MULTISCALE OPTIMIZATION

Achi Brandt

Department of Computer Science
and Applied Mathematics
Weizmann Institute of Science
Rehovot, Israel

Collaboration with Dorit Ron

Oren Livne

Minimize $E(x_1, x_2, \ldots, x_n)$

- x_j. real / integer / ± 1 /
- Large n
- Sparse:

$$E \quad \sum_k \varphi_k(\text{small number of } x_j\text{'s})$$

- Mostly local: $\varphi_k(\text{neighboring } x_j\text{'s})$

  or asymptotically smooth $\varphi_k(x_i, x_j)$

Relaxation:

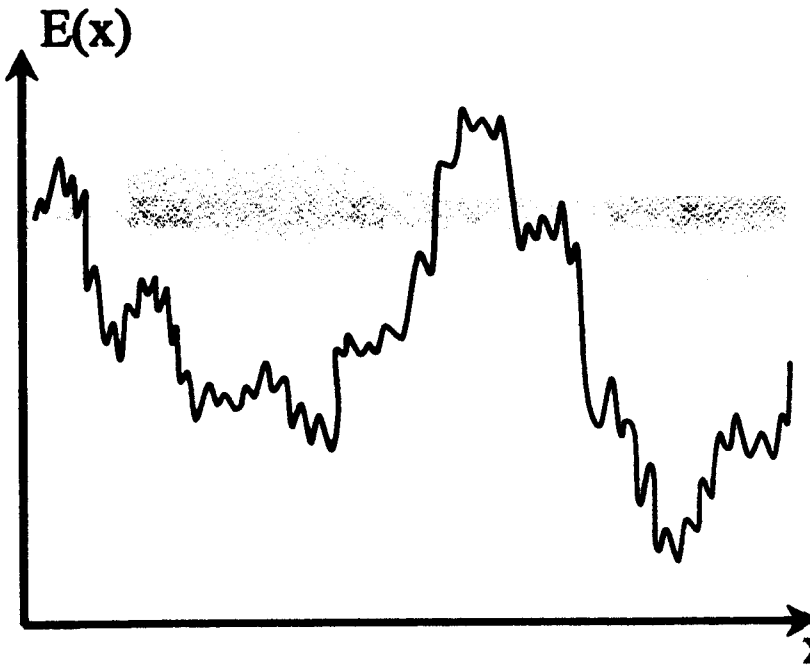
Steps of small-scale improvements
e.g., point-by-point minimization

- Slow convergence of large-scale features
  (e.g., smooth components)

