Composing Stepsize Schedules for Minimizing Smooth Convex Functions

Alex L. Wang, Purdue University

Joint work with:
Benjamin Grimmer, Johns Hopkins University
Kevin Shu, California Institute of Technology

ICCOPT, July 2025

Smooth Convex Minimization and Gradient Descent

- Want to minimize $f : \mathbb{R}^d \to \mathbb{R}$ assuming

Smooth Convex Minimization and Gradient Descent

- Want to minimize $f : \mathbb{R}^d \to \mathbb{R}$ assuming
  - f is convex

Smooth Convex Minimization and Gradient Descent

- Want to minimize $f : \mathbb{R}^d \to \mathbb{R}$ assuming
  - f is convex
  - f is 1-smooth: $\|\nabla f(x) - \nabla f(y)\| \leq \|x - y\| \quad \forall x, y$

Smooth Convex Minimization and Gradient Descent

- Want to minimize $f : \mathbb{R}^d \to \mathbb{R}$ assuming
  - f is convex
  - f is 1-smooth: $\|\nabla f(x) - \nabla f(y)\| \leq \|x - y\| \quad \forall x, y$
  - f has a minimizer $x_\star$ with minimum value $f_\star$

Smooth Convex Minimization and Gradient Descent

- Want to minimize $f : \mathbb{R}^d \to \mathbb{R}$ assuming
  - f is convex
  - f is 1-smooth: $\|\nabla f(x) - \nabla f(y)\| \leq \|x - y\| \quad \forall x, y$
  - f has a minimizer $x_\star$ with minimum value $f_\star$
- **Gradient descent method**: Given instance (f, x_0), repeatedly set

$$x_{i+1} = x_i - h_i \nabla f(x_i)$$

for a vector of fixed stepsizes $\mathbf{h} = [h_0, h_1, \ldots, h_{N-1}]$

Smooth Convex Minimization and Gradient Descent

- Want to minimize $f : \mathbb{R}^d \to \mathbb{R}$ assuming
  - f is convex
  - f is 1-smooth: $\|\nabla f(x) - \nabla f(y)\| \leq \|x - y\| \quad \forall x, y$
  - f has a minimizer $x_\star$ with minimum value $f_\star$
- **Gradient descent method**: Given instance (f, x_0), repeatedly set

$$x_{i+1} = x_i - h_i \nabla f(x_i)$$

  for a vector of fixed stepsizes $\mathbf{h} = [h_0, h_1, \ldots, h_{N-1}]$
- **Goal**: Pick stepsizes $\mathbf{h}$ to minimize

$$\text{Performance metric :} \qquad \sup_{(f, x_0)} \frac{f_N - f_\star}{\frac{1}{2}\|x_0 - x_\star\|^2}$$

Smooth Convex Minimization and Gradient Descent

- Want to minimize $f : \mathbb{R}^d \to \mathbb{R}$ assuming
  - f is convex
  - f is 1-smooth: $\|\nabla f(x) - \nabla f(y)\| \leq \|x - y\| \quad \forall x, y$
  - f has a minimizer $x_\star$ with minimum value $f_\star$
- **Gradient descent method**: Given instance (f, x_0), repeatedly set

$$x_{i+1} = x_i - h_i \nabla f(x_i)$$

  for a vector of fixed stepsizes $\mathbf{h} = [h_0, h_1, \ldots, h_{N-1}]$
- **Goal**: Pick stepsizes $\mathbf{h}$ to minimize

$$\text{Performance metric :} \qquad \sup_{(f,x_0)} \frac{f_N - f_\star}{\frac{1}{2}\|x_0 - x_\star\|^2}$$

