UNIVERSITY OF MINNESOTA TWIN CITIES

Relaxed gradient-type descent methods

Yousef Saad

University of Minnesota

ICERM - Providence, RI

May 7, 2026

First

- Joint work with: Jean-Paul Chehab (Amiens, Fr) Gaspard Kemlin (Amiens, Fr) & Marcos Raydan (NovaMath, Pt)
- Work supported by NSF
- Preprint: “Eigenvector-based acceleration strategies for gradient-type methods”, J.-P. Chehab, G. Kemlin & M. Raydan. arXiv:2601.11145 [math.NA]

University of Minnesota

Introduction

Problem: $\min_{x \in \mathbb{R}^n} f(x)$ $\quad f : \mathbb{R}^n \rightarrow \mathbb{R}$ continuously differentiable

Gradient descent method: $x_{k+1} = x_k - \alpha_k g_k$ where $g_k = \nabla f(x_k)$

$\alpha_k ==$ steplength

- Cauchy's steepest descent (SD): $\alpha_k^{\text{SD}} := \arg\min_\alpha f(x_k - \alpha g_k)$
- Minimal residual descent (MR): $\alpha_k^{\text{MR}} := \arg\min_\alpha \|\nabla f(x_k - \alpha g_k)\|_2$
- Observation: poor practical behavior of iteration ('zigzag') – in both cases.

University of Minnesota

Break the zig-zag

- *Akaike* ['59]: Asymptotically search direction alternates within the 2-dimens. subspace spanned by largest, smallest eigenvectors.
- Idea: break the zig-zag by adding randomness
- *Raydan* and *Svaiter* ['02]: convergence can be improved by randomly relaxing α_k by $\sigma \in (0, 2)$
- See also a recent paper by *L. MacDonald, R. Murray, and R. Tappenden* ['25]

University of Minnesota

The quadratic case

Consider: $f(x) = \frac{1}{2}x^\top Ax - b^\top x$ $\quad b \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times n}$, A is SPD

- *Gradient:* $g(x) \equiv \nabla f(x) = Ax - b$, *Global minimizer of* f: $x^* = A^{-1}b$

Two Gradient Descent methods:

Steepest Descent:

$$\alpha_k^{\text{SD}} = \frac{g_k^\top g_k}{g_k^\top A g_k} \quad \text{(SD)}$$

Minimal Residual:

$$\alpha_k^{\text{MR}} = \frac{g_k^\top A g_k}{g_k^\top A^2 g_k} \quad \text{(MR)}$$

- Relaxed gradient-type method: $x_{k+1} = x_k - \sigma \alpha_k g_k$

where: $\alpha_k \equiv \alpha_k^{\text{SD}}$ or α_k^{MR}, $\quad \sigma =$ relaxation param. $\neq 1$

University of Minnesota

Relation with shifted power method

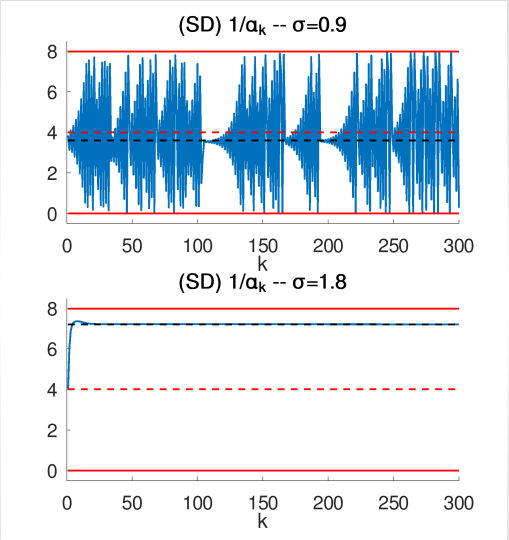
Notice that:

$$g_{k+1} = g_k - \sigma\alpha_k A g_k$$
$$= (I - \sigma\alpha_k A) g_k \quad \text{so:} \rightarrow$$

$$g_{k+1} = \left(\prod_{i=0}^{k} (I - \sigma\alpha_i A) \right) g_0$$

- When $\sigma = 1$: asymptotically, gradients will tend to oscillate in a subspace of dim. 2. [largest, smallest eigenvectors]
- What happens when $\sigma \neq 1$?
- Experiment: a 900×900 matrix - discretization of a Laplacean on 30×30 grid.
- Notation: $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ = eigenvalues of A $U = [u_1, u_2, \cdots, u_n]$ = matrix of unit eigenvectors.

University of Minnesota



Rayleigh quotients $(1/\alpha_k)$ for $\sigma = 0.9$ (top) and $\sigma = 1.8$ (bottom), for SD.

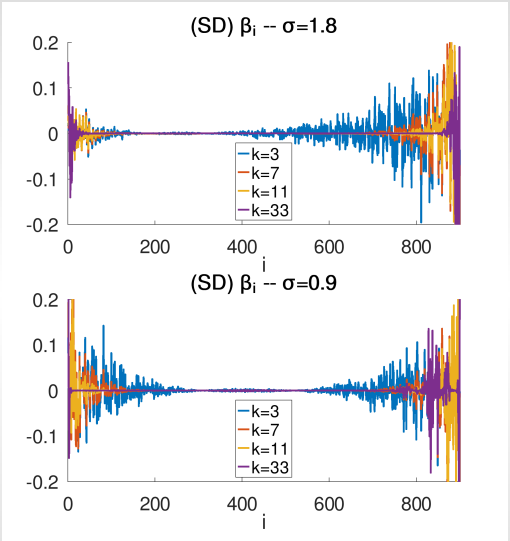
Top and bottom lines = λ_n, λ_1.

