Iterative methods with retards for the solution of large-scale linear systems

PhD Thesis Defense

Qinmeng Zou

June 14, 2019

Jury Members

Examiners: Raphaël Couturier
Eric Monacelli
Juliette Ryan
Pierre Spitéri
Damien Tromeur-Dervout

Supervisor: Frédéric Magoulès

Outline

Introduction

Introduction: Gradient iterations

Let A be an N-dimensional symmetric positive definite (SPD) matrix
To solve:

$$Ax = b$$

Find solution $x_* \iff$ minimize quadratic function [*Temple, 1938*]:

$$f(x) = \frac{1}{2}x^\intercal Ax - b^\intercal x$$

$x^\intercal$: transpose of x
Let gradient vector $g_n = Ax_n - b$. Gradient methods, first proposed by [*Cauchy, 1847*], further formalized by [*Kantorovitch, 1945*]:

$$x_{n+1} = x_n - \alpha_n g_n$$

$\alpha_n > 0$: step, steplength, stepsize

Introduction: Basic gradient methods

Monotone methods: computing $\{\alpha_n\}$ such that $f(x_n)$ decreases monotonically

Alternate methods: two or more expressions for α_n

$\Longrightarrow$ Basic methods: monotone (or monotone + non-alternate)

- steepest descent [*Cauchy, 1847*]:

$$\alpha_n^{\mathsf{SD}} = (g_n^\intercal g_n)/(g_n^\intercal A g_n)$$

- minimal gradient [*Krasnosel'skii and Krein, 1952*]:

$$\alpha_n^{\mathsf{MG}} = (g_n^\intercal A g_n)/(g_n^\intercal A^2 g_n)$$

- asymptotically optimal [*Dai and Yang, 2006*]:

$$\alpha_n^{\mathsf{AO}} = \|g_n\| \,/\, \|A g_n\|$$

- other choices: relaxed steepest descent, alternate minimization, Dai-Yuan method, etc.

Introduction: Gradient methods with retards

- Barzilai-Borwein method [*Barzilai and Borwein, 1988*]:

$$\alpha_n^{\mathsf{BB}} = (g_{n-1}^{\mathsf{T}} g_{n-1})/(g_{n-1}^{\mathsf{T}} A g_{n-1})$$

  convergence: [*Raydan, 1993*][*Dai and Liao, 2002*]

- a general framework [*Friedlander et al., 1999*]:

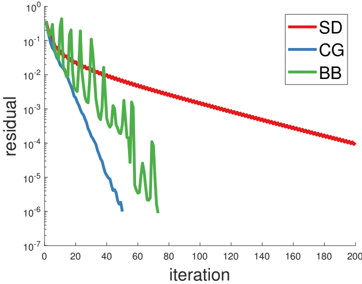
$$\alpha_n^{\mathsf{GMR}} = (g_{\tau(n)}^{\mathsf{T}} A^{\rho(n)} g_{\tau(n)})/(g_{\tau(n)}^{\mathsf{T}} A^{\rho(n)+1} g_{\tau(n)})$$

  $\tau(n)$: $\bar{n},\, \bar{n}+1,\, \ldots,\, n-1,\, n; \quad \bar{n} = \max\{0,\, n-m\}; \quad m > 0$
  $\rho(n)$: $q_1,\, \ldots,\, q_m; \quad q_j \geq 0$
  convergence: [*Friedlander et al., 1999*]

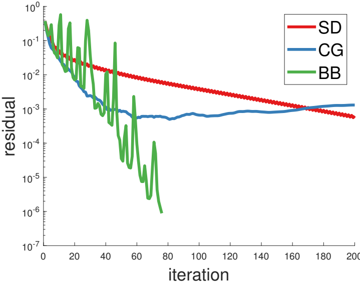
- other choices: adaptive Barzilai-Borwein, Yuan variants, steepest descent with alignment, etc.

$\Longrightarrow$ non-monotone
$\Longrightarrow$ efficient and error-tolerant

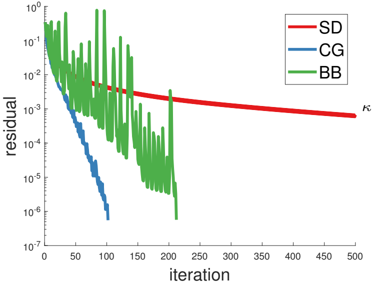
Introduction: Gradient methods with retards



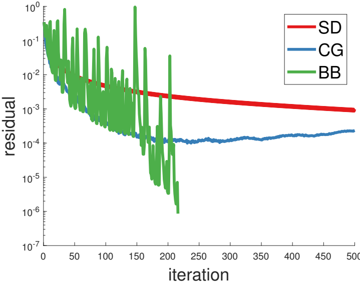
$\kappa = 10^2$



$\kappa = 10^2$
with
perturbation



$\kappa = 10^3$



$\kappa = 10^3$
with
perturbation

Introduction: Objectives

Our work: modern use of gradient iterations

Motivation: the progress of steepest descent

- asymptotic properties [*Nocedal et al., 2002*]
- steepest descent with alignment [*De Asmundis et al., 2013*]

$\Longrightarrow$ Alignment: asymptotic properties + retards. Using alternately basic steplengths and auxiliary steplengths with retards.

Introduction: Objectives

Our work: modern use of gradient iterations

Motivation: the progress of steepest descent

- asymptotic properties [*Nocedal et al., 2002*]
- steepest descent with alignment [*De Asmundis et al., 2013*]

$\Longrightarrow$ Alignment: asymptotic properties + retards. Using alternately basic steplengths and auxiliary steplengths with retards. Example:

$$\alpha_n^{\mathsf{A}_0} = \left(\frac{1}{\alpha_{n-1}^{\mathsf{SD}}} + \frac{1}{\alpha_n^{\mathsf{SD}}} \right)^{-1} \qquad \alpha_n^{\mathsf{SDA}} = \begin{cases} \alpha_n^{\mathsf{SD}}, & n \bmod (d_1 + d_2) < d_1 \\ \alpha_n^{\mathsf{A}_0}, & n \bmod (d_1 + d_2) = d_1 \\ \alpha_{n-1}^{\mathsf{SDA}}, & \text{otherwise} \end{cases}$$

with d_1, $d_2 \geq 1$ [*De Asmundis et al., 2013*]

Introduction: Objectives

Our work: modern use of gradient iterations

Motivation: the progress of steepest descent

- asymptotic properties [*Nocedal et al., 2002*]
- steepest descent with alignment [*De Asmundis et al., 2013*]

$\Longrightarrow$ Alignment: asymptotic properties $+$ retards. Using alternately basic steplengths and auxiliary steplengths with retards. Example:

$$\alpha_n^{\mathsf{A}_0} = \left(\frac{1}{\alpha_{n-1}^{\mathsf{SD}}} + \frac{1}{\alpha_n^{\mathsf{SD}}} \right)^{-1} \quad \alpha_n^{\mathsf{SDA}} = \begin{cases} \alpha_n^{\mathsf{SD}}, & n \bmod (d_1 + d_2) < d_1 \\ \alpha_n^{\mathsf{A}_0}, & n \bmod (d_1 + d_2) = d_1 \\ \alpha_{n-1}^{\mathsf{SDA}}, & \text{otherwise} \end{cases}$$

with d_1, $d_2 \geq 1$ [*De Asmundis et al., 2013*]

Existing analyses only support steepest descent iterations
Can we extend these techniques to other basic gradient methods?

Spectral properties

Spectral properties: A general steplength

Recall that: $x_{n+1} = x_n - \alpha_n g_n$ where $g_n = Ax_n - b$
$\implies g_{n+1} = g_n - \alpha_n A g_n = (I - \alpha_n A) g_n$ $\quad$ I: identity matrix

Spectral properties: A general steplength

Recall that: $x_{n+1} = x_n - \alpha_n g_n$ where $g_n = Ax_n - b$
$\implies g_{n+1} = g_n - \alpha_n A g_n = (I - \alpha_n A) g_n$ $\qquad I$: identity matrix

A is symmetric $\implies \exists\, N$ orthonormal eigenvectors $\{v_i\}$ corresponding to eigenvalues $\{\lambda_i\}$

$$g_n = \sum_{i=1}^{N} \zeta_{i,n} v_i \implies \boxed{\zeta_{i,n+1} = (1 - \alpha_n \lambda_i)\zeta_{i,n}}$$

Spectral properties: A general steplength

Recall that: $x_{n+1} = x_n - \alpha_n g_n$ where $g_n = Ax_n - b$
$\implies g_{n+1} = g_n - \alpha_n A g_n = (I - \alpha_n A) g_n$ $\quad$ I: identity matrix

A is symmetric $\implies \exists\, N$ orthonormal eigenvectors $\{v_i\}$ corresponding to eigenvalues $\{\lambda_i\}$

$$g_n = \sum_{i=1}^{N} \zeta_{i,n} v_i \implies \boxed{\zeta_{i,n+1} = (1 - \alpha_n \lambda_i)\zeta_{i,n}}$$

Consider the steplength

$$\bar{\alpha}_n = (g_n^\intercal A^\rho g_n)/(g_n^\intercal A^{\rho+1} g_n), \quad \rho \geq 0$$

$\rho = 0$: steepest descent. $\rho = 1$: minimal gradient

$$\zeta_{i,n+1} = \left(1 - \frac{g_n^\intercal A^\rho g_n}{g_n^\intercal A^{\rho+1} g_n}\lambda_i\right)\zeta_{i,n}$$

Spectral properties: A general steplength

$$\zeta_{i,n+1} = (1 - \frac{g_n^\intercal A^\rho g_n}{g_n^\intercal A^{\rho+1} g_n}\lambda_i)\zeta_{i,n}$$

By orthogonality: $g_n^\intercal A g_n = (\sum_{i=1}^N \zeta_{i,n} v_i)(\sum_{i=1}^N \zeta_{i,n}\lambda_i v_i) = \sum_{i=1}^N \lambda_i \zeta_{i,n}^2$

$$\zeta_{i,n+1} = (1 - \frac{\sum_{i=1}^N \lambda_i^\rho \zeta_{i,n}^2}{\sum_{i=1}^N \lambda_i^{\rho+1} \zeta_{i,n}^2}\lambda_i)\zeta_{i,n}$$

Spectral properties: A general steplength

$$\zeta_{i,n+1} = (1 - \frac{g_n^\intercal A^\rho g_n}{g_n^\intercal A^{\rho+1} g_n}\lambda_i)\zeta_{i,n}$$