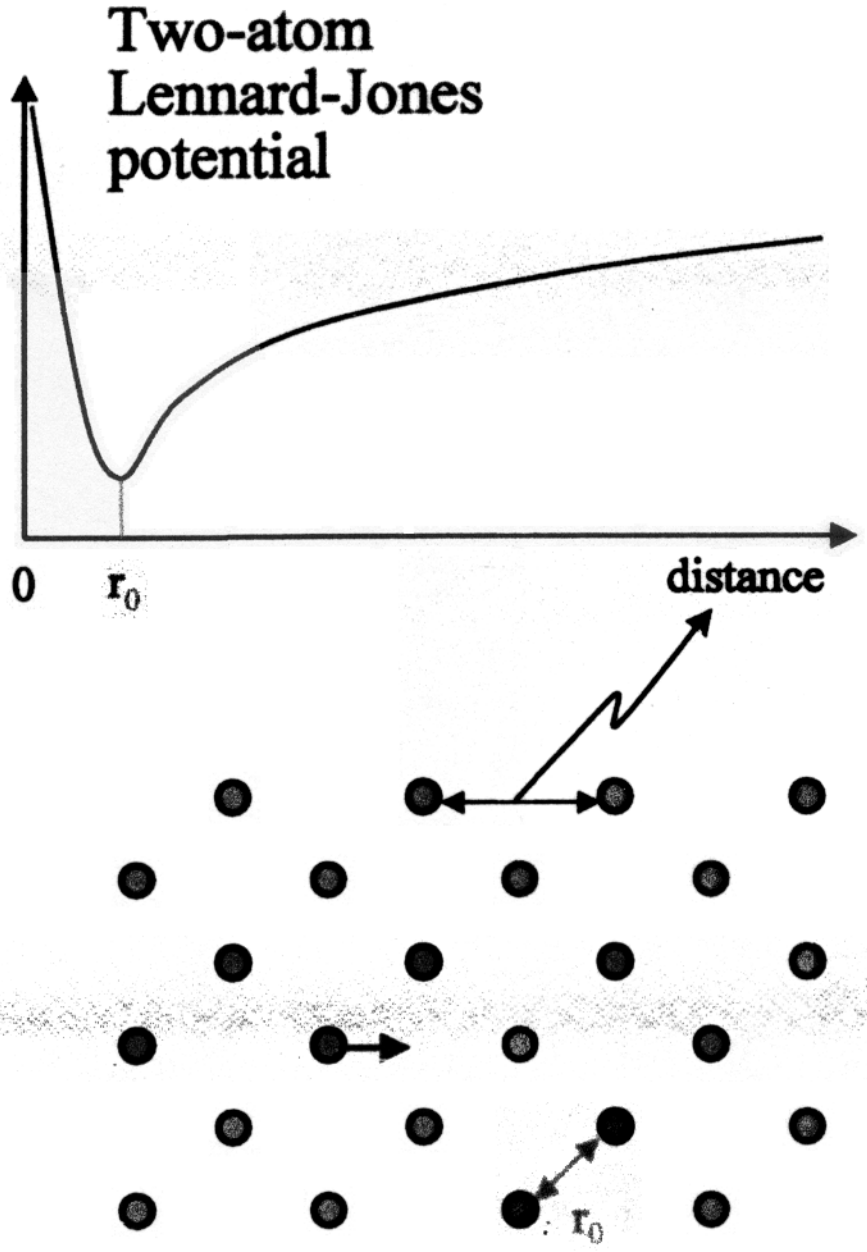
- Local-minimum traps

  Simulated annealing: $P(x) \quad e^{-E(x)/T}$

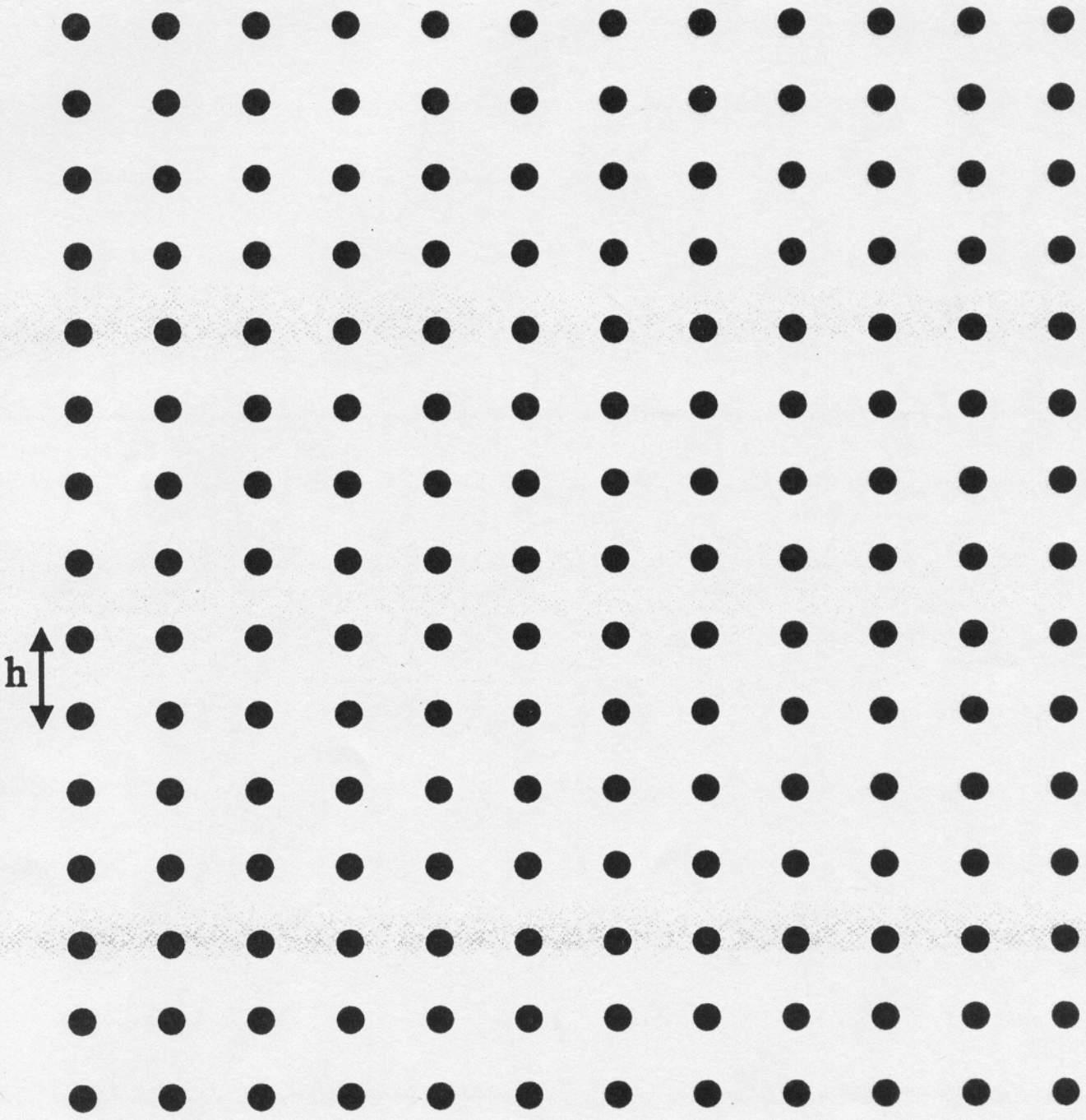
  - Slow equilibration of large-scale features
  - Nested multiscale attraction basins