Dashed black line = $\sigma(\lambda_1 + \lambda_n)/2$.

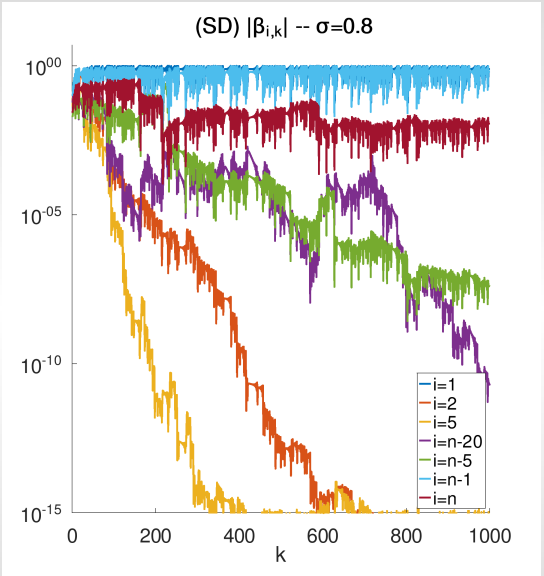
Dashed red line = $(\lambda_1 + \lambda_n)/2$.

Results with MR are very similar

University of Minnesota



Components of the normalized residuals in the eigenbasis for iterations $k = 3, 7, 11, 33$ with the MR (left) and SD (right) algorithms using $\sigma = 1.8$ (top) or $\sigma = 0.9$ (bottom).



Convergence of $|\beta_{i,k}|$ for different i's and $\sigma = 0.8$ in the case of SD.

- Residuals supported by a few extremal modes: mode $i = 1$ and a few of the highest modes – i between n and $n - 5$ – All intermediate modes vanish asymptotically.

University of Minnesota

Spectral acceleration

Eigenvector acceleration techniques

- Idea: exploit the observation that g_k becomes close to an eigenvector
- A simple lemma for motivation. Assumption $f(x) = \frac{1}{2}x^\top Ax - b^\top x$, A SPD.

Lemma

If the scheme $x_{k+1} = x_k - \sigma\alpha_k g_k$ is used to minimize $f(x)$ with $\sigma \in (0, 2)$, α_k obtained via (SD) or (MR), and at some iteration $k \geq 1$ the gradient vector g_k, is an eigenvector of A ($Ag_k = \lambda g_k$) then setting $\sigma = 1$ at iteration k results in: $x_{k+1} = x^* = A^{-1}b$.

- If g_k 'stumbles' on an exact eigenvector at step k then $x_{k+1} = x_*$

University of Minnesota

Proposition (Inexact eigenvector case)

Assume that the relaxed GD iteration with either steplength (SD) or (MR) is applied at step k and let ρ_k be the residual of the vector g_k considered as an approximate eigenvector of A:

$$\rho_k = \frac{(A - \alpha_k^{-1} I)g_k}{\|g_k\|}.$$

Then the gradient of the next iterate satisfies the equality:

$$\|g_{k+1}\| = |\alpha_k| \times \|\rho_k\| \times \|g_k\|$$

in which α_k stands for either $\alpha_k^{(SD)}$ or $\alpha_k^{(MR)}$.

University of Minnesota

Lemma

From one step to the next the residual norms in MR are related as follows when $\sigma = 1$:

$$\frac{\|g_{k+1}\|^2}{\|g_k\|^2} = \sin^2 \angle(g_k, Ag_k)$$

From one step to the next the A^{-1} residual norms in SD are related as follows when $\sigma = 1$:

$$\frac{\|g_{k+1}\|_{A^{-1}}^2}{\|g_k\|_{A^{-1}}^2} = \sin^2 \angle_{A^{-1}}(g_k, Ag_k)$$

University of Minnesota

Proposition

Let g_k be an approximate eigenvector such that $Ag_k = \mu_k g_k + w_k$ where w_k is orthogonal to g_k for MR, and A^{-1}-orthogonal to g_k for SD. If we assume that

$$\text{MR:} \quad \frac{\|w_k\|}{\|g_k\|} \leq \epsilon, \qquad \text{SD:} \quad \frac{\|w_k\|_{A^{-1}}}{\|g_k\|_{A^{-1}}} \leq \epsilon$$

then the pair of vectors g_k and Ag_k satisfy:

$$\text{MR: } \sin \angle(g_k, Ag_k) \leq \frac{\epsilon}{\sqrt{\mu_k^2 + \epsilon^2}} \qquad \text{SD: } \sin \angle_{A^{-1}}(g_k, Ag_k) \leq \frac{\epsilon}{\sqrt{\mu_k^2 + \epsilon^2}}$$

University of Minnesota

- Example: $\epsilon \leq \mu_k/\sqrt{3}$ $\rightarrow$ reduction by a factor 2 in residual norm
- Idea: Monitor eigenvector-residual $\frac{\alpha_k}{\|g_k\|}\left\|Ag_k - \alpha_k^{-1} g_k\right\|$ at each step
- When residual is close to an eigenvector $\rightarrow$ set $\sigma = 1$
- Otherwise set $\sigma \neq 1$ (fixed)

Algorithm: Eigenvector Acceleration Scheme

1: Start: initial guess x_0; set $r_0 = b - Ax_0$, $p_0 = Ar_0$;