- Notation: $f_i = f(x_i)$ and $g_i = \nabla f(x_i)$

Review of Stepsize Choices: Short Stepsizes

- Per-iteration descent $f_{i+1} < f_i$ is guaranteed if $h_i \in (0, 2)$

Review of Stepsize Choices: Short Stepsizes

- Per-iteration descent $f_{i+1} < f_i$ is guaranteed if $h_i \in (0, 2)$
- **[Folklore]**: If we take $\mathbf{h} = \mathbf{1}_N$, then

$$\sup_{(f, x_0)} \frac{f_N - f_\star}{\frac{1}{2}\left\|x_0 - x_\star\right\|^2} \leq \frac{1}{2N+1}$$

Review of Stepsize Choices: Short Stepsizes

- Per-iteration descent $f_{i+1} < f_i$ is guaranteed if $h_i \in (0,2)$
- **[Folklore]**: If we take $\mathbf{h} = \mathbf{1}_N$, then

$$\sup_{(f,x_0)} \frac{f_N - f_\star}{\frac{1}{2}\left\|x_0 - x_\star\right\|^2} \leq \frac{1}{2N+1}$$

- **[Drori, Teboulle '12] [Taylor, Hendrickx, Glineur '15]**: Performance Estimation Programming (PEP)

Review of Stepsize Choices: Short Stepsizes

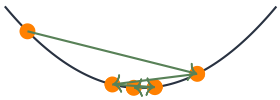
- Per-iteration descent $f_{i+1} < f_i$ is guaranteed if $h_i \in (0, 2)$
- [Folklore]: If we take $\mathbf{h} = \mathbf{1}_N$, then

$$\sup_{(f,x_0)} \frac{f_N - f_\star}{\frac{1}{2}\|x_0 - x_\star\|^2} \leq \frac{1}{2N+1}$$

- [Drori, Teboulle '12] [Taylor, Hendrickx, Glineur '15]: Performance Estimation Programming (PEP)
- [Drori, Teboulle '12] [Rotaru, Glineur, Patrinos '24] [Kim '24]: If $\alpha \in (0, 2)$ and we take $\mathbf{h} = \alpha \mathbf{1}_N$, then

$$\sup_{(f,x_0)} \frac{f_N - f_\star}{\frac{1}{2}\|x_0 - x_\star\|^2} = \max\left((\alpha - 1)^{2N}, \frac{1}{1 + 2N\alpha}\right)$$

Review of Stepsize Choices: Short Stepsizes



$$x_N = \prod_i (1-h_i)x_0$$

$$\frac{f_N - f_\star}{\frac{1}{2}\|x_0 - x_\star\|^2} = \prod_i (h_i - 1)^2$$

- Per-iteration descent $f_{i+1} < f_i$ is guaranteed if $h_i \in (0,2)$
- **[Folklore]**: If we take $\mathbf{h} = \mathbf{1}_N$, then

$$\sup_{(f,x_0)} \frac{f_N - f_\star}{\frac{1}{2}\|x_0 - x_\star\|^2} \leq \frac{1}{2N+1}$$

- **[Drori, Teboulle '12] [Taylor, Hendrickx, Glineur '15]**: Performance Estimation Programming (PEP)
- **[Drori, Teboulle '12] [Rotaru, Glineur, Patrinos '24] [Kim '24]**: If $\alpha \in (0,2)$ and we take $\mathbf{h} = \alpha\mathbf{1}_N$, then

$$\sup_{(f,x_0)} \frac{f_N - f_\star}{\frac{1}{2}\|x_0 - x_\star\|^2} = \max\left((\alpha-1)^{2N}, \frac{1}{1+2N\alpha}\right)$$

Review of Stepsize Choices: Short Stepsizes

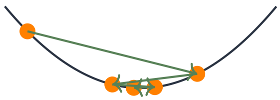
- Per-iteration descent $f_{i+1} < f_i$ is guaranteed if $h_i \in (0, 2)$
- **[Folklore]**: If we take $\mathbf{h} = \mathbf{1}_N$, then

$$\sup_{(f,x_0)} \frac{f_N - f_\star}{\frac{1}{2}\|x_0 - x_\star\|^2} \le \frac{1}{2N+1}$$