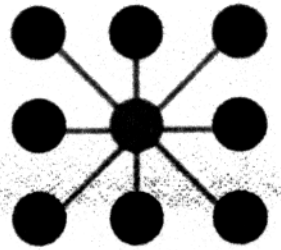


Required: Multiscale steps


Two-atom
Lennard-Jones
potential
0
r_0
distance
small
step
r_0

fine grid





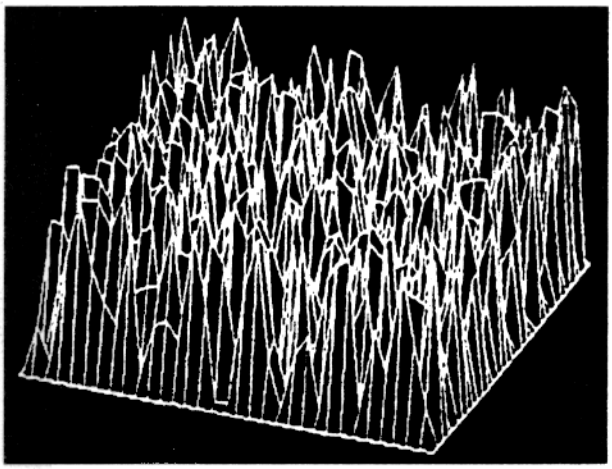
Local Equation

e.g., $u_\bullet$ = average of $u_\bullet$'s + $f_\bullet$ (known)

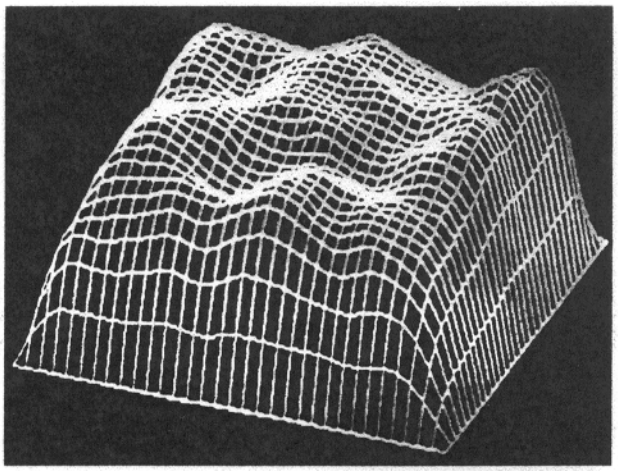
approximating Poisson eq. $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = F$