By orthogonality: $g_n^\intercal A g_n = (\sum_{i=1}^N \zeta_{i,n} v_i)(\sum_{i=1}^N \zeta_{i,n}\lambda_i v_i) = \sum_{i=1}^N \lambda_i \zeta_{i,n}^2$

$$\zeta_{i,n+1} = (1 - \frac{\sum_{i=1}^N \lambda_i^\rho \zeta_{i,n}^2}{\sum_{i=1}^N \lambda_i^{\rho+1} \zeta_{i,n}^2}\lambda_i)\zeta_{i,n}$$

Let $\hat{p}_{i,n} = \lambda_i^\rho \zeta_{i,n}^2$, and let $p_{i,n} = \hat{p}_{i,n} / \sum_{j=1}^N \hat{p}_{j,n}$. Idea: $\boxed{\sum_{i=1}^N p_{i,n} = 1}$

$$\begin{aligned} p_{i,n+1} &= \frac{\lambda_i^\rho \zeta_{i,n+1}^2}{\sum_{j=1}^N \lambda_j^\rho \zeta_{j,n+1}^2} = (1 - \frac{\sum_{j=1}^N \lambda_j^\rho \zeta_{j,n}^2}{\sum_{j=1}^N \lambda_j^{\rho+1} \zeta_{j,n}^2}\lambda_i)^2 \frac{\lambda_i^\rho \zeta_{i,n}^2}{\sum_{j=1}^N \lambda_j^\rho \zeta_{j,n+1}^2} \\ &= \quad \cdots \quad = \frac{(\lambda_i - \sum_{j=1}^N \lambda_j p_{j,n})^2 p_{i,n}}{\sum_{l=1}^N (\lambda_l - \sum_{j=1}^N \lambda_j p_{j,n})^2 p_{l,n}} \end{aligned}$$

Spectral properties: Probability measure

Let $\bar{\lambda}^{(n)} = \sum_{j=1}^{N} \lambda_j p_{j,n}$. Then

$$p_{i,n+1} = \frac{(\lambda_i - \bar{\lambda}^{(n)})^2 p_{i,n}}{\sum_{l=1}^{N} (\lambda_l - \bar{\lambda}^{(n)})^2 p_{l,n}}$$

Let Λ be a random variable with outcomes $\{\lambda_i\}$. From a probability point of view:

- $\{p_{i,n}\}_{i \in \{1, \ldots, N\}}$: probability distribution for Λ
- $\bar{\lambda}^{(n)}$: expectation of Λ
- $\sum_{l=1}^{N} (\lambda_l - \bar{\lambda}^{(n)})^2 p_{l,n}$: variance of Λ

Let $S_k^{(n)} = \sum_{l=1}^{N} (\lambda_l - \bar{\lambda}^{(n)})^k p_{l,n}$. Then

$$S_2^{(n+1)} - S_2^{(n)} = \frac{\sum_{i=1}^{N} (\lambda_i - \bar{\lambda}^{(n+1)})^2 (\lambda_i - \bar{\lambda}^{(n)})^2 p_{i,n} - (S_2^{(n)})^2}{S_2^{(n)}}$$

Spectral properties: Probability measure

$$\boxed{S_k^{(n)} = \sum_{l=1}^{N} (\lambda_l - \bar{\lambda}^{(n)})^k p_{l,n}}$$

$$S_2^{(n+1)} - S_2^{(n)} = \quad \cdots \quad = \frac{S_4^{(n)} S_2^{(n)} - (S_3^{(n)})^2 - (S_2^{(n)})^3}{(S_2^{(n)})^2}$$

Spectral properties: Probability measure

$$\boxed{S_k^{(n)} = \sum_{l=1}^{N} (\lambda_l - \bar{\lambda}^{(n)})^k p_{l,n}}$$

$$S_2^{(n+1)} - S_2^{(n)} = \quad \cdots \quad = \frac{S_4^{(n)} S_2^{(n)} - (S_3^{(n)})^2 - (S_2^{(n)})^3}{(S_2^{(n)})^2}$$

Assume $\boxed{\lambda_1 < \cdots < \lambda_N}$ without loss of generality [*Fletcher, 2005*]

$$\det M_2^{(n)} = S_4^{(n)} S_2^{(n)} - (S_3^{(n)})^2 - (S_2^{(n)})^3 = \begin{vmatrix} 1 & S_1^{(n)} & S_2^{(n)} \\ S_1^{(n)} & S_2^{(n)} & S_3^{(n)} \\ S_2^{(n)} & S_3^{(n)} & S_4^{(n)} \end{vmatrix} \quad (S_1^{(n)} = 0)$$

$M_m^{(n)}$: moment matrix of size $m+1$

- positive semi-definiteness $\Longrightarrow$ $\det M_m^{(n)} \geq 0$
- equality holds $\iff$ m or fewer points with $p_{i,n} > 0$ [*Lindsay, 1989*]

Spectral properties: Probability measure

$\boxed{S_k^{(n)} = \sum_{l=1}^{N} (\lambda_l - \bar{\lambda}^{(n)})^k p_{l,n}}$ $\boxed{\lambda_1 < \cdots < \lambda_N}$ $\boxed{p_{i,n} = \frac{\lambda_i^\rho \zeta_{i,n}^2}{\sum_{j=1}^{N} \lambda_j^\rho \zeta_{j,n}^2}}$

$$S_2^{(n+1)} - S_2^{(n)} = (\det M_2^{(n)})/(S_2^{(n)})^2$$

Assume: at least two i such that $p_{i,n} > 0$. Then

- $S_2^{(n+1)} \geq S_2^{(n)}$
- equality holds $\iff$ only two i with $p_{i,n} > 0$

Spectral properties: Probability measure

$\boxed{S_k^{(n)} = \sum_{l=1}^N (\lambda_l - \bar{\lambda}^{(n)})^k p_{l,n}}$ $\boxed{\lambda_1 < \cdots < \lambda_N}$ $\boxed{p_{i,n} = \frac{\lambda_i^\rho \zeta_{i,n}^2}{\sum_{j=1}^N \lambda_j^\rho \zeta_{j,n}^2}}$

$$S_2^{(n+1)} - S_2^{(n)} = (\det M_2^{(n)})/(S_2^{(n)})^2$$

Assume: at least two i such that $p_{i,n} > 0$. Then

- $S_2^{(n+1)} \geq S_2^{(n)}$
- equality holds $\iff$ only two i with $p_{i,n} > 0$

From Akaike's theorem (1959):

$$n \to \infty : \quad \begin{array}{ll} p_{i,n} > 0 & \text{if } i = 1 \text{ or } N \\ p_{i,n} = 0 & \text{if } i \in \{2, \ldots, N-1\} \\ \{p_{1,n+1}, p_{N,n+1}\} = \{p_{N,n}, p_{1,n}\} & \end{array}$$

$\{p_{i,n}\}$ tends to alternate between two subsets

Similar result - [*Akaike, 1959*] for steepest descent ($\rho = 0$)

Spectral properties: Asymptotic results

$\boxed{\bar{\alpha}_n = (g_n^\intercal A^\rho g_n)/(g_n^\intercal A^{\rho+1} g_n)}$ $\boxed{p_{i,n} = (\lambda_i^\rho \zeta_{i,n}^2)/(\sum_{j=1}^N \lambda_j^\rho \zeta_{j,n}^2)}$

$\{p_{i,n}\}$ tends to alternate between two subsets

In other words the following limits hold:

$$\lim_{n\to\infty} p_{i,2n} = \begin{cases} \frac{1}{1+c^2} & i = 1 \\ \frac{c^2}{1+c^2} & i = N \\ 0 & \text{otherwise} \end{cases} \qquad \lim_{n\to\infty} p_{i,2n+1} = \begin{cases} \frac{c^2}{1+c^2} & i = 1 \\ \frac{1}{1+c^2} & i = N \\ 0 & \text{otherwise} \end{cases}$$

for some constant c

Spectral properties: Asymptotic results

$\boxed{\bar{\alpha}_n = (g_n^\intercal A^\rho g_n)/(g_n^\intercal A^{\rho+1} g_n)}$ $\boxed{p_{i,n} = (\lambda_i^\rho \zeta_{i,n}^2)/(\sum_{j=1}^N \lambda_j^\rho \zeta_{j,n}^2)}$

$\{p_{i,n}\}$ tends to alternate between two subsets

In other words the following limits hold:

$$\lim_{n\to\infty} p_{i,2n} = \begin{cases} \frac{1}{1+c^2} & i = 1 \\ \frac{c^2}{1+c^2} & i = N \\ 0 & \text{otherwise} \end{cases} \qquad \lim_{n\to\infty} p_{i,2n+1} = \begin{cases} \frac{c^2}{1+c^2} & i = 1 \\ \frac{1}{1+c^2} & i = N \\ 0 & \text{otherwise} \end{cases}$$

for some constant c

For $\bar{\alpha}_n = 1/(\sum_{i=1}^N \lambda_i p_{i,n})$, substituting $p_{i,2n}$ and $p_{i,2n+1}$ yields

$$\boxed{\lim_{n\to\infty} \bar{\alpha}_{2n} = \frac{1+c^2}{\lambda_1(1+c^2\kappa)}} \qquad \boxed{\lim_{n\to\infty} \bar{\alpha}_{2n+1} = \frac{1+c^2}{\lambda_1(c^2+\kappa)}}$$

where $\kappa = \lambda_N/\lambda_1$. Independent of ρ

Spectral properties: Asymptotic results

For gradient vectors:

$$\lim_{n\to\infty} \frac{\left\|g_{n+1}\right\|^2}{\left\|g_n\right\|^2} = \frac{\sum_{i=1}^{N} \left(1-\bar{\alpha}_n\lambda_i\right)^2 \zeta_{i,n}^2}{\sum_{i=1}^{N} \zeta_{i,n}^2}$$

$$\lim_{n\to\infty} \frac{\left\|g_{2n+1}\right\|^2}{\left\|g_{2n}\right\|^2} = \frac{\left(1-\frac{1+c^2}{1+c^2\kappa}\right)^2 \lambda_1^{-\rho}\frac{1}{1+c^2} + \left(1-\frac{(1+c^2)\kappa}{1+c^2\kappa}\right)^2 \lambda_N^{-\rho}\frac{c^2}{1+c^2}}{\lambda_1^{-\rho}\frac{1}{1+c^2} + \lambda_N^{-\rho}\frac{c^2}{1+c^2}}$$

$$= \quad \cdots \quad = \frac{c^2(\kappa-1)^2(1+c^2\kappa^\rho)}{(c^2+\kappa^\rho)(1+c^2\kappa)^2}$$

$$\lim_{n\to\infty} \frac{\left\|g_{2n+2}\right\|^2}{\left\|g_{2n+1}\right\|^2} = \frac{c^2(\kappa-1)^2(c^2+\kappa^\rho)}{(c^2+\kappa)^2(1+c^2\kappa^\rho)} \qquad \left(= \lim_{n\to\infty} \frac{\left\|g_{2n+1}\right\|^2}{\left\|g_{2n}\right\|^2} \text{ if } \rho = 1\right)$$