2: Choose: σ with $0 < \sigma < 1$, and $0 < \epsilon_{\text{eig}} < 1$

3: **while** (not converged) **do**

4: $\quad \alpha_k = \dfrac{p_k^\top r_k}{p_k^\top p_k}$

5: $\quad$ **if** $\left(\frac{\alpha_k}{\|r_k\|} \, \|p_k - \alpha_k^{-1} r_k\| < \epsilon_{\text{eig}} \right)$ **then** $\boxed{\sigma_k = 1}$ **else** $\boxed{\sigma_k = \sigma}$ **end if**

6: $\quad x_{k+1} = x_k + \sigma_k \alpha_k r_k$

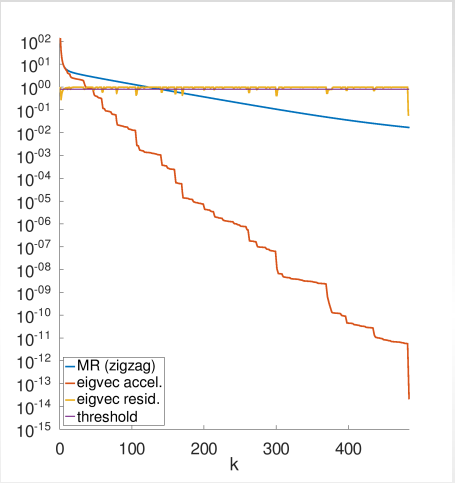
7: $\quad r_{k+1} = r_k - \sigma_k \alpha_k p_k$

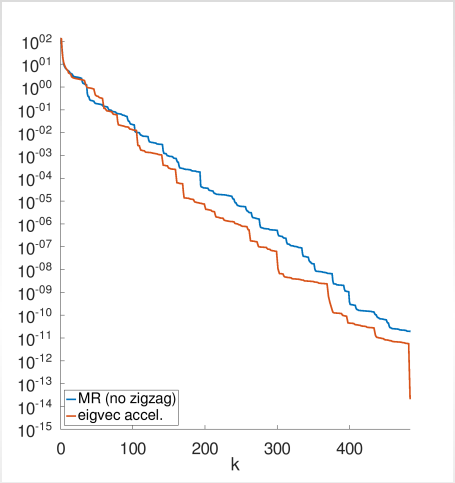
8: $\quad p_{k+1} = Ar_{k+1}$

9: **end while**

$\sigma = 0.8$

$\epsilon_{\text{eig}} = 0.8$





Left: blue line == standard MR (zigzag). *Right:* blue line == MR with σ_k = constant.

Compare eigenv. accel. vs 'no-zigzag' (right).

University of Minnesota

Lanczos-based Acceleration

- Push previous idea a bit further: use a subspace instead of single vector

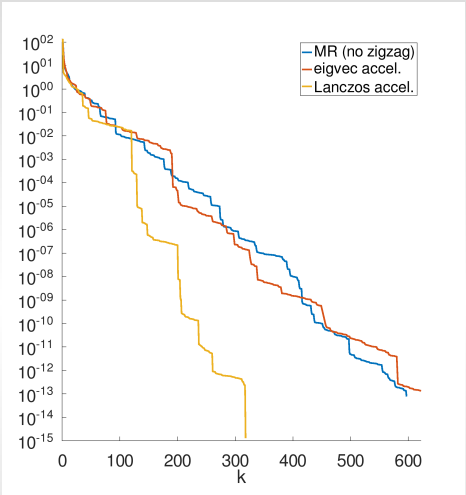
Procedure:

1. Use the same detection criterion as before: $\frac{\alpha_k}{\|r_k\|} \|Ar_k - \alpha_k^{-1} r_k\| < \epsilon_{\text{eig}}$
2. If satisfied, start a short Lanczos iteration starting with r_k.
3. Result: $V_m, T_m = V_m^\top A V_m$
4. Compute: $x_{k+1} = x_k + V_m y_m$ with $y_m = \text{argmin}_{z \in \mathbb{R}^m} \|r_k - A V_m z\|^2$

University of Minnesota

Algorithm: Lanczos Based Acceleration (LBA)

1: Start: initial guess x_0; set $r_0 = b - Ax_0$ and $p_0 = Ar_0$;
2: **while** (not converged) **do**
3: $\alpha_k = p_k^\top r_k / p_k^\top p_k$
4: **if** $\left(\frac{\alpha_k}{\|r_k\|} \, \|p_k - \alpha_k^{-1} r_k\| < \epsilon_{\text{eig}} \right)$ **then**
5: $V_m, T_m = \textbf{Lanczos}(A, r_k, m)$
6: Solve $y_m = \textbf{argmin}_{z \in \mathbb{R}^m} \|r_k - AV_m z\|$
7: $x_{k+1} = x_k + V_m y_m$; $\quad r_{k+1} = r_k - AV_m y_m$
8: **else**
9: $x_{k+1} = x_k + \sigma \alpha_k r_k$; $\quad r_{k+1} = r_k - \sigma \alpha_k p_k$
10: **end if**
11: $p_k = Ar_k$
12: **end while**



	mat-vec's
MR (no zigzag)	597
Eigenvector accel.	622
Lanczos-based accel.	396

Comparison of total Mat-vec's

Lanczos based acceleration ($\sigma = 0.8$, $\epsilon_{\text{eig}} = 0.8$); Lanczos with $m = 5$.