Influence of (pointwise) Gauss-Seidel relaxation on the error

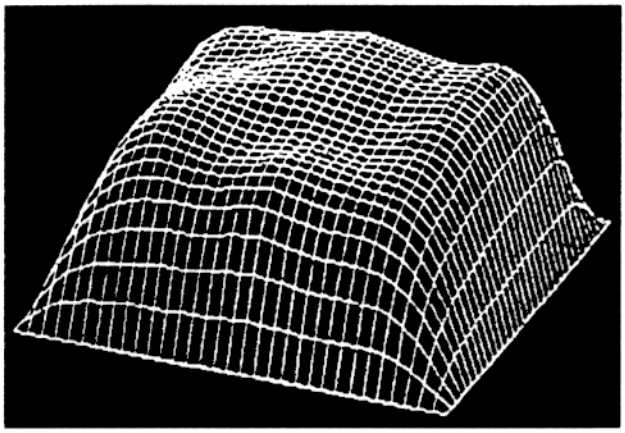
Poisson equation, *uniform grid*



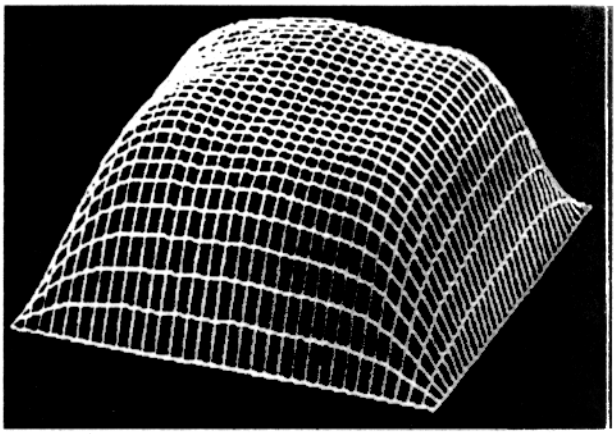
Error of initial guess



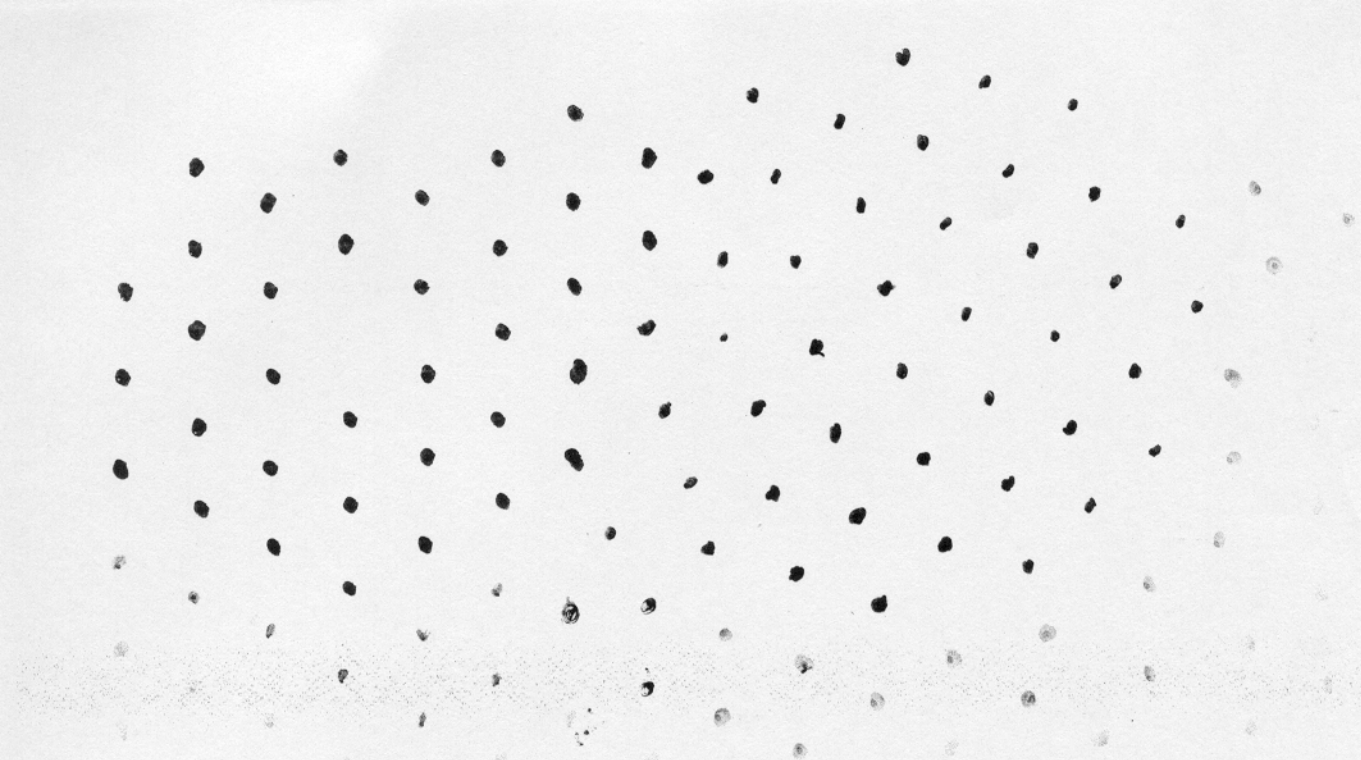
Error after 5 relaxations



Error after 10 relaxations



Error after 15 relaxations



Linear System $\underset{n \times n}{A}x = b,$

Minimize $\frac{1}{2}x^T A\, x - b^T x$

Relaxation: Fast convergence of most eigenmodes.