$\rho = 0$: the case of steepest descent [*Nocedal et al., 2002*]

Spectral properties: Asymptotic results

Other results in a nutshell:

$$\lim_{n\to\infty} \frac{g_{2n+1}^{\mathsf{T}} A^{\gamma} g_{2n+1}}{g_{2n}^{\mathsf{T}} A^{\gamma} g_{2n}} = \frac{c^2(\kappa-1)^2(1+c^2\kappa^{\rho-\gamma})}{(c^2+\kappa^{\rho-\gamma})(1+c^2\kappa)^2}$$

$$\lim_{n\to\infty} \frac{g_{2n+2}^{\mathsf{T}} A^{\gamma} g_{2n+2}}{g_{2n+1}^{\mathsf{T}} A^{\gamma} g_{2n+1}} = \frac{c^2(\kappa-1)^2(c^2+\kappa^{\rho-\gamma})}{(c^2+\kappa)^2(1+c^2\kappa^{\rho-\gamma})}$$

$\rho - \gamma = 1 \implies$ first = second

$\rho = 0$:

$$\lim_{n\to\infty} \frac{\|g_{n+2}\|^2}{\|g_n\|^2} = \lim_{n\to\infty} \frac{f(x_{n+1}) - f(x_*)}{f(x_n) - f(x_*)} = \frac{c^2(\kappa-1)^2}{(c^2+\kappa)(1+c^2\kappa)}$$

$\rho = 1$:

$$\lim_{n\to\infty} \frac{f(x_{2n+2}) - f(x_*)}{f(x_{2n}) - f(x_*)} = \lim_{n\to\infty} \frac{\|g_{n+1}\|^4}{\|g_n\|^4} = \frac{c^4(\kappa-1)^4}{(c^2+\kappa)^2(1+c^2\kappa)^2}$$

Fast methods

Fast methods: Constant steplength

Solutions with $\bar{\alpha}_n$: zigzag in two directions

- slow convergence
- badly affected by ill-conditioning

Fast methods: Constant steplength

Solutions with $\bar{\alpha}_n$: zigzag in two directions

- slow convergence
- badly affected by ill-conditioning

Idea: use spectral properties. Consider a constant steplength $\hat{\alpha}$ such that

$$\hat{\alpha} \leq \frac{2}{\lambda_1 + \lambda_N}$$

$\hat{\alpha} < 2/\lambda_N \leq 2\alpha_n^{\mathrm{SD}} \implies$ convergence [*Raydan and Svaiter, 2002*]

Recall that: $\zeta_{i,n+1} = (1 - \alpha_n \lambda_i)\zeta_{i,n}$

$\alpha_n = \hat{\alpha} \implies \zeta_{i,n+1} = (1 - \hat{\alpha}\lambda_i)^{n+1}\zeta_{i,0}$. Then

$$\lim_{n\to\infty} \frac{\zeta_{i,n}}{\zeta_{1,n}} = \frac{\zeta_{i,0}}{\zeta_{1,0}} \lim_{n\to\infty} \left(\frac{1 - \hat{\alpha}\lambda_i}{1 - \hat{\alpha}\lambda_1} \right)^n$$

Fast methods: Constant steplength

Let

$$\varphi_i = \frac{1-\hat{\alpha}\lambda_i}{1-\hat{\alpha}\lambda_1} = \frac{\lambda_i}{\lambda_1} - \frac{\lambda_i - \lambda_1}{\lambda_1(1-\hat{\alpha}\lambda_1)}$$

$|\varphi_i| < 1 \iff (\lambda_i + \lambda_1)\hat{\alpha} < 2.$ Hence

$$\hat{\alpha} < \frac{2}{\lambda_1 + \lambda_N} \implies |\varphi_i| < 1 \implies \lim_{n\to\infty} \left(\frac{1-\hat{\alpha}\lambda_i}{1-\hat{\alpha}\lambda_1}\right)^n = 0 \qquad \forall i$$

$\implies \lim_{n\to\infty}(\zeta_{i,n})/(\zeta_{1,n}) = 0$

Gradient vector tends to be aligned with v_1 when $\hat{\alpha} < 2/(\lambda_1 + \lambda_N)$

Fast methods: Constant steplength

Let

$$\varphi_i = \frac{1-\hat{\alpha}\lambda_i}{1-\hat{\alpha}\lambda_1} = \frac{\lambda_i}{\lambda_1} - \frac{\lambda_i - \lambda_1}{\lambda_1(1-\hat{\alpha}\lambda_1)}$$

$|\varphi_i| < 1 \iff (\lambda_i + \lambda_1)\hat{\alpha} < 2$. Hence

$$\hat{\alpha} < \frac{2}{\lambda_1 + \lambda_N} \implies |\varphi_i| < 1 \implies \lim_{n\to\infty}\left(\frac{1-\hat{\alpha}\lambda_i}{1-\hat{\alpha}\lambda_1}\right)^n = 0 \qquad \forall i$$

$\implies \lim_{n\to\infty}(\zeta_{i,n})/(\zeta_{1,n}) = 0$

Gradient vector tends to be aligned with v_1 when $\hat{\alpha} < 2/(\lambda_1 + \lambda_N)$

$$\hat{\alpha} = \frac{2}{\lambda_1 + \lambda_N} \implies \varphi_N = -1 \implies \lim_{n\to\infty}\frac{\zeta_{N,n}}{\zeta_{1,n}} = \frac{\zeta_{N,0}}{\zeta_{1,0}}(-1)^n$$

. . . and others remain 0

Gradient vector tends to zigzag in two directions when $\hat{\alpha} = 2/(\lambda_1 + \lambda_N)$

Fast methods: Minimal gradient with alignment

Heuristics:

- zigzag in two directions with $\bar{\alpha}_n = (g_n^\intercal A^\rho g_n)/(g_n^\intercal A^{\rho+1} g_n)$
- zigzag in two directions with $\hat{\alpha} = 2/(\lambda_1 + \lambda_N)$
- align with an eigendirection with $\hat{\alpha} < 2/(\lambda_1 + \lambda_N)$

Strategy:

- step 1: use $\bar{\alpha}_n$ - sink into two-dimensions - $i = 1, N$
- step 2: use constant steplength $\hat{\alpha}$ - align with v_1

Fast methods: Minimal gradient with alignment

Heuristics:

- zigzag in two directions with $\bar{\alpha}_n = (g_n^\intercal A^\rho g_n)/(g_n^\intercal A^{\rho+1} g_n)$
- zigzag in two directions with $\hat{\alpha} = 2/(\lambda_1 + \lambda_N)$
- align with an eigendirection with $\hat{\alpha} < 2/(\lambda_1 + \lambda_N)$

Strategy:

- step 1: use $\bar{\alpha}_n$ - sink into two-dimensions - $i = 1,\ N$
- step 2: use constant steplength $\hat{\alpha}$ - align with v_1

Consider the steplength

$$\alpha_n^{\mathsf{A}_\rho} = \left(\frac{1}{\bar{\alpha}_{n-1}} + \frac{1}{\bar{\alpha}_n}\right)^{-1} \qquad (\rho = 0 : \text{ steepest descent with alignment})$$

$$\lim_{n\to\infty} \alpha_n^{\mathsf{A}_\rho} = \left(\frac{\lambda_1(1+c^2\kappa)}{1+c^2} + \frac{\lambda_1(c^2+\kappa)}{1+c^2}\right)^{-1} = \frac{1}{\lambda_1 + \lambda_N}$$

Fast methods: Minimal gradient with alignment

Heuristics:

- zigzag in two directions with $\bar{\alpha}_n = (g_n^\intercal A^\rho g_n)/(g_n^\intercal A^{\rho+1} g_n)$
- zigzag in two directions with $\hat{\alpha} = 2/(\lambda_1 + \lambda_N)$
- align with an eigendirection with $\hat{\alpha} < 2/(\lambda_1 + \lambda_N)$

Strategy:

- step 1: use $\bar{\alpha}_n$ - sink into two-dimensions - $i = 1, N$
- step 2: use constant steplength $\hat{\alpha}$ - align with v_1

$\rho = 1$: minimal gradient with alignment

$$\alpha_n^{\mathsf{A}_1} = \left(\frac{1}{\alpha_{n-1}^{\mathsf{MG}}} + \frac{1}{\alpha_n^{\mathsf{MG}}} \right)^{-1} \qquad \alpha_n^{\mathsf{MGA}} = \begin{cases} \alpha_n^{\mathsf{MG}}, & n \bmod (d_1 + d_2) < d_1 \\ \alpha_n^{\mathsf{A}_1}, & n \bmod (d_1 + d_2) = d_1 \\ \alpha_{n-1}^{\mathsf{MGA}}, & \text{otherwise} \end{cases}$$

with $d_1, d_2 \geq 1$

Fast methods: Asymptotically optimal with alignment

Heuristics:

- zigzag in two directions with $\bar{\alpha}_n = (g_n^\intercal A^\rho g_n)/(g_n^\intercal A^{\rho+1} g_n)$
- zigzag in two directions with $\hat{\alpha} = 2/(\lambda_1 + \lambda_N)$
- align with an eigendirection with $\hat{\alpha} < 2/(\lambda_1 + \lambda_N)$

Fast methods: Asymptotically optimal with alignment

Heuristics:

- zigzag in two directions with $\bar{\alpha}_n = (g_n^\intercal A^\rho g_n)/(g_n^\intercal A^{\rho+1} g_n)$
- zigzag in two directions with $\hat{\alpha} = 2/(\lambda_1 + \lambda_N)$
- align with an eigendirection with $\hat{\alpha} < 2/(\lambda_1 + \lambda_N)$