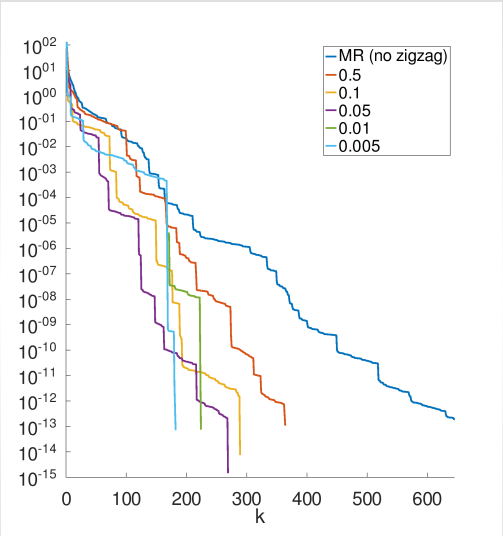
University of Minnesota

- Goal: set number of Lanczos steps dynamically

Algorithm: Adaptive Lanczos

1: Start: initial guess r; matrix A; relative tolerance `reltol`
2: **for** $i = 1 \ldots m$ **do**
3: perform i-th step of Lanczos: $\rightarrow V_i, H_i$
4: Solve $y_i = \mathbf{argmin}_{z \in \mathbb{R}^i} \|r - AV_i z\|$
5: **if** $\|r - AV_i y_i\| \leq$ `reltol` $\times \|r\|$ **then**
6: break
7: **end if**
8: **end for**

- Note: Givens rotations exploited to perform test in Line 5.



adaptive			
reltol	its	mat- -vecs	Lanczos calls
.5	363	425	21
.1	288	345	10
.05	268	361	11
.01	223	276	*7*
.005	*181*	*264*	8

Subspace dim. in Lanczos: $\boldsymbol{m = 10}$.

Perf. for different `reltol`'s

Nonquadratic case

Nonquadratic case

- So far: quadratic case only. But: in this case → Conjugate Gradient is best
- Goal: extend strategies seen for general nonlinear case
- Residual r_k becomes $\nabla f(x_k)$
- We will use MR – so $$\alpha_k = \frac{g_k^\top H_k g_k}{g_k^\top H_k^2 g_k} = \frac{g_k^\top (H_k g_k)}{(H_k g_k)^\top (H_k g_k)}$$ (One 'mat-vec')
- $H_k g_k$ computed from Frechet diff. $$H_k g_k \approx \frac{\nabla f(x_k + h g_k) - g_k}{h}$$ where $h > 0$ is a small parameter

University of Minnesota

Global convergence strategy

- Extensions of the MR method with globalization technique - in convex case
- Large Scale problems: use 'undemanding' line search (LS) globalization
- Quadratic case: α_k obtained from exact line search (closed form formula)
- In non-quadratic case $\rightarrow$ inexact line search: adapted from *W. La Cruz and G. Noguera '09*
- Steplength t_k must satisfy: $f(x_k + t_k\, d_k) \leq f(x_k) - \gamma t_k^2 \|g_k\|^2 + \eta_k,$ where $\gamma >$ is a small scalar, $\eta_k > 0$ is s.t. $\sum_{k\in\mathbb{N}} \eta_k \leq \eta < \infty$
- Search dir. d_k is either $-g_k$ or $V_m y_m$ with $t_k = 1$ (Lanczos update)

University of Minnesota

Algorithm: Line search (backtracking)

1: Input: $\sigma \in (0,1)$, $0 < \gamma < 1$, x_k, α_k, g_k, d_k, and $\eta_k > 0$

2: Set $\widehat{d}_k = \sigma\alpha_k d_k$, $x_+ = x_k + \widehat{d}_k$, $\delta = g_k^\top \widehat{d}_k$, and $\beta = 1$

3: **while** $(f(x_+) > f(x_k) - \gamma(\beta\sigma\alpha_k)^2\|g_k\|^2 + \eta_k)$ **do**

4: $\quad \beta_{\text{temp}} = -\frac{1}{2}\beta^2\delta/(f(x_+) - f(x_k) - \beta\delta)$

5: $\quad$ **if** $\beta_{\text{temp}} \in [\beta\sigma_1, \beta\sigma_2]$ **then** $\beta := \beta_{\text{temp}}$ **else** $\beta := \frac{\beta}{2}$ **endif**

6: $\quad x_+ = x_k + \beta\widehat{d}_k$

7: **end while**

8: Output: $\sigma\alpha_k \leftarrow \beta\sigma\alpha_k$ and $x_{k+1} \leftarrow x_+$

University of Minnesota

Numerical examples

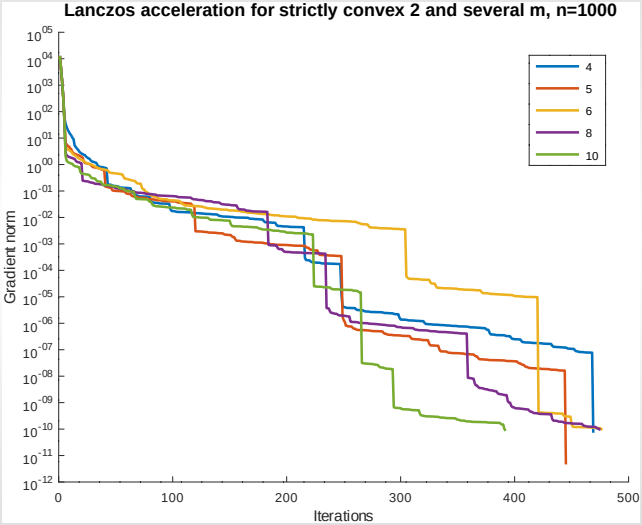
(a) *(Strictly convex 2)* $x \in \mathbb{R}^n \mapsto f(x) = \sum_{i=1}^{n} \frac{i}{10}\left(\exp(x_i) - x_i\right);$