Slows when $\|A\, e\| \ll \|A\|\,\|e\|$, $e = x - x_{current}$

$r = b - Ax_{current}$

Remaining Problem: $Ae = r$, $\underset{n \times m}{e} \approx \uparrow x^c$ ($\uparrow$: $m \times n$, x^c: m)

typically $m < n/2$

Interpolation
Collective motions

Coarse unknowns

$\Rightarrow$ Minimize $\frac{1}{2}(\uparrow x^c)^T A\, (\uparrow x^c) - r^T (\uparrow x^c)$

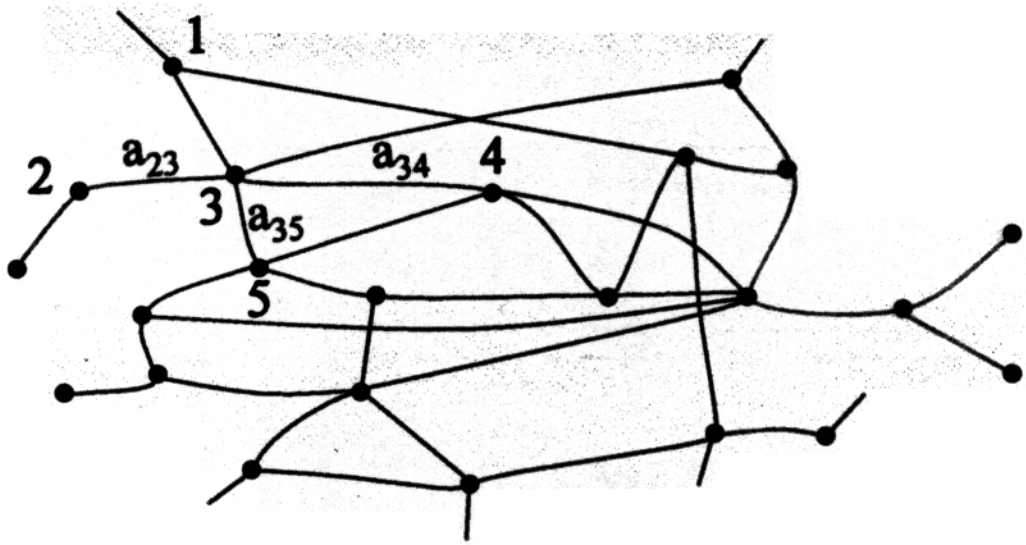
$A^c x^c = b^c$, $A^c = \uparrow^T A \uparrow$, $b^c = \uparrow^T r$