Recall that: $\alpha_n^{\mathsf{AO}} = \|g_n\| \,/\, \|Ag_n\|$

- $\lim_{n\to\infty} \alpha_n^{\mathsf{AO}} = 2/(\lambda_1 + \lambda_N)$ [*Dai and Yang, 2006*]
- Let $\tilde{\alpha}_n = \theta\alpha_n^{\mathsf{AO}}$ where $0 < \theta < 1$

Alternating between α_n^{AO} and $\tilde{\alpha}_n$ achieves the same effect

Fast methods: Asymptotically optimal with alignment

Heuristics:

- zigzag in two directions with $\bar{\alpha}_n = (g_n^\intercal A^\rho g_n)/(g_n^\intercal A^{\rho+1} g_n)$
- zigzag in two directions with $\hat{\alpha} = 2/(\lambda_1 + \lambda_N)$
- align with an eigendirection with $\hat{\alpha} < 2/(\lambda_1 + \lambda_N)$

Recall that: $\alpha_n^{\mathsf{AO}} = \|g_n\| \,/\, \|Ag_n\|$

- $\lim_{n\to\infty} \alpha_n^{\mathsf{AO}} = 2/(\lambda_1 + \lambda_N)$ [*Dai and Yang, 2006*]
- Let $\tilde{\alpha}_n = \theta \alpha_n^{\mathsf{AO}}$ where $0 < \theta < 1$

Alternating between α_n^{AO} and $\tilde{\alpha}_n$ achieves the same effect

asymptotically optimal with alignment:

$$\alpha_n^{\mathsf{AOA}} = \begin{cases} \alpha_n^{\mathsf{AO}}, & n \bmod (d_1 + d_2) < d_1 \\ \tilde{\alpha}_n, & n \bmod (d_1 + d_2) = d_1 \\ \alpha_{n-1}^{\mathsf{AOA}}, & \text{otherwise} \end{cases}$$

with $d_1,\, d_2 \geq 1$

Fast methods: Generalized Yuan steplength

Recall that $\rho = 1$ (minimal gradient) leads to:

$$\boxed{\lim_{n\to\infty} \alpha_{2n}^{\mathsf{MG}} = \frac{1+c^2}{\lambda_1(1+c^2\kappa)}} \qquad \boxed{\lim_{n\to\infty} \alpha_{2n+1}^{\mathsf{MG}} = \frac{1+c^2}{\lambda_1(c^2+\kappa)}}$$

$$\boxed{\lim_{n\to\infty} \frac{g_{2n+1}^{\mathsf{T}} A g_{2n+1}}{g_{2n}^{\mathsf{T}} A g_{2n}} = \frac{c^2(\kappa-1)^2}{(1+c^2\kappa)^2}} \qquad \boxed{\lim_{n\to\infty} \frac{g_{2n+2}^{\mathsf{T}} A g_{2n+2}}{g_{2n+1}^{\mathsf{T}} A g_{2n+1}} = \frac{c^2(\kappa-1)^2}{(c^2+\kappa)^2}}$$

Fast methods: Generalized Yuan steplength

Recall that $\rho = 1$ (minimal gradient) leads to:

$$\boxed{\lim_{n\to\infty} \alpha_{2n}^{\mathsf{MG}} = \frac{1+c^2}{\lambda_1(1+c^2\kappa)}} \qquad \boxed{\lim_{n\to\infty} \alpha_{2n+1}^{\mathsf{MG}} = \frac{1+c^2}{\lambda_1(c^2+\kappa)}}$$

$$\boxed{\lim_{n\to\infty} \frac{g_{2n+1}^{\mathsf{T}} A g_{2n+1}}{g_{2n}^{\mathsf{T}} A g_{2n}} = \frac{c^2(\kappa-1)^2}{(1+c^2\kappa)^2}} \qquad \boxed{\lim_{n\to\infty} \frac{g_{2n+2}^{\mathsf{T}} A g_{2n+2}}{g_{2n+1}^{\mathsf{T}} A g_{2n+1}} = \frac{c^2(\kappa-1)^2}{(c^2+\kappa)^2}}$$

$$\Longrightarrow \lim_{n\to\infty} \frac{g_{2n+2}^{\mathsf{T}} A g_{2n+2}}{\left(\alpha_{2n+1}^{\mathsf{MG}}\right)^2 g_{2n+1}^{\mathsf{T}} A g_{2n+1}} = \lim_{n\to\infty} \frac{g_{2n+1}^{\mathsf{T}} A g_{2n+1}}{\left(\alpha_{2n}^{\mathsf{MG}}\right)^2 g_{2n}^{\mathsf{T}} A g_{2n}} = \frac{\lambda_1^2 c^2(\kappa-1)^2}{(1+c^2)^2}$$

$$\lim_{n\to\infty} \frac{1}{\alpha_{n-1}^{\mathsf{MG}} \alpha_n^{\mathsf{MG}}} = \frac{\lambda_1^2(1+c^2\kappa)(c^2+\kappa)}{(1+c^2)^2}$$

$$\Longrightarrow \lim_{n\to\infty} \left(\frac{1}{\alpha_{n-1}^{\mathsf{MG}} \alpha_n^{\mathsf{MG}}} - \frac{g_n^{\mathsf{T}} A g_n}{\left(\alpha_{n-1}^{\mathsf{MG}}\right)^2 g_{n-1}^{\mathsf{T}} A g_{n-1}} \right) = \quad \cdots \quad = \lambda_1 \lambda_N$$

Fast methods: Generalized Yuan steplength

Let

$$\alpha_n^{\mathsf{Y}_1} = 2\left(\sqrt{\left(\frac{1}{\alpha_{n-1}^{\mathsf{MG}}} - \frac{1}{\alpha_n^{\mathsf{MG}}}\right)^2 + \frac{4g_n^{\mathsf{T}}Ag_n}{\left(\alpha_{n-1}^{\mathsf{MG}}\right)^2 g_{n-1}^{\mathsf{T}}Ag_{n-1}}} + \frac{1}{\alpha_{n-1}^{\mathsf{MG}}} + \frac{1}{\alpha_n^{\mathsf{MG}}}\right)^{-1}$$

Since

$$\lim_{n\to\infty}\left(\frac{1}{\alpha_{n-1}^{\mathsf{MG}}} + \frac{1}{\alpha_n^{\mathsf{MG}}}\right)^{-1} = \frac{1}{\lambda_1+\lambda_N}$$

$$\lim_{n\to\infty}\left(\frac{1}{\alpha_{n-1}^{\mathsf{MG}}\alpha_n^{\mathsf{MG}}} - \frac{g_n^{\mathsf{T}}Ag_n}{\left(\alpha_{n-1}^{\mathsf{MG}}\right)^2 g_{n-1}^{\mathsf{T}}Ag_{n-1}}\right) = \lambda_1\lambda_N$$

It follows that

$$\lim_{n\to\infty}\alpha_n^{\mathsf{Y}_1} = 2\left(\sqrt{(\lambda_1+\lambda_N)^2 - 4\lambda_1\lambda_N} + \lambda_1 + \lambda_N\right)^{-1} = \frac{1}{\lambda_N}$$

Fast methods: Generalized Yuan steplength

Classical Yuan steplength [*Yuan, 2006*]:

- inspired by fast gradient methods for two-dimensional matrices
- spectral properties: first provided by [*De Asmundis et al., 2014*]
- only for steepest descent

Fast methods: Generalized Yuan steplength

Classical Yuan steplength [*Yuan, 2006*]:

- inspired by fast gradient methods for two-dimensional matrices
- spectral properties: first provided by [*De Asmundis et al., 2014*]
- only for steepest descent

Generalized Yuan steplength: based on

$$\boxed{\bar{\alpha}_n = \frac{g_n^{\mathsf{T}} A^{\rho} g_n}{g_n^{\mathsf{T}} A^{\rho+1} g_n}} \quad \text{and} \quad \boxed{\bar{\beta}_n = \frac{g_{n+1}^{\mathsf{T}} A^{\rho} g_{n+1}}{g_n^{\mathsf{T}} A^{\rho} g_n}}$$

$$\Longrightarrow \boxed{\lim_{n\to\infty} \alpha_n^{\mathsf{Y}^{\rho}} = 1/\lambda_N}$$

Fast methods: Generalized Yuan steplength

Classical Yuan steplength [*Yuan, 2006*]:

- inspired by fast gradient methods for two-dimensional matrices
- spectral properties: first provided by [*De Asmundis et al., 2014*]
- only for steepest descent

Generalized Yuan steplength: based on

$$\boxed{\bar{\alpha}_n = \frac{g_n^{\mathsf{T}} A^\rho g_n}{g_n^{\mathsf{T}} A^{\rho+1} g_n}} \quad \text{and} \quad \boxed{\bar{\beta}_n = \frac{g_{n+1}^{\mathsf{T}} A^\rho g_{n+1}}{g_n^{\mathsf{T}} A^\rho g_n}}$$

$$\Longrightarrow \boxed{\lim_{n\to\infty} \alpha_n^{\mathsf{Y}^\rho} = 1/\lambda_N}$$

Notice that:

$$\zeta_{i,n+1} = (1 - \alpha_n \lambda_i)\zeta_{i,n} = \prod_{j=0}^{n} (1 - \alpha_j \lambda_i)\zeta_{i,0}$$

$\alpha_j = 1/\lambda_N \implies \zeta_{N,j+1}, \zeta_{N,j+2}, \dots = 0$: vanish forever!