(b) *(Logistic loss)* $x \in \mathbb{R}^n \mapsto f(x) = \frac{\kappa}{2}\|x\|^2 + \sum_{i=1}^{p} \log\left(1 + \exp(-(x^\top z_i) y_i)\right)$

- Initial guess for (a) = componenwise random in the interval $[0, 3]$
- For (b) initial guess is = vector of all ones; $z_i \in \mathbb{R}^n$ are random vectors; y_i are random ± 1, $\kappa = 0.1$
- The h in Frechet derivative is set to: $h = \frac{10^{-5}}{\min\{1, \max\{10^{-3}, 10^5 \|g_k\|\}\}}$

University of Minnesota

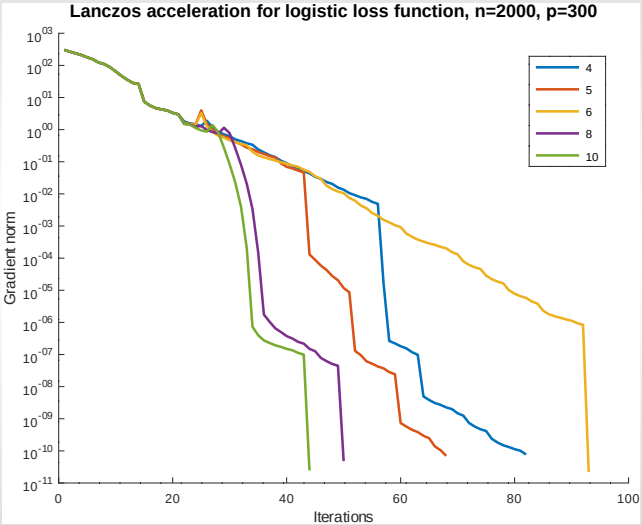
Varying m – problem (a) $n = 1000$



m	its.	Grad Evals	Calls to Lanczos
4	468	973	9
5	444	934	9
6	476	995	7
8	474	1021	9
10	391	863	8

University of Minnesota

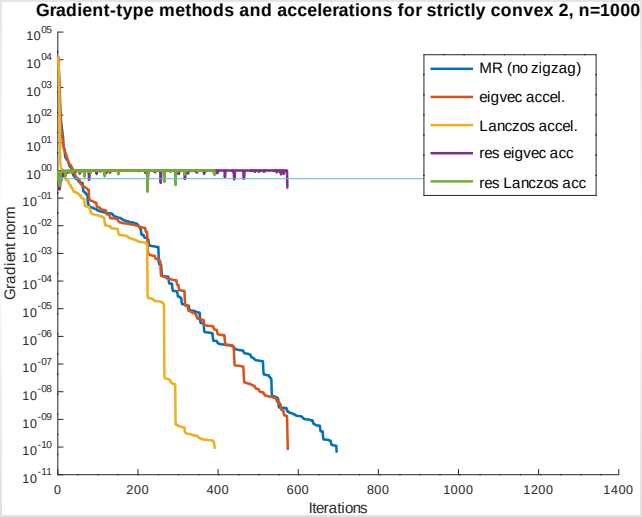
Varying m – problem (b) $n = 2,000, p = 300$



m	its.	Grad Evals	Calls to Lanczos
4	81	187	6
5	67	170	7
6	92	215	5
8	49	187	11
10	43	197	11

University of Minnesota

Problem (a) $n = 1,000$, Fixed $m = 5$ for Lanczos

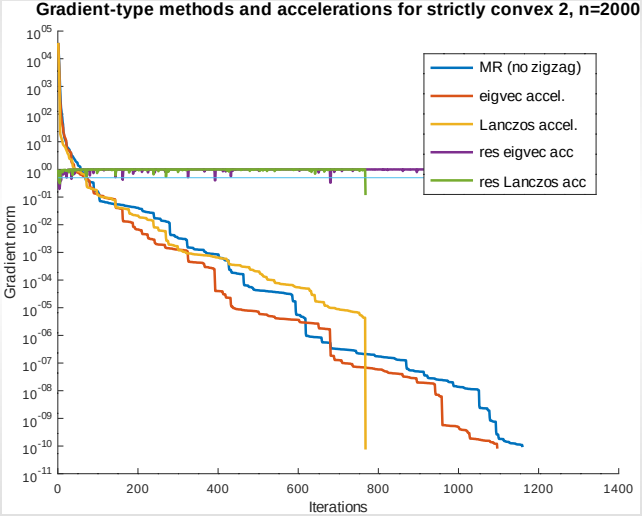


Accel. scheme	its.	Grad EvalS
MR (no zigzag)	695	1391
Eigenvector	573	1147
Lanczos (9)	444	934

Note: Top curves = eigenvector residuals

University of Minnesota

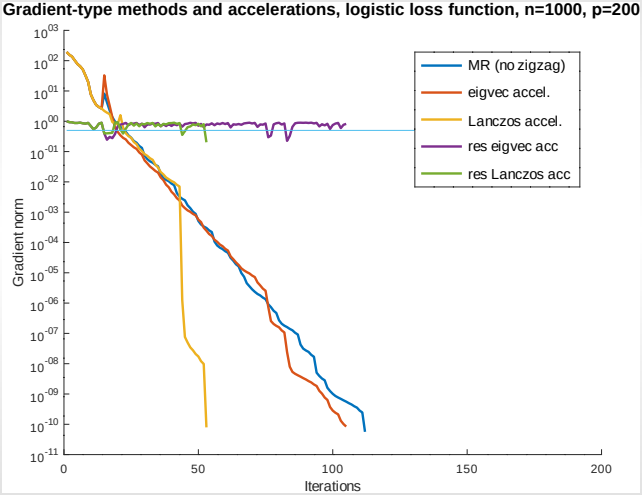
Problem (a) $n = 2,000$, Fixed $m = 5$ for Lanczos