Required: Good $\uparrow$

Classical ALGEBRAIC MULTIGRID (AMG) 1982

$Ax = b$, "M matrix"

$$a_{ii}\, x_i = \sum_{j \neq i} a_{ij}\, x_j + b_i, \qquad a_{ij} = a_{ji} \geq 0 \;\; (i = 1,\ldots,n)$$

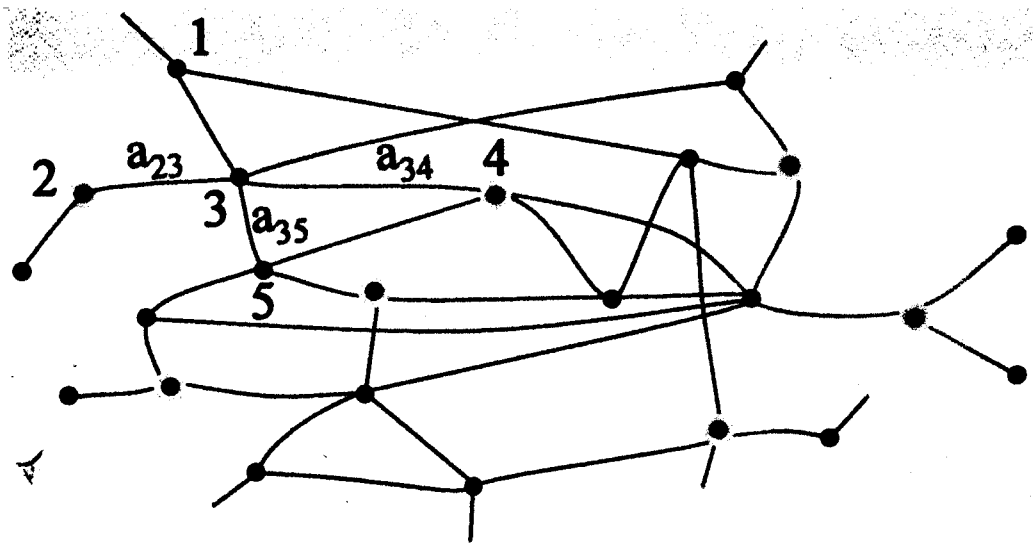


- Relaxation $\Rightarrow$ approximation $\tilde{x}$

Classical ALGEBRAIC MULTIGRID (AMG) 1982

$Ax = b$, "M matrix"

$$a_{ii} x_i = \sum_{j \neq i} a_{ij} x_j + b_i, \qquad a_{ij} = a_{ji} \geq 0 \quad (i = 1,\ldots,n)$$



Coarse variables - a subset

$$(\uparrow x^c)_3 = (a_{23} x_2 + a_{34} x_4)/(a_{23} + a_{34})$$

AMG Cycle

- Relaxation $\Rightarrow$ approximation $\tilde{x}$
- Solve $A^c x^c = \uparrow^T (b - A\tilde{x})$ (recursion)
- $\tilde{x} \leftarrow \tilde{x} + \uparrow x^c$

Graph problems

Partition: min cut

Clustering

Image segmentation

Routing

Linear arrangement: bandwidth, cutwidth

Graph drawing

⋮

Coarsening: weighted aggregation

Recursion: inherited couplings (like AMG)

Modified by properties of coarse aggregates

General principle: **Multilevel objectives**

Beyond M matrices

Weighted Graphs

Required interpolation ↑ :

For every y such that $\|Ay\| << \|A\|\|y\|$

there exists y^c such that $\|y - \uparrow y^c\| \le \dfrac{\|Ay\|}{\|A\|}$

Create ↑ by best fit to a set

of "test vectors" $x^{(1)}, x^{(2)}, \ldots, x^{(m)}$

with small $\|A\, x^{(k)}\| \,/\, \|x^{(k)}\|$

$\forall\, i$ minimize $\displaystyle\sum_{k=1}^{m} \|A\, x^{(k)}\|^{-2} \cdot \|x_i^{(k)} - \sum_j x_j^{(k)c}\|^2$

Bootstrap AMG (BAMG) 2000

Obtaining $x^{(k)}$ with small $\|A x^{(k)}\| / \|x^{(k)}\|$

by applying the evolving AMG solver itself

to the equation $Ax = 0$

First relaxation only...