- **[Drori, Teboulle '12] [Taylor, Hendrickx, Glineur '15]**:
  Performance Estimation Programming (PEP)
- **[Drori, Teboulle '12] [Rotaru, Glineur, Patrinos '24] [Kim '24]**:
  If $\alpha \in (0, 2)$ and we take $\mathbf{h} = \alpha \mathbf{1}_N$, then

$$\sup_{(f,x_0)} \frac{f_N - f_\star}{\frac{1}{2}\|x_0 - x_\star\|^2} = \max\left((\alpha - 1)^{2N}, \frac{1}{1 + 2N\alpha} \right)$$

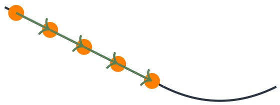
$f(x) = \frac{1}{2}x^2$



$$x_N = \prod_i (1 - h_i) x_0$$

$$\frac{f_N - f_\star}{\frac{1}{2}\|x_0 - x_\star\|^2} = \prod_i (h_i - 1)^2$$

$f(x) =$ Huber



$$x_N = x_0 + \delta \sum_i h_i$$

$$\frac{f_N - f_\star}{\frac{1}{2}\|x_0 - x_\star\|^2} = (1 + 2\sum_i h_i)^{-1}$$

Review of Stepsize Choices: Short Stepsizes

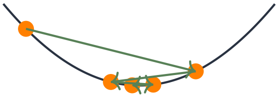
- Per-iteration descent $f_{i+1} < f_i$ is guaranteed if $h_i \in (0,2)$
- **[Folklore]**: If we take $\mathbf{h} = \mathbf{1}_N$, then

$$\sup_{(f,x_0)} \frac{f_N - f_\star}{\frac{1}{2}\|x_0 - x_\star\|^2} \leq \frac{1}{2N+1}$$

- **[Drori, Teboulle '12] [Taylor, Hendrickx, Glineur '15]**: Performance Estimation Programming (PEP)
- **[Drori, Teboulle '12] [Rotaru, Glineur, Patrinos '24] [Kim '24]**: If $\alpha \in (0,2)$ and we take $\mathbf{h} = \alpha \mathbf{1}_N$, then

$$\sup_{(f,x_0)} \frac{f_N - f_\star}{\frac{1}{2}\|x_0 - x_\star\|^2} = \max\left((\alpha - 1)^{2N}, \frac{1}{1 + 2N\alpha} \right) \approx \frac{1}{4N}$$

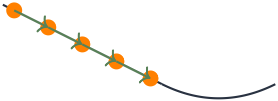
$f(x) = \frac{1}{2}x^2$



$$x_N = \prod_i (1 - h_i) x_0$$

$$\frac{f_N - f_\star}{\frac{1}{2}\|x_0 - x_\star\|^2} = \prod_i (h_i - 1)^2$$

$f(x) =$ Huber



$$x_N = x_0 + \delta \sum_i h_i$$

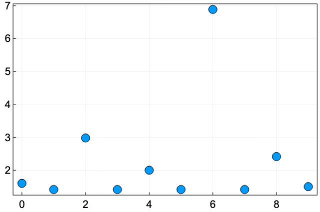
$$\frac{f_N - f_\star}{\frac{1}{2}\|x_0 - x_\star\|^2} = (1 + 2\sum_i h_i)^{-1}$$

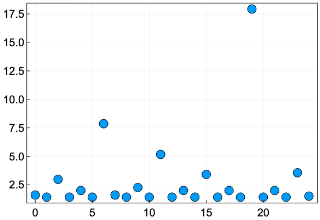
Review of Stepsize Choices: Long Stepsizes

- **[Altschuler '18] [Daccache 2019] [Eloi '22]**:
  Identification of optimal $\mathbf{h}$ for $N = 1, 2, 3$