Accel. scheme	its.	Grad EvalS
MR (no zigzag)	1159	2319
Eigenvector	1096	2193
Lanczos (6)	767	1565

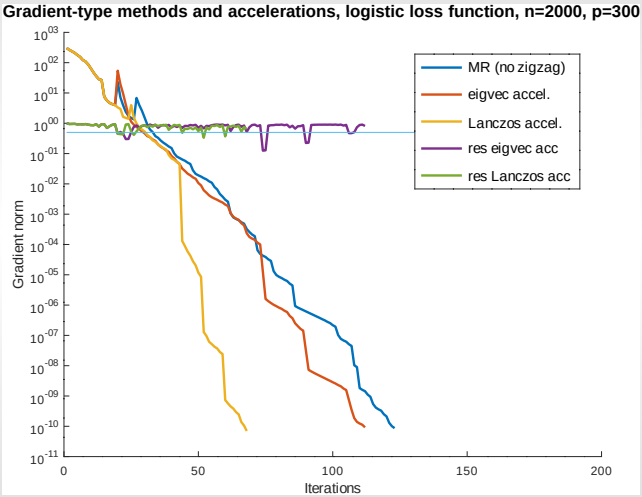
University of Minnesota

Problem (b) $n = 1,000, p = 200$, Fixed $m = 5$ for Lanczos



Acceleration	its.	Grad Eval
MR (no zigzag)	111	223
Eigenvector	104	209
Lanczos (8)	52	145

Problem (b) $n = 2,000, p = 300$, Fixed $m = 5$ for Lanczos



Acceleration	its.	Grad Eval
MR (no zigzag)	122	245
Eigenvector	111	223
Lanczos (7)	67	170

Concluding remarks

1. Fascinating behavior of relaxed GD methods ...
2. ... which can be nicely exploited
3. Main point: for certain σ's, residual tends to lie in a small dimensional eigenspace. Projection method (Lanczos) perfect for this case
4. To do: Can this be helpful in Machine Learning? [in theory and in practice]
5. To do: Fully analyze behavior

University of Minnesota

Concluding remarks

1. Fascinating behavior of relaxed GD methods ...
2. ... which can be nicely exploited
3. Main point: for certain σ's, residual tends to lie in a small dimensional eigenspace. Projection method (Lanczos) perfect for this case
4. To do: Can this be helpful in Machine Learning? [in theory and in practice]
5. To do: Fully analyze behavior

Thank you!

University of Minnesota

If time permits: Analysis

Analysis of relaxed GD in quadratic case

- Recall that $r_{k+1} = (I - \sigma\alpha_k A)r_k \quad \rightarrow \quad r_{k+1} = \left(\prod_{i=0}^{k}(I - \sigma\alpha_i A)\right) r_0$
- Observed: residuals concentrates on a few extreme eigenvectors.
- Define: $\mu(A, v) \equiv \dfrac{v^\top A v}{v^\top v}, \qquad \nu(A, v) \equiv \dfrac{v^\top A^2 v}{v^\top A v},$
- Note: $\alpha_k^{(SD)} = \dfrac{1}{\mu(A, r_k)}, \qquad \alpha_k^{(MR)} = \dfrac{1}{\nu(A, r_k)}$
- Denote the components of r_k in the eigenbasis of A by $\beta_{1,k}, \ldots, \beta_{n,k}$
- Normalize by A-norm for MR, 2-norm for SD:

$$\text{(SD)} : \sum_{i=1}^{n} \beta_{i,k}^2 = 1, \quad \text{(MR)} : \sum_{i=1}^{n} \beta_{i,k}^2 \lambda_i = 1.$$

University of Minnesota

- Rayleigh quotients:

$$\text{(SD)} : \mu(A, r_k) = \sum_{i=1}^{n} \beta_{i,k}^2 \lambda_i; \qquad \text{(MR)} : \nu(A, r_k) = \sum_{i=1}^{n} \beta_{i,k}^2 \lambda_i^2.$$

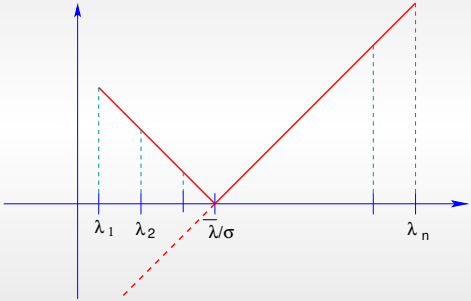
- These are convex combinations of eigenvalues $\rightarrow$ Both $\in [\lambda_1, \lambda_n]$.
- We will denote by $\bar{\lambda}_{\beta,k}$ either $\mu(A, r_k)$ (SD) or $\nu(A, r_k)$ (MR)
- Rewrite $r_{k+1} = (I - \sigma\alpha_k A) r_k$ as $r_{k+1} = \frac{1}{s_k}\left(\frac{1}{\sigma\alpha_k} I - A\right) r_k$ where $s_k =$ scaling
- Note: $1/\alpha_k == \bar{\lambda}_{\beta,k} \rightarrow$ Define $\xi_k \equiv \frac{\bar{\lambda}_{\beta,k}}{\sigma} \rightarrow$

$$r_{k+1} = \frac{1}{s_k}(\xi_k I - A) r_k$$

- Apart from scaling : shifted power method. Shift ξ_k changes at each step.
- Components of new residual: ($s_k \equiv$ normalization factor)

$$\beta_{i,k+1} = \frac{1}{s_k} (\xi_k - \lambda_i) \beta_{i,k}, \quad 1 \le i \le n$$