Fast methods: Generalized Yuan steplength

Heuristics:

- zigzag in two directions with $\bar{\alpha}_n = (g_n^\intercal A^\rho g_n)/(g_n^\intercal A^{\rho+1} g_n)$
- zigzag in two directions with $\hat{\alpha} = 2/(\lambda_1 + \lambda_N)$
- align with an eigendirection with $\hat{\alpha} < 2/(\lambda_1 + \lambda_N)$
- align with an eigendirection with $\alpha_n^{\mathsf{Y}_\rho}$

$$\alpha_n^{\mathsf{Y}_\rho} = 2\left(\sqrt{\left(\frac{1}{\bar{\alpha}_{n-1}} - \frac{1}{\bar{\alpha}_n}\right)^2 + \frac{4\bar{\beta}_{n-1}}{(\bar{\alpha}_{n-1})^2}} + \frac{1}{\bar{\alpha}_{n-1}} + \frac{1}{\bar{\alpha}_n}\right)^{-1}$$

Fast methods: Generalized Yuan steplength

Heuristics:

- zigzag in two directions with $\bar{\alpha}_n = (g_n^\intercal A^\rho g_n)/(g_n^\intercal A^{\rho+1} g_n)$
- zigzag in two directions with $\hat{\alpha} = 2/(\lambda_1 + \lambda_N)$
- align with an eigendirection with $\hat{\alpha} < 2/(\lambda_1 + \lambda_N)$
- align with an eigendirection with $\alpha_n^{\mathsf{Y}_\rho}$

$$\alpha_n^{\mathsf{Y}_\rho} = 2\left(\sqrt{\left(\frac{1}{\bar{\alpha}_{n-1}} - \frac{1}{\bar{\alpha}_n}\right)^2 + \frac{4\bar{\beta}_{n-1}}{(\bar{\alpha}_{n-1})^2}} + \frac{1}{\bar{\alpha}_{n-1}} + \frac{1}{\bar{\alpha}_n}\right)^{-1}$$

$\rho = 1$, $\gamma = 1$: minimal gradient aligned with constant steplength

$$\alpha_n^{\mathsf{MGC}} = \begin{cases} \alpha_n^{\mathsf{MG}}, & n \bmod (d_1 + d_2) < d_1 \\ \alpha_n^{\mathsf{Y}_1}, & n \bmod (d_1 + d_2) = d_1 \\ \alpha_{n-1}^{\mathsf{MGC}}, & \text{otherwise} \end{cases}$$

with d_1, $d_2 \geq 1$

Fast methods: Convergence

Alignment methods: step $\alpha_n = \alpha_{n-1}$ yields retards

$$\boxed{\tilde{\alpha}_{\tau(n)} = \theta \frac{\left\| g_{\tau(n)} \right\|}{\left\| A g_{\tau(n)} \right\|}} \quad \boxed{\alpha^{\mathsf{A}_\rho}_{\tau(n)} = \left(\frac{1}{\bar{\alpha}_{\tau(n)-1}} + \frac{1}{\bar{\alpha}_{\tau(n)}} \right)^{-1}}$$

$$\boxed{\alpha^{\mathsf{Y}_\rho}_{\tau(n)} = 2 \left(\sqrt{\left(\frac{1}{\bar{\alpha}_{\tau(n)-1}} - \frac{1}{\bar{\alpha}_{\tau(n)}} \right)^2 + \frac{4\bar{\beta}_{\tau(n)-1}}{(\bar{\alpha}_{\tau(n)-1})^2}} + \frac{1}{\bar{\alpha}_{\tau(n)-1}} + \frac{1}{\bar{\alpha}_{\tau(n)}} \right)^{-1}}$$

where $\tau(n) \leq n$

Fast methods: Convergence

Alignment methods: step $\alpha_n = \alpha_{n-1}$ yields retards

$$\boxed{\tilde{\alpha}_{\tau(n)} = \theta \frac{\left\| g_{\tau(n)} \right\|}{\left\| A g_{\tau(n)} \right\|}} \quad \boxed{\alpha^{\mathsf{A}_\rho}_{\tau(n)} = \left(\frac{1}{\bar{\alpha}_{\tau(n)-1}} + \frac{1}{\bar{\alpha}_{\tau(n)}} \right)^{-1}}$$

$$\boxed{\alpha^{\mathsf{Y}_\rho}_{\tau(n)} = 2 \left(\sqrt{\left(\frac{1}{\bar{\alpha}_{\tau(n)-1}} - \frac{1}{\bar{\alpha}_{\tau(n)}} \right)^2 + \frac{4\bar{\beta}_{\tau(n)-1}}{(\bar{\alpha}_{\tau(n)-1})^2}} + \frac{1}{\bar{\alpha}_{\tau(n)-1}} + \frac{1}{\bar{\alpha}_{\tau(n)}} \right)^{-1}}$$

where $\tau(n) \leq n$

Convergence framework of [*Dai, 2003*]:

- add orthogonal transformation such that $A = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$
- add scale factor such that $\lambda_1 = 1$
- use Property A to prove convergence

Gradient method: invariant under orthogonal transformation [*Fletcher, 2005*]

Fast methods: Convergence

Recall that: $g_n = \sum_{i=1}^{N} \zeta_{i,n} v_i$. $A = \mathrm{diag}(\lambda_1, \ldots, \lambda_N) \implies g_{i,n} = \zeta_{i,n}$

Definition:

If $\exists m_0 \in \mathbb{N}$, $\exists c_1, c_2 > 0$, such that $\forall \mu \in \{1, \ldots, N-1\}$, $\forall \varepsilon > 0$, $\forall j \in \{0, \ldots, \min\{n, m_0\}\}$:

- $\lambda_1 \leq \alpha_n^{-1} \leq c_1$
- if $\sum_{i=1}^{\mu} g_{i,n-j}^2 \leq \varepsilon$ and $g_{\mu+1,n-j}^2 \geq c_2 \varepsilon$, then $\alpha_n^{-1} \geq \frac{2}{3} \lambda_{\mu+1}$

then the steplength α_n has Property A

If the steplength α_n has Property A, then the sequence $\{\|g_n\|\}$ generated by a gradient method converges to zero [*Dai, 2003*]

Fast methods: Convergence

$\exists m_0 \in \mathbb{N}$, $\exists c_1, c_2 > 0$, such that $\forall \mu \leq N-1$, $\forall \varepsilon > 0$, $\forall j \leq \min\{n, m_0\}$:

- $\lambda_1 \leq \alpha_n^{-1} \leq c_1$
- if $\sum_{i=1}^{\mu} g_{i,n-j}^2 \leq \varepsilon$ and $g_{\mu+1,n-j}^2 \geq c_2 \varepsilon$, then $\alpha_n^{-1} \geq \frac{2}{3}\lambda_{\mu+1}$

$\boxed{\tilde{\alpha}_{\tau(n)} = \theta \left\| g_{\tau(n)} \right\| / \left\| A g_{\tau(n)} \right\|}$ with $0 < \theta \leq 1$

$$\tilde{\alpha}_{\tau(n)} = \theta \left(\frac{g_{\tau(n)}^\mathsf{T} g_{\tau(n)}}{g_{\tau(n)}^\mathsf{T} A^2 g_{\tau(n)}} \right)^{\frac{1}{2}} = \theta \left(\frac{g_{\tau(n)}^\mathsf{T} A g_{\tau(n)}}{g_{\tau(n)}^\mathsf{T} A^2 g_{\tau(n)}} \cdot \frac{g_{\tau(n)}^\mathsf{T} g_{\tau(n)}}{g_{\tau(n)}^\mathsf{T} A g_{\tau(n)}} \right)^{\frac{1}{2}}$$

Rayleigh quotient: $\lambda_1 < R(A, u) = (u^\mathsf{T} A u)/(u^\mathsf{T} u) < \lambda_N \implies$

$$\theta/\lambda_N \leq \tilde{\alpha}_{\tau(n)} \leq \theta/\lambda_1 \leq 1/\lambda_1$$

$\implies \boxed{c_1 = \lambda_N/\theta}$ the first condition follows

Fast methods: Convergence

$\exists m_0 \in \mathbb{N}$, $\exists c_1, c_2 > 0$, such that $\forall \mu \leq N-1$, $\forall \varepsilon > 0$, $\forall j \leq \min\{n, m_0\}$:

- $\lambda_1 \leq \alpha_n^{-1} \leq c_1$
- if $\sum_{i=1}^{\mu} g_{i,n-j}^2 \leq \varepsilon$ and $g_{\mu+1,n-j}^2 \geq c_2 \varepsilon$, then $\alpha_n^{-1} \geq \frac{2}{3}\lambda_{\mu+1}$

Let m_0 be the maximum retard and notice that $\tau(n) = n - j$:

$$\theta \cdot \left(\tilde{\alpha}_{\tau(n)}\right)^{-1} = \left(\frac{g_{\tau(n)}^{\mathsf{T}} A^2 g_{\tau(n)}}{g_{\tau(n)}^{\mathsf{T}} g_{\tau(n)}}\right)^{\frac{1}{2}} = \left(\frac{\sum_{i=1}^{N} g_{i,n-j}^2 \lambda_i^2}{\sum_{i=1}^{\mu} g_{i,n-j}^2 + \sum_{i=\mu+1}^{N} g_{i,n-j}^2}\right)^{\frac{1}{2}}$$

$$\geq \left(\frac{\lambda_{\mu+1}^2 \sum_{i=\mu+1}^{N} g_{i,n-j}^2}{\sum_{i=1}^{\mu} g_{i,n-j}^2 + \sum_{i=\mu+1}^{N} g_{i,n-j}^2}\right)^{\frac{1}{2}} \geq \left(\frac{\lambda_{\mu+1}^2 g_{\mu+1,n-j}^2}{\epsilon + g_{\mu+1,n-j}^2}\right)^{\frac{1}{2}} \geq \left(\frac{\lambda_{\mu+1}^2 c_2}{1 + c_2}\right)^{\frac{1}{2}}$$

$$\boxed{c_2 = 0.8} \implies \left(\tilde{\alpha}_{\tau(n)}\right)^{-1} \geq (1/\theta) \cdot \left(4\lambda_{\mu+1}^2/9\right)^{\frac{1}{2}} \geq (2\lambda_{\mu+1})/3$$

$\implies$ the second condition follows $\implies$ convergence of AOA

Fast methods: Convergence

For minimal gradient with alignment:

$$\boxed{\alpha^{\mathsf{A}_\rho}_{\tau(n)} = \left(\frac{1}{\bar{\alpha}_{\tau(n)-1}} + \frac{1}{\bar{\alpha}_{\tau(n)}}\right)^{-1}} \quad \boxed{\bar{\alpha}_{\tau(n)} = \frac{g^{\mathsf{T}}_{\tau(n)} A^{\rho} g_{\tau(n)}}{g^{\mathsf{T}}_{\tau(n)} A^{\rho+1} g_{\tau(n)}}}$$

$$\frac{1}{2\lambda_N} \leq \frac{\min\{\bar{\alpha}_{\tau(n)-1},\, \bar{\alpha}_{\tau(n)}\}}{2} \leq \alpha^{\mathsf{A}_\rho}_{\tau(n)} \leq \min\{\bar{\alpha}_{\tau(n)-1},\, \bar{\alpha}_{\tau(n)}\} \leq \frac{1}{\lambda_1}$$

