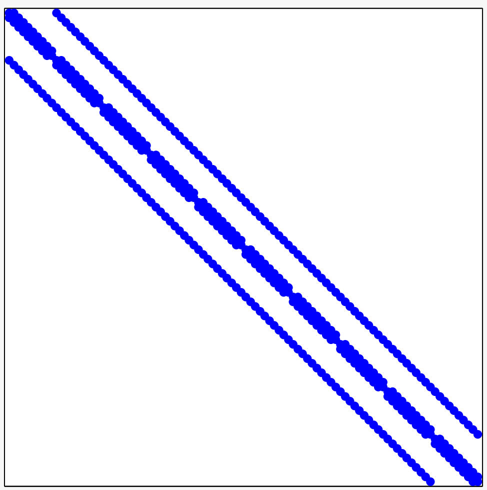
Asymptotic average reduction factor (Axelsson 1996) is

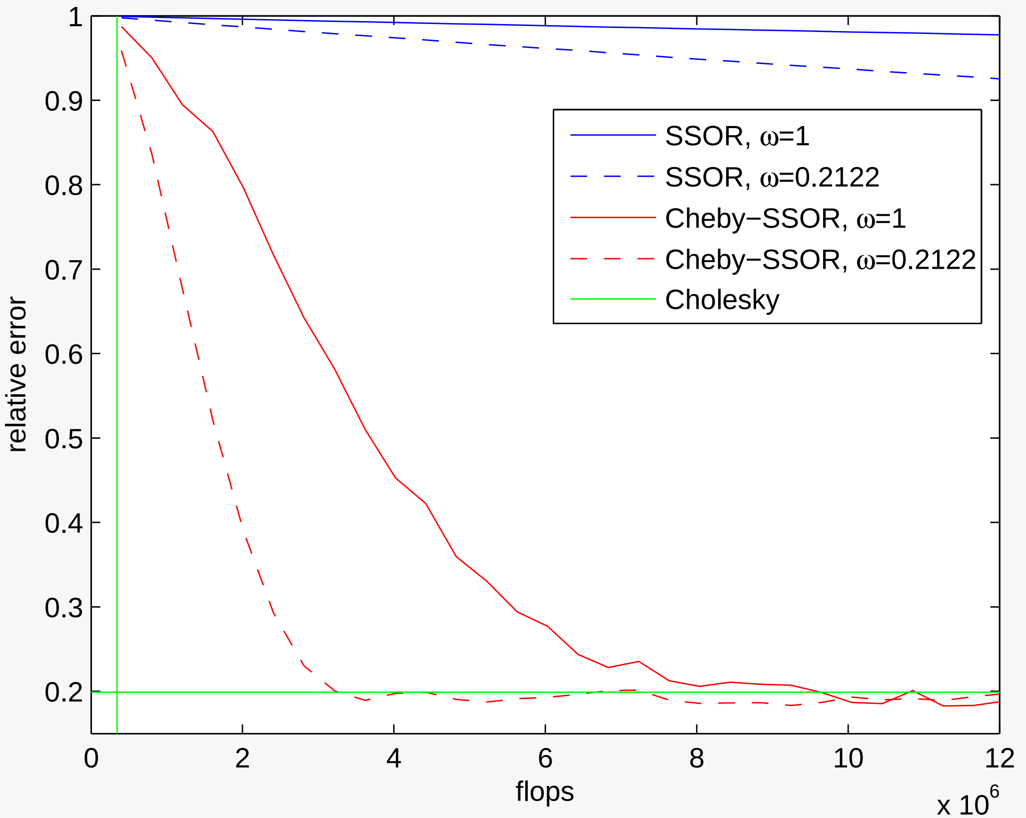
$$\sigma = \frac{1-\sqrt{\lambda_1/\lambda_n}}{1+\sqrt{\lambda_1/\lambda_n}}$$

Axelsson 1996, F & Parker 2012

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

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





$\approx 10^4$ 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 $\mathbf{P} = \Sigma^{-1}$ where $\Sigma = \mathrm{cov}(\pi(\sigma|y))$
Don't have $\mathbf{P}$, so adapt to it

Initialize $\mu = \sigma_{\mathsf{MAP}}$ and $\mathbf{P} =$ Hessian of $-\log \pi$ at σ_{MAP}

Algorithm 2 *At state σ^l with values for τ and α:*

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

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

Conclusions

For Gaussians

- Gibbs sampling is fundamentally equivalent to Gauss-Seidel
- Accelerators for linear solvers also accelerate Gibbs sampling

For EIT (and other inverse problems)

- $\mathbf{A}^{\mathsf{T}} C A$ structure allows fast conditional updates
- Convergence in mean and covariance can be accelerated