- Magnitude of each eigen-component of r_k 'amplified' by $\frac{1}{s_k}|\xi_k - \lambda_i|$.
- Recall: $\xi_k \equiv \frac{\bar{\lambda}_{\beta,k}}{\sigma}$ and $\xi_k \in [\lambda_1/\sigma, \lambda_n/\sigma]$



- Amplification factors for relaxed SD / MR scheme
- Largest factor : associated with either λ_1 or with λ_n.
- Intuition: if the ξ_k's are well distributed $\Rightarrow$ for $\lambda_i \neq 1, n$, $\beta_{i,k} \rightarrow 0$

Lemma:

Let $(\beta_{i,k})_{k\in\mathbb{N}}$ be the i-th component of r_k. Assume that $\beta_{1,0}\beta_{n,0} \neq 0$. Select ξ_k from interval $(\frac{\lambda_1}{\sigma}, \frac{\lambda_n}{\sigma})$, with $\sigma \in (0,2)$, at each step k. Then, for $i = 1, \ldots, n$

$$\begin{cases} \limsup_{k\to+\infty} \left(\prod_{j=0}^{k} \frac{|\xi_j - \lambda_i|}{|\xi_j - \lambda_1|} \right)^{1/k} < 1 \quad \Rightarrow \quad \lim_{k\to+\infty} \frac{|\beta_{i,k}|}{|\beta_{1,k}|} = 0 \\ \limsup_{k\to+\infty} \left(\prod_{j=0}^{k} \frac{|\xi_j - \lambda_i|}{|\xi_j - \lambda_n|} \right)^{1/k} < 1 \quad \Rightarrow \quad \lim_{k\to+\infty} \frac{|\beta_{i,k}|}{|\beta_{n,k}|} = 0. \end{cases}$$

- Note: the shifts are not necessarily those of the SD/MR schemes

University of Minnesota

Lemma:

Let $(\xi_k)_{k\in\mathbb{N}}$ be a sequence of independent random variables, uniformly distributed in $(\frac{\lambda_1}{\sigma}, \frac{\lambda_n}{\sigma})$ for $\sigma \in (0, 2)$.

$$\sigma < 1 \Rightarrow \forall\, i \neq 1\colon \limsup_{k\to+\infty}\left(\prod_{j=0}^{k}\frac{|\xi_j-\lambda_i|}{|\xi_j-\lambda_1|}\right)^{1/k} < 1 \Rightarrow \lim_{k\to\infty}|\beta_{i,k}| = 0$$

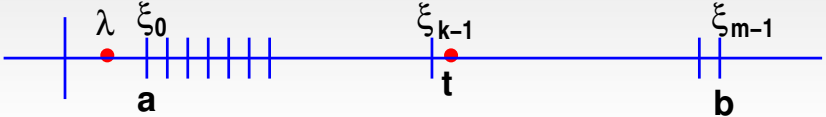
$$\sigma > 1 \Rightarrow \forall\, i \neq n\colon \limsup_{k\to+\infty}\left(\prod_{j=0}^{k}\frac{|\xi_j-\lambda_i|}{|\xi_j-\lambda_n|}\right)^{1/k} < 1 \Rightarrow \lim_{k\to\infty}|\beta_{i,k}| = 0$$

$$\sigma = 1 \Rightarrow \forall i \neq 1, n : \limsup_{k\to+\infty}\left(\prod_{j=0}^{k}\frac{|\xi_j-\lambda_i|}{|\xi_j-\boxed{\lambda_{1|n}}|}\right)^{1/k} < 1 \Rightarrow \lim_{k\to\infty}|\beta_{i,k}| = 0 \text{ and}$$

$|\beta_{1,k}|$ & $|\beta_{n,k}|$ cannot both converge to 0.

A determinisc analysis

- Assume a deterministic situation where we preselect m shifts in interval $[a, b]$



$$\xi_0 = a, \quad \xi_{m-1} = b, \quad \lambda_i \in [a, b] \text{ for } i > 1, \quad \lambda \equiv \lambda_1 < a$$

- Notation: $t \in (a, b)$ – meant to represent an arbitrary λ_i
- Interested in analyzing:

$$\frac{\prod_{i=0}^{m-1} |t - \xi_i|}{\prod_{i=0}^{m-1} |\lambda - \xi_i|} \equiv \frac{|p(t)|}{|p(\lambda)|} \quad \text{with} \quad p(t) = \prod_{i=0}^{m-1} (t - \xi_i).$$

University of Minnesota

- Assumption: m shifts ξ_i uniformely distributed in $[a, b]$:

$$\xi_i = a + ih, \quad i = 0, \cdots, m-1, \quad \text{with:} \quad h \equiv (b-a)/(m-1).$$

- Notation: ξ_{k-1} = largest ξ_i to the left of t $\quad \max_{\xi_i \leq t} \xi_i = \xi_{k-1}.$ $\quad (k \geq 1)$

- Define: $\quad \delta_a = (t - \xi_{k-1})/h; \quad \delta = (a - \lambda)/h.$ $\quad$ (Note: $0 \leq \delta_a < 1$)

Theorem:

Under previous assumptions and notation the ratio of the amplification factors $p(t)$ and $p(\lambda)$ is such that:

$$\frac{|p(t)|}{|p(\lambda)|} = \frac{\Gamma(k+\delta_a)}{\Gamma(\delta_a)} \times \frac{\Gamma(m-k+1-\delta_a)}{\Gamma(1-\delta_a)} \times \frac{\Gamma(\delta)}{\Gamma(m+\delta)} \tag{1}$$

$$\approx c \frac{\left(\frac{k}{m}\right)^{\delta_a}}{(1-\frac{k}{m})^{\delta_a - 1}} \times \frac{m^{1-\delta}}{\binom{m-1}{k-1}} \tag{2}$$

where Γ represents the *Gamma* function and $c = \frac{\Gamma(\delta)}{\Gamma(\delta_a)\Gamma(1-\delta_a)}$.