$\boxed{c_1 = 2\lambda_N}$ $\Longrightarrow$ the first condition follows

$$\left(\alpha^{\mathsf{A}_\rho}_{\tau(n)}\right)^{-1} \geq \frac{1}{\min\{\bar{\alpha}_{\tau(n)-1},\, \bar{\alpha}_{\tau(n)}\}} \geq \frac{1}{\bar{\alpha}_{\tau(n)}} = \frac{g^{\mathsf{T}}_{\tau(n)} A^{\rho+1} g_{\tau(n)}}{g^{\mathsf{T}}_{\tau(n)} A^{\rho} g_{\tau(n)}}$$

$$\geq \quad \cdots \quad \geq \frac{\lambda_{\mu+1} c_2}{1 + c_2}$$

$\boxed{c_2 = 2}$ $\Longrightarrow$ the second condition follows $\Longrightarrow$ convergence of MGA

Fast methods: Convergence

For minimal gradient aligned with constant steplength:

$$\boxed{\bar{\alpha}_{\tau(n)} = \frac{g_{\tau(n)}^{\mathsf{T}} A^{\rho} g_{\tau(n)}}{g_{\tau(n)}^{\mathsf{T}} A^{\rho+1} g_{\tau(n)}}} \quad \boxed{\bar{\beta}_{\tau(n)} = \frac{g_{\tau(n)+1}^{\mathsf{T}} A^{\rho} g_{\tau(n)+1}}{g_{\tau(n)}^{\mathsf{T}} A^{\rho} g_{\tau(n)}}}$$

$$\boxed{\alpha_{\tau(n)}^{\mathsf{Y}_\rho} = 2\left(\sqrt{\left(\frac{1}{\bar{\alpha}_{\tau(n)-1}} - \frac{1}{\bar{\alpha}_{\tau(n)}}\right)^2 + \frac{4\bar{\beta}_{\tau(n)-1}}{(\bar{\alpha}_{\tau(n)-1})^2}} + \frac{1}{\bar{\alpha}_{\tau(n)-1}} + \frac{1}{\bar{\alpha}_{\tau(n)}}\right)^{-1}}$$

$$\alpha_{\tau(n)}^{\mathsf{Y}_\rho} < \quad \cdots \quad < \min\{\bar{\alpha}_{\tau(n)-1}, \bar{\alpha}_{\tau(n)}\} \leq \frac{1}{\lambda_1}$$

Then by Kantorovich inequality

$$\bar{\beta}_{\tau(n)} = \quad \cdots \quad = \frac{g_{\tau(n)}^{\mathsf{T}} A^{\rho} g_{\tau(n)} \cdot g_{\tau(n)}^{\mathsf{T}} A^{\rho+2} g_{\tau(n)}}{\left(g_{\tau(n)}^{\mathsf{T}} A^{\rho+1} g_{\tau(n)}\right)^2} - 1 \leq \frac{(\lambda_N - \lambda_1)^2}{4\lambda_N \lambda_1}$$

Fast methods: Convergence

It follows that

$$\alpha_{\tau(n)}^{\mathsf{Y}_\rho} \geq 2\left(\sqrt{(\lambda_N - \lambda_1)^2 + \kappa(\lambda_N - \lambda_1)^2} + 2\lambda_N\right)^{-1} \qquad \text{constant}$$

$\Longrightarrow$ the first condition follows

$$\left(\alpha_{\tau(n)}^{\mathsf{Y}_\rho}\right)^{-1} \geq \frac{1}{\min\{\bar{\alpha}_{\tau(n)-1},\, \bar{\alpha}_{\tau(n)}\}} \qquad \text{same as MGA}$$

$\Longrightarrow$ the second condition follows $\Longrightarrow$ convergence of MGC

Fast methods: Convergence

It follows that

$$\alpha_{\tau(n)}^{\mathsf{Y}_\rho} \geq 2\left(\sqrt{(\lambda_N-\lambda_1)^2+\kappa(\lambda_N-\lambda_1)^2}+2\lambda_N\right)^{-1} \qquad \text{constant}$$

$\Longrightarrow$ the first condition follows

$$\left(\alpha_{\tau(n)}^{\mathsf{Y}_\rho}\right)^{-1} \geq \frac{1}{\min\{\bar{\alpha}_{\tau(n)-1},\, \bar{\alpha}_{\tau(n)}\}} \qquad \text{same as MGA}$$

$\Longrightarrow$ the second condition follows $\Longrightarrow$ convergence of MGC

REMARK:

- convergence for alternate methods: all steplengths have Property A [*Dai, 2003*]
- steepest descent and minimal gradient have Property A [*Dai, 2003*]
- convergence results can also be proved by contradiction without using Dai's theorem. Idea: [*Raydan, 1993*]

Applications

Applications: Splitting methods

Let $H = (A + A^{\mathsf{H}})/2$ and $S = (A - A^{\mathsf{H}})/2$. A^{H}: conjugate transpose. Let A be non-Hermitian positive definite, i.e., H is Hermitian positive definite

Hermitian and skew-Hermitian splitting [*Bai et al., 2003*]:

$$\begin{cases} (\gamma I + H)x_{n+\frac{1}{2}} = (\gamma I - S)x_n + b \\ (\gamma I + S)x_{n+1} = (\gamma I - H)x_{n+\frac{1}{2}} + b \end{cases}$$

with $\gamma > 0$.

Applications: Splitting methods

Let $H = (A + A^{\mathsf{H}})/2$ and $S = (A - A^{\mathsf{H}})/2$. A^{H}: conjugate transpose. Let A be non-Hermitian positive definite, i.e., H is Hermitian positive definite

Hermitian and skew-Hermitian splitting [*Bai et al., 2003*]:

$$\begin{cases} (\gamma I + H)x_{n+\frac{1}{2}} = (\gamma I - S)x_n + b \\ (\gamma I + S)x_{n+1} = (\gamma I - H)x_{n+\frac{1}{2}} + b \end{cases}$$

with $\gamma > 0$. Stationary iterative form: $x_{n+1} = Tx_n + p$

$$T = (\gamma I + S)^{-1}(\gamma I - H)(\gamma I + H)^{-1}(\gamma I - S)$$

Let

$$\hat{T} = (\gamma I - H)(\gamma I + H)^{-1}(\gamma I - S)(\gamma I + S)^{-1}$$

By similarity invariance, T and $\hat{T}$ have the same eigenvalues

$S^{\mathsf{H}} = -S \implies (\gamma I - S)(\gamma I + S)^{-1}$ is unitary (Cayley transform)

Applications: Splitting methods

Let $\rho(\cdot)$ be the spectral radius and let $\sigma(\cdot)$ be the spectrum:

$$\rho(T) = \rho(\hat{T}) \leq \left\|(\gamma I - H)(\gamma I + H)^{-1}\right\| = \max_{\lambda \in \sigma(H)} \frac{|\lambda - \gamma|}{|\lambda + \gamma|}$$

Choosing optimal parameter $\gamma_* = \sqrt{\lambda_1(H)\lambda_N(H)}$ leads to

$$\rho(T) \leq \frac{\sqrt{\kappa(H)} - 1}{\sqrt{\kappa(H)} + 1}$$

[*Bai et al., 2003*]. $\lambda_i(\cdot)$: eigenvalues. $\kappa(\cdot)$: condition number

Applications: Splitting methods

Let $\rho(\cdot)$ be the spectral radius and let $\sigma(\cdot)$ be the spectrum:

$$\rho(T) = \rho(\hat{T}) \leq \left\|(\gamma I - H)(\gamma I + H)^{-1}\right\| = \max_{\lambda \in \sigma(H)} \frac{|\lambda - \gamma|}{|\lambda + \gamma|}$$

Choosing optimal parameter $\gamma_* = \sqrt{\lambda_1(H)\lambda_N(H)}$ leads to

$$\rho(T) \leq \frac{\sqrt{\kappa(H)} - 1}{\sqrt{\kappa(H)} + 1}$$

[*Bai et al., 2003*]. $\lambda_i(\cdot)$: eigenvalues. $\kappa(\cdot)$: condition number

Idea: estimate γ_* by gradient iterations

- use asymptotic results of gradient methods
- not necessary to get an exact estimate

Applications: Splitting methods

First approach: Let

$$\Gamma_{\rho,n} = \frac{1}{\bar{\alpha}_{n-1}\bar{\alpha}_n} - \frac{\bar{\beta}_{n-1}}{(\bar{\alpha}_{n-1})^2}$$

Recall that

$$\bar{\alpha}_n = \frac{g_n^{\mathsf{T}} A^{\rho} g_n}{g_n^{\mathsf{T}} A^{\rho+1} g_n} \quad \text{and} \quad \bar{\beta}_n = \frac{g_{n+1}^{\mathsf{T}} A^{\rho} g_{n+1}}{g_n^{\mathsf{T}} A^{\rho} g_n}$$

Let

$$Q_{\rho,n}(\alpha) = \Gamma_{\rho,n}\alpha^2 - (\alpha_n^{\mathsf{A}_\rho})^{-1}\alpha + 1$$

Notice that the roots

$$\frac{2}{(\alpha_n^{\mathsf{A}_\rho})^{-1} + \sqrt{(\alpha_n^{\mathsf{A}_\rho})^{-2} - 4\Gamma_{\rho,n}}} \quad \text{and} \quad \frac{2}{(\alpha_n^{\mathsf{A}_\rho})^{-1} - \sqrt{(\alpha_n^{\mathsf{A}_\rho})^{-2} - 4\Gamma_{\rho,n}}}$$

are positive and $Q_{\rho,n}(0) = 1$. Then $\Gamma_{\rho,n} > 0$ and

$$\boxed{\gamma_* = \lim_{n\to\infty} \sqrt{\Gamma_{\rho,n}}}$$

Applications: Splitting methods

Second approach: Let $\mathcal{M} = \gamma I + H$. Then $\lambda_i(\mathcal{M}) = \gamma + \lambda_i(H)$

Choosing an arbitrary γ, it follows that

$$
\begin{aligned}
\gamma_* &= \sqrt{\lambda_1(H)\lambda_N(H)} = \sqrt{(\lambda_1(\mathcal{M}) - \gamma)(\lambda_N(\mathcal{M}) - \gamma)} \\
&= \sqrt{\lambda_1(\mathcal{M})\lambda_N(\mathcal{M}) - \gamma(\lambda_1(\mathcal{M}) + \lambda_N(\mathcal{M})) + \gamma^2} \\
&= \sqrt{\lim_{n\to\infty} \Gamma_{\rho,n} - \gamma \lim_{n\to\infty} (\alpha_n^{\mathsf{A}_\rho})^{-1} + \gamma^2}
\end{aligned}
$$

Then

$$
\boxed{\gamma_* = \lim_{n\to\infty} \sqrt{\Gamma_{\rho,n} - \gamma(\alpha_n^{\mathsf{A}_\rho})^{-1} + \gamma^2}}
$$