Then multilevel cycling...

Ever decreasing $\|A x^{(k)}\| / \|x^{(k)}\|$

- Fast-self-correcting coarsening system
- Enables minimizing # interpolation points
  $\Rightarrow$ maximizing sparseness of coarse levels.
- Generality, fast solution

Multilevel Monte Carlo $P(x) \sim e^{-E(x)/T}$

Coarse movements x^c: $P(x^c) \sim e^{-E(x^c)/T}$

Designing ↑ (i.e., shapes of coarse moves, (probability-weighted) aggregates, etc.)

through the simulation itself:

First simple MC…

Then multilevel cycling…

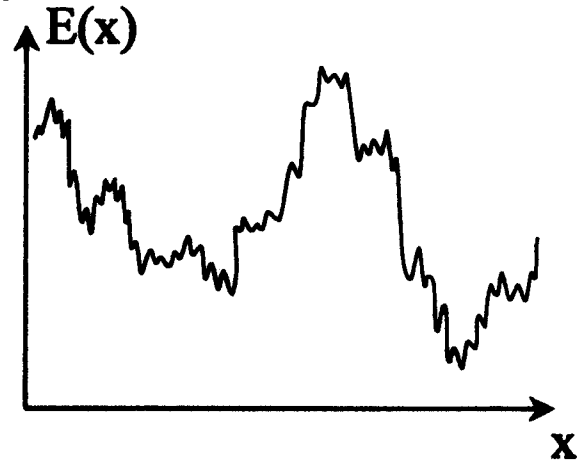
Ever decreasing I

- $E(\uparrow x^c)$ modified by aggregate properties

Multilevel

RE-SHAPING

of moves



A large-scale move is decided

only after "reshaping":

Optimizing around it at all finer scales

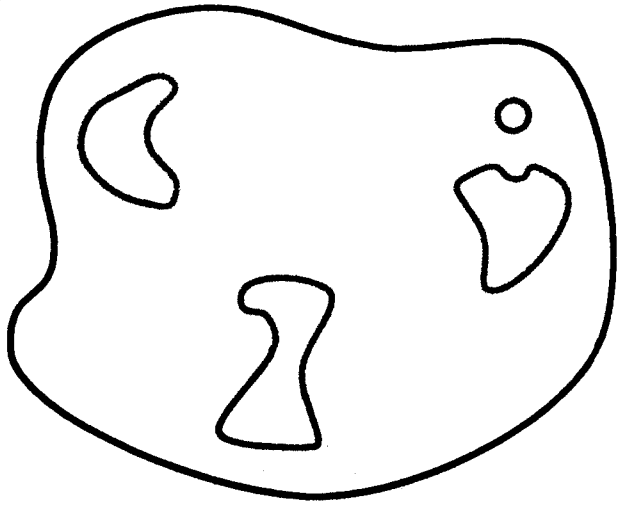
- Starting at finest
- Recursion
- Brief annealing at each scale
- Top parallel-processing efficiency

Fast search over large-scale basins

Even with "stupid" moves: Spin glasses, L-J solids

Very large system
at $T > 0 \Rightarrow$

High probability for some
small-scale moves
away from optimum



$\Downarrow$

Recombinations

with stored low-energy configurations

- In reshaping processes, too

Population algorithms

Multiple aggregation rules

(cf. indefinite BAMG)

Multiscale Optimization Strategies

- **Relaxation**
- **Hierarchical** (weighted) **aggregation**
- **Multiscale objectives:** properties of intermediate aggregates determine further aggregations
- **Bootstrap:** (weighted) aggregates formed by the experience of the hierarchical solver itself
- **Annealing:** Monte-Carlo with decreasing T, identifying increasingly lower moves
- **Reshaping** of large-scale moves at all finer scales
- **Recombinations** and population algorithms