Note: (1) can be defined by continuation when $\delta_a = 0$ i.e., when $t = \xi_{k-1}$: analysis $\Rightarrow$ a value of zero for the right-hand side of (1) as expected.

University of Minnesota

- Examine case when number of points $m \to \infty$ and eigenvalue $\lambda_i = t$ is fixed.
- Largest $\xi_{k-1} \leq t$ will converge to t
- $k \to \infty$ but ratio $k/m \to (t-a)/(b-a) \quad \Rightarrow$
- Product (2) behaves like a constant times $\boxed{m^{1-\delta} / \binom{m-1}{k-1}}$ and

1. $\delta \equiv (a-\lambda)/h > 1$ for m large enough – so $\Rightarrow m^{1-\delta} \to 0$
2. When $k = 1$ then $\binom{m-1}{k-1}^{-1} \equiv 1$
3. When $k > 1$ then $\lim_{m\to\infty} \binom{m-1}{k-1}^{-1} \to 0$

- Note: we can can be applied tp case where more than eigenvalue $< a$

Relaxing uniform distribution: Quasi-uniform distribution

- Can adapt theorem by relaxing the uniform distribution assumption.
- We will require the ξ_i's to obey a quasi-uniform distribution:

$$\xi_0 = a, \quad \xi_{m-1} = b \quad \text{and} \quad 0 < \underline{h} \leq \xi_i - \xi_{i-1} \leq \bar{h}, \quad i = 1, \cdots, m-1.$$

- Define $\bar{\delta}_a = (t - \xi_{k-1})/\bar{h}; \quad \underline{\delta} = (a - \lambda)/\underline{h}.$
- Then can state a similar theorem - with following asymptotic result:

$$\frac{|p(t)|}{|p(\lambda)|} \approx \tilde{c} \frac{\left(\frac{k}{m}\right)^{\bar{\delta}_a}}{(1 - \frac{k}{m})^{\bar{\delta}_a - 1}} \times \frac{m^{1-\underline{\delta}}}{\binom{m-1}{k-1}}$$

University of Minnesota

Simplified analysis of the relaxed SD iteration

- So far we assumed a certain distribution of the shifts.
- It is difficult to prove that the shifts will satisfy a given distribution
- We often (not always) observe that the final residuals tend to have only 2 large eigen-components $\Rightarrow$
- Explore case where we have only two nonzero components $\beta_{1,0}$ and $\beta_{n,0}$.
- The $\beta_{i,k}$'s evolve as:

$$\beta_{1,k+1} = \frac{1}{s_k}\left(\frac{\bar{\lambda}_{\beta,k}}{\sigma} - \lambda_1\right)\beta_{1,k}, \qquad \text{and} \qquad \beta_{n,k+1} = \frac{1}{s_k}\left(\frac{\bar{\lambda}_{\beta,k}}{\sigma} - \lambda_n\right)\beta_{n,k}.$$

Simplified analysis of the relaxed SD iteration

- Notation:

$$\gamma \equiv \frac{\lambda_1 + \lambda_n}{2}, \quad h \equiv \frac{\lambda_n - \lambda_1}{2}, \quad a_k \equiv \frac{\bar{\lambda}_{\beta,k}}{\sigma} - \gamma \quad \Rightarrow$$

$$\frac{\bar{\lambda}_{\beta,k}}{\sigma} - \lambda_1 = a_k + h, \quad \frac{\bar{\lambda}_{\beta,k}}{\sigma} - \lambda_n = a_k - h.$$

- Then we get:

$$\beta_{1,k+1} = \frac{1}{s_k}(a_k + h)\,\beta_{1,k}, \qquad \beta_{n,k+1} = \frac{1}{s_k}(a_k - h)\,\beta_{n,k}.$$

Theorem:

Under previous notation and assumptions:

$$\left(\frac{\bar{\lambda}_{\beta,k+1}}{\sigma} - \gamma\right) = \left(\frac{\bar{\lambda}_{\beta,k}}{\sigma} - \gamma\right)\left[1 - 2\frac{(\lambda_n - \bar{\lambda}_{\beta,k})(\bar{\lambda}_{\beta,k} - \lambda_1)}{\sigma s_k^2}\right]$$

$\Rightarrow \bar{\lambda}_{\beta,k+1}$ will change sides relative to $\sigma\gamma$ whenever

$$\sigma s_k^2 < 2(\lambda_n - \bar{\lambda}_{\beta,k})(\bar{\lambda}_{\beta,k} - \lambda_1)$$

When $\sigma = 1$ Inequality *always* satisfied $\Rightarrow \bar{\lambda}_{\beta,k}$ will oscillate around γ. When $\sigma \geq 2\lambda_n/(\lambda_1 + \lambda_n)$ then $\bar{\lambda}_{\beta,k}$ converges. When $\bar{\lambda}_{\beta,k}$ converges its limit is either λ_1, or λ_n or $\sigma\gamma$.

University of Minnesota