Note:

- iterations for $\mathcal{M}$, not H, useful when H is not explicit
- should give an estimate γ in the beginning

Applications: s-dimensional methods

Let A be symmetric definite positive. Let s be a positive integer. Then

$$x_{n+1} = x_n - \alpha_n^{(1)} g_n - \cdots - \alpha_n^{(s)} A^{s-1} g_n$$

Minimizing quadratic function f yields

$$\begin{pmatrix} \omega_n^{(1)} & \omega_n^{(2)} & \cdots & \omega_n^{(s)} \\ \omega_n^{(2)} & \omega_n^{(3)} & \cdots & \omega_n^{(s+1)} \\ \vdots & \vdots & & \vdots \\ \omega_n^{(s)} & \omega_n^{(s+1)} & \cdots & \omega_n^{(2s-1)} \end{pmatrix} \begin{pmatrix} \alpha_n^{(1)} \\ \alpha_n^{(2)} \\ \vdots \\ \alpha_n^{(s)} \end{pmatrix} = \begin{pmatrix} \omega_n^{(0)} \\ \omega_n^{(1)} \\ \vdots \\ \omega_n^{(s-1)} \end{pmatrix}$$

where $\omega_n^{(j)} = g_n^{\intercal} A^j g_n$ [*Forsythe, 1968*]

Applications: s-dimensional methods

Let A be symmetric definite positive. Let s be a positive integer. Then

$$x_{n+1} = x_n - \alpha_n^{(1)} g_n - \cdots - \alpha_n^{(s)} A^{s-1} g_n$$

Minimizing quadratic function f yields

$$\begin{pmatrix} \omega_n^{(1)} & \omega_n^{(2)} & \cdots & \omega_n^{(s)} \\ \omega_n^{(2)} & \omega_n^{(3)} & \cdots & \omega_n^{(s+1)} \\ \vdots & \vdots & & \vdots \\ \omega_n^{(s)} & \omega_n^{(s+1)} & \cdots & \omega_n^{(2s-1)} \end{pmatrix} \begin{pmatrix} \alpha_n^{(1)} \\ \alpha_n^{(2)} \\ \vdots \\ \alpha_n^{(s)} \end{pmatrix} = \begin{pmatrix} \omega_n^{(0)} \\ \omega_n^{(1)} \\ \vdots \\ \omega_n^{(s-1)} \end{pmatrix}$$

where $\omega_n^{(j)} = g_n^{\intercal} A^j g_n$ [*Forsythe, 1968*]

Idea: add retards. Cyclic steepest descent:

$$\alpha_n^{\mathsf{CSD}} = \begin{cases} \alpha_n^{\mathsf{SD}}, & n \bmod d = 0 \\ \alpha_{n-1}, & \text{otherwise} \end{cases}$$

with $d \geq 2$ [*Friedlander et al., 1999*]

Applications: s-dimensional methods

Recall the framework of gradient methods with retards:

$$\alpha_n^{\mathsf{GMR}} = (g_{\tau(n)}^{\mathsf{T}} A^{\rho(n)} g_{\tau(n)})/(g_{\tau(n)}^{\mathsf{T}} A^{\rho(n)+1} g_{\tau(n)})$$

Observe that $(\omega_{\tau(n)}^{(j)})/(\omega_{\tau(n)}^{(j+1)})$ satisfies this form with $0 \leq j \leq 2s-1$

Cyclic s-dimensional steepest descent:

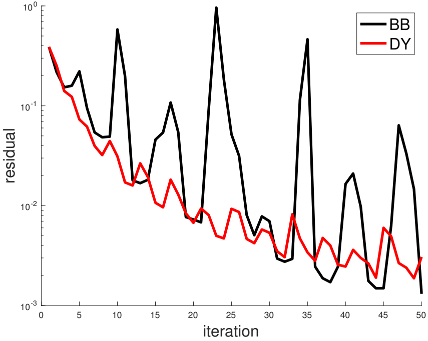
- step 1: compute $g_n, \ldots, A^s g_n$, $\{\omega_n^{(j)}\}$ and $\{\alpha_n^{(i)}\}$
- step 2: update $x_{n+1} = x_n - \alpha_n^{(1)} g_n - \cdots - \alpha_n^{(s)} A^{s-1} g_n$
- step 3: compute a steplength $\hat{\alpha} = (\omega_n^{(j)})/(\omega_n^{(j+1)})$, update n
- step 4: update multiple times g_n, $x_{n+1} = x_n - \hat{\alpha} g_n$, update n

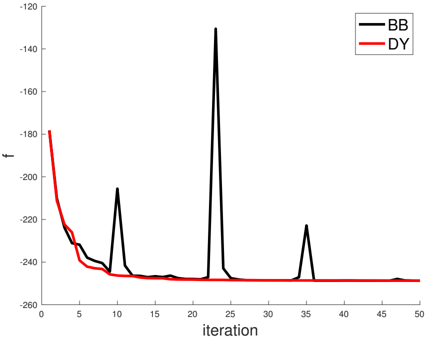
Damped version: update multiple times by using decreasing $\hat{\alpha}_n$ based on

$$\frac{\omega_n^{(2s-2)}}{\omega_n^{(2s-1)}} < \cdots < \frac{\omega_n^{(0)}}{\omega_n^{(1)}} \qquad \text{(by Cauchy–Schwarz inequality)}$$

Experimental results

Experimental results: Monotone vs non-monotone





Left: residual. Right: quadratic function. Recall that: $f(x) = \frac{1}{2}x^{\mathsf{T}}Ax - b^{\mathsf{T}}x$
Barzilai-Borwein: non-monotone. Dai-Yuan: monotone:

$$\alpha_n^{\mathsf{DY}} = \begin{cases} \alpha_n^{\mathsf{SD}}, & 2 \\ \alpha_n^{\mathsf{Y}_0}, & 2 \end{cases}$$

(right column: number of runs)

Monotone methods may oscillate in residual curves

Experimental results: Spectral behavior

$A = \mathrm{diag}(1,\ 10,\ 100,\ 1000)$
$x_* = (1,\ 1,\ 1,\ 1)^\intercal \quad x_0 = 0$

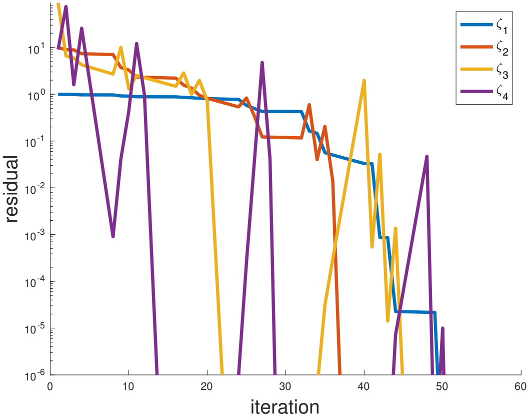
Use MGC:
$d_1 = 4,\ d_2 = 4 \implies$

$$\alpha_n^{\mathsf{MGC}} = \begin{cases} \alpha_n^{\mathsf{MG}}, & 4 \\ \alpha_n^{\mathsf{Y}_1}, & 1 \\ \alpha_{n-1}^{\mathsf{MGC}}, & 3 \end{cases}$$

Recall that:

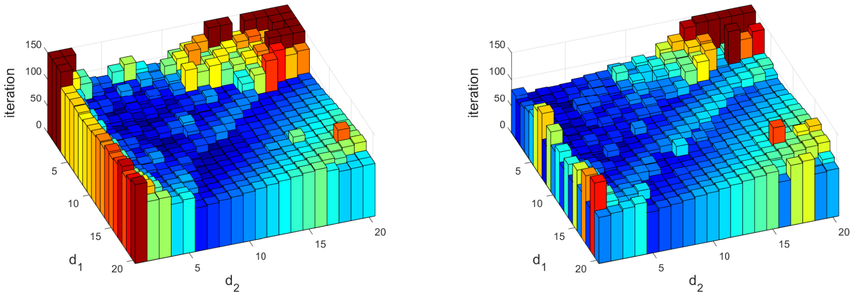
$$\lim_{n\to\infty} \alpha_n^{\mathsf{Y}_1} = 1/\lambda_N$$

$$\zeta_{i,n+1} = (1 - \alpha_n^{\mathsf{MGC}}\lambda_i)\zeta_{i,n}$$



ζ_4 decreases, then ζ_3, ζ_2, ζ_1 decrease in turn

Experimental results: Impact of parameters



Left: AOA ($\theta = 0.5$). Right: MGC

A: generated by MATLAB function `sprandsym`, $N = 100$, $\kappa = 100$

x_* generated randomly from $(-10, 10)$ $\quad b = Ax_*$

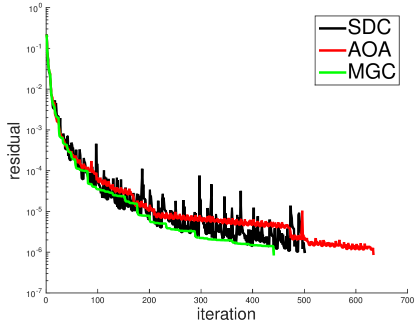
$x_0 = 0 \quad \|g_n\| < 10^{-6}\,\|g_0\|$ (default hereafter)

- small d_2: quasi-monotone, no benefit
- small d_1 and large d_2: oscillating, loss of precision

Experimental results: Comparison of alignment methods

Conditioning	SDA	SDC	AOA	MGA	MGC
$\kappa = 10^2$	70	76	80	74	75
$\kappa = 10^3$	194	182	227	209	190
$\kappa = 10^4$	619	475	547	515	488
$\kappa = 10^5$	1381	1273	1490	1321	1251

$d_1 = 4,\ d_2 = 4$
Avg. results: 10 tests
Random problems
$N = 10^3$

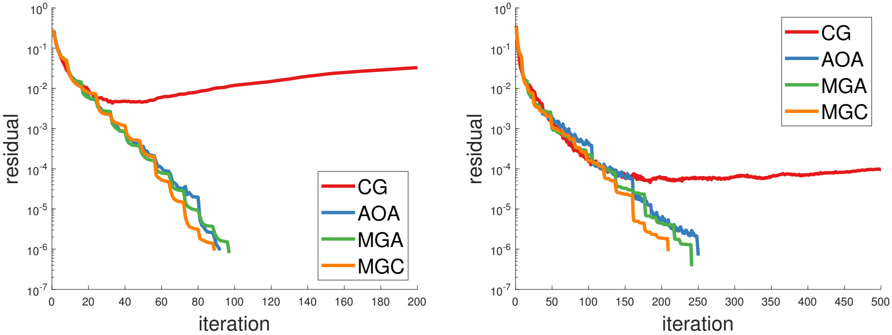


SDA: SD+A_0+retards
SDC: SD+Y_0+retards
AOA: AO+0.5AO+retards
MGA: MG+A_1+retards
MGC: MG+Y_1+retards

$A = \text{tridiag}(-\frac{1}{h^2}, \frac{2}{h^2}, -\frac{1}{h^2})$
$h = 11/N,\ N = 10^5$

SDC and MGC are the best

Experimental results: Comparison of alignment methods



Comparison with conjugate gradient: with perturbation

$\tilde{A}x = b \quad \tilde{A} = A + \delta V \quad \delta = 1/\kappa$

A: generated by `sprandsym`. V: generated by `sprand` (nonsymmetric)

$N = 100$. Left: $\kappa = 10^2$. Right: $\kappa = 10^3$

Conjugate gradient is sensitive to perturbation, our methods not

Experimental results: Comparison of alignment methods

Comparison with conjugate gradient: large-scale

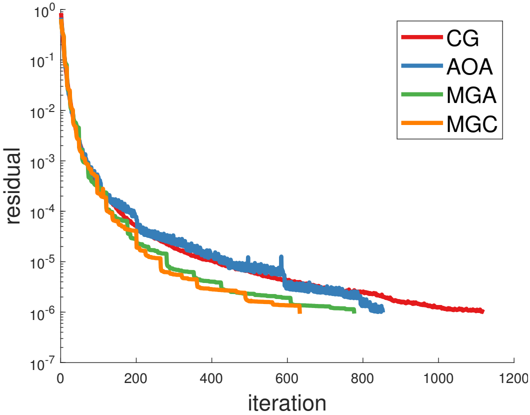
A, b: drawn from SuiteSparse (ID: 2544)

3D mechanical problem

$N = 1564794$
Nonzeros $= 114165372$

$\lambda_1 = 4.542 \times 10^{-1}$
$\lambda_N = 5.566 \times 10^{7}$

$\kappa = 1.225 \times 10^{8}$



Our methods work well for the large-scale problem

MGC is the best

Experimental results: Application to splitting methods

Parameter estimation by

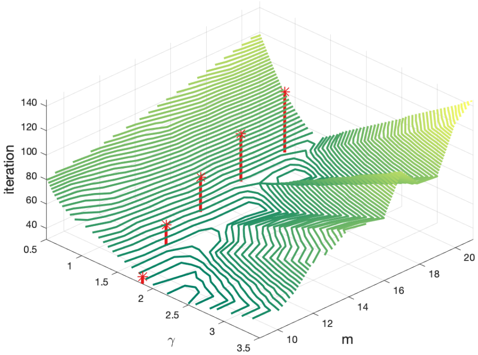
$$\gamma_* = \lim_{n\to\infty} \sqrt{\Gamma_{\rho,n}}$$

Consider equation below on unit cube with $\theta > 0$

Finite difference on $m \times m \times m$ grid

$b \in (-10,\, 10)+\iota(-10,\, 10)$

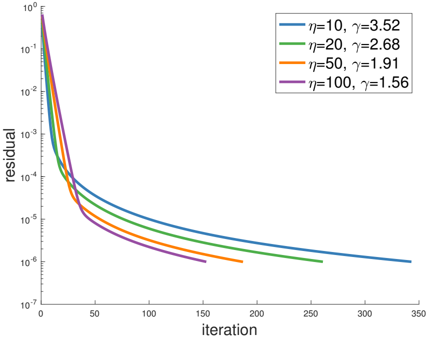
Red lines: optimal parameters for different m

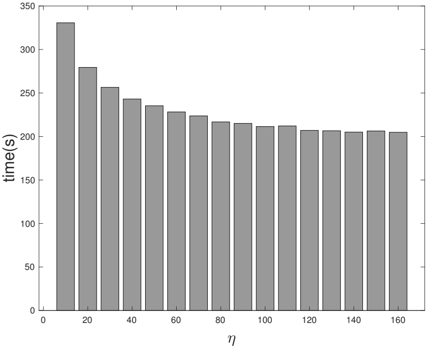


$$-\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}\right)+\theta\left(\frac{\partial u}{\partial x}+\frac{\partial u}{\partial y}+\frac{\partial u}{\partial z}\right)=q$$

γ_* is a good estimation, leading to the fast algorithm of the splitting method

Experimental results: Application to splitting methods





η: number of iterations of steepest descent before the splitting method

Left: convergence with different η. Right: total time (avg. of 10 tests)

Total time $=$ time of steepest descent $+$ time of splitting iterations

$m = 128 \implies N = 2097152$

Spectral properties of steepest descent accelerate the splitting method

Experimental results: Cyclic s-dimensional steepest descent

Random problem

$N = 100, \ \kappa = 500$

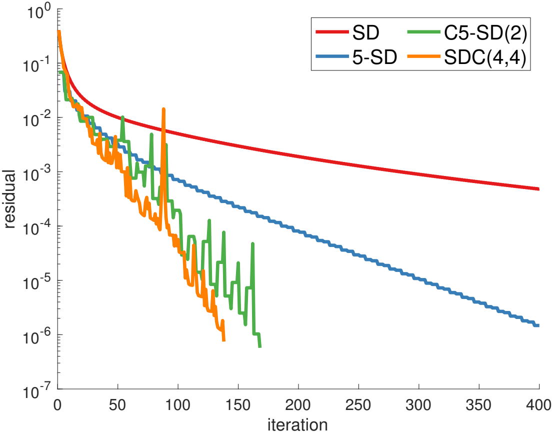
$x_* \in (-10, \ 10)$

Recall C5-SD(2):

$$x_{n+1} = x_n - \sum_{i=1}^{5} \alpha_n^{(i)} A^{i-1} g_n$$

$$\hat{\alpha} = (\omega_n^{(0)})/(\omega_n^{(1)})$$

$$x_{n+2} = x_{n+1} - \hat{\alpha} g_{n+1}$$



Cyclic version C5-SD(2) is much faster than the original version 5-SD

Experimental results: Parallel computing

	64 processors			128 processors		
method	iter	time(s)	residual	iter	time(s)	residual
BB	720	23.649	7.9×10^{-7}	625	16.246	8.9×10^{-7}
CSD(4)	818	19.672	8.5×10^{-7}	709	13.594	8.4×10^{-7}
SDC(4,4)	453	14.503	9.4×10^{-7}	445	11.796	9.1×10^{-7}
C5-SD(4)	768	22.925	9.2×10^{-7}	545	14.002	9.7×10^{-7}
	256 processors			512 processors		
method	iter	time(s)	residual	iter	time(s)	residual
BB	601	14.179	9.4×10^{-7}	762	18.017	8.0×10^{-7}
CSD(4)	569	9.591	8.9×10^{-7}	598	9.844	8.0×10^{-7}
SDC(4,4)	429	10.505	8.9×10^{-7}	429	10.269	9.8×10^{-7}
C5-SD(4)	671	14.998	8.0×10^{-7}	700	15.522	9.9×10^{-7}

Concluding remarks

Concluding remarks: Summary of gradient methods

Taxonomy:

- basic methods = monotone (monotone + non-alternate)
- gradient methods with retards = basic methods + retards
- alignment methods = asymptotic properties + retards
- fast methods = GMRs + alignment methods + special cases
- special cases: Dai-Yuan, adaptive methods, LMSD, etc.
- cyclic methods: gradient methods with the steplengths $\alpha_n = \alpha_{n-1}$

Effective gradient methods:

- never use basic methods as solvers, but for asymptotic properties
- alignment methods perform better than classical GMRs
- alignment methods are competitive with conjugate gradient
- GMRs are competitive with conjugate gradient in low precision
- GMRs and alignment methods are less sensitive to perturbation
- cyclic methods perform well in parallel computing

Concluding remarks: Contribution

Primary contributions (gradient methods):

We have proposed a cyclic gradient method:

- Q. Zou, F. Magoulès, A new cyclic gradient method adapted to large-scale linear systems, in Proceedings of 17th DCABES, 2018

We have proposed several spectral properties and three alignment methods:

- Q. Zou, F. Magoulès, Gradient methods with alignment for linear systems without Cauchy step, in progress

We have proposed several approaches to estimate parameter in the HSS method:

- Q. Zou, F. Magoulès, Parameter estimation in HSS method using gradient iterations, in progress

We have proposed cyclic s-dimensional gradient methods and more properties:

- Q. Zou, F. Magoulès, Reducing the effect of global synchronization in delayed gradient methods for symmetric linear systems, in progress
- Q. Zou, F. Magoulès, Parallel iterative methods with retards for linear systems, in Proceedings of 6th PARENG, 2019

Concluding remarks: Contribution

Other contributions (asynchronous iterative methods):

We have formalized the asynchronous Laplace transform method:

- F. Magoulès, Q. Zou, Asynchronous Time-Parallel Method based on Laplace Transform, Int. J. Comput. Math., in review

We have implemented asynchronous convergence detection using modified recursive doubling:

- Q. Zou, F. Magoulès, Convergence detection of asynchronous iterations based on modified recursive doubling, in Proceedings of 17th DCABES, 2018

We have applied asynchronous Parareal method to the Black-Scholes equation:

- F. Magoulès, G. Gbikpi-Benissan, Q. Zou, Asynchronous iterations of Parareal algorithm for option pricing models, Mathematics, 2018
- Q. Zou, G. Gbikpi-Benissan, F. Magoulès, Asynchronous Parareal algorithm applied to European option pricing, in Proceedings of 16th DCABES, 2017
- Q. Zou, G. Gbikpi-Benissan, F. Magoulès, Asynchronous communications library for the parallel-in-time solution of Black-Scholes Equation, in Proceedings of 16th DCABES, 2017

Concluding remarks: Future work

Gradient methods:

- Convergence rate analysis of lagged gradient methods
- Lagged gradient methods for unconstrained and constrained optimization problems
- Limited memory gradient methods

Matrix analysis:

- Multiple right-hand sides and linear matrix equations, tensor techniques
- Structured matrices, matrix polynomials, matrix functions
- Low-rank approximation, data science

Parallel computing:

- Asynchronous iterative methods
- Communication-avoiding methods and preconditioners

Other active areas ...