Review of Stepsize Choices: Long Stepsizes

- **[Altschuler '18] [Daccache 2019] [Eloi '22]**:
  Identification of optimal $\mathbf{h}$ for $N = 1, 2, 3$
- **[Das Gupta, Van Parys, Ryu '23]**:
  Numerical identification of optimal* $\mathbf{h}$ for $N = 1, \ldots, 25$



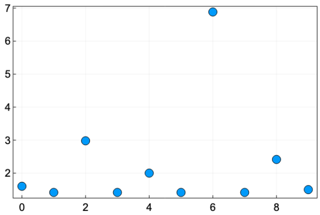


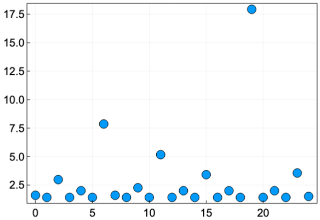
Review of Stepsize Choices: Long Stepsizes

- **[Altschuler '18] [Daccache 2019] [Eloi '22]**:
  Identification of optimal $\mathbf{h}$ for $N = 1, 2, 3$
- **[Das Gupta, Van Parys, Ryu '23]**:
  Numerical identification of optimal* $\mathbf{h}$ for $N = 1, \ldots, 25$
- **[Altschuler, Parrilo '23a,b] [Grimmer, Shu, Wang '24a,b] [Zhang, Jiang 2024]**:
  Identification of conjectured optimal $\mathbf{h}$ for all N

$$\sup_{(f, x_0)} \frac{f_N - f_\star}{\frac{1}{2}\|x_0 - x_\star\|^2} = O\left(\frac{1}{N^{1.2716}}\right)$$

where $1.2716 = \log_2(1 + \sqrt{2})$



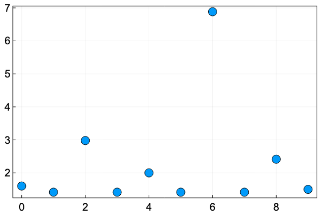


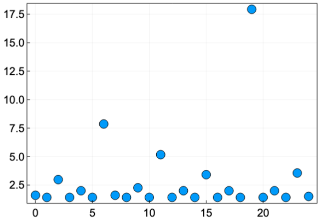
Review of Stepsize Choices: Long Stepsizes

- **[Altschuler '18] [Daccache 2019] [Eloi '22]**:
  Identification of optimal $\mathbf{h}$ for $N = 1, 2, 3$
- **[Das Gupta, Van Parys, Ryu '23]**:
  Numerical identification of optimal* $\mathbf{h}$ for
  $N = 1, \ldots, 25$
- **[Altschuler, Parrilo '23a,b] [Grimmer, Shu, Wang '24a,b] [Zhang, Jiang 2024]**:
  Identification of conjectured optimal $\mathbf{h}$ for all N

$$\sup_{(f, x_0)} \frac{f_N - f_\star}{\frac{1}{2}\|x_0 - x_\star\|^2} = O\left(\frac{1}{N^{1.2716}}\right)$$

  where $1.2716 = \log_2(1 + \sqrt{2})$
- **[Bok, Altschuler '24, '25] [Wang et al. '24] [Zhang et al. '24]**:
  Additional extensions





Definitions and Results

Composable Sequences

Definition

$\mathbf{h}$ is f-composable with rate η if

$$\sup_{(f,x_0)} \frac{f_N - f_\star}{\frac{1}{2}\left\|x_0 - x_\star\right\|^2} = \eta \quad \text{where } \eta = \frac{1}{1 + 2\sum_i h_i} = \prod_i (h_i - 1)^2$$

Composable Sequences

Definition

$\mathbf{h}$ is f-composable with rate η if

$$\sup_{(f,x_0)} \frac{f_N - f_\star}{\frac{1}{2}\left\|x_0 - x_\star\right\|^2} = \eta \quad \text{where } \eta = \frac{1}{1 + 2\sum_i h_i} = \prod_i (h_i - 1)^2$$

Definition

$\mathbf{h}$ is s-composable with rate η if

$$\sup_{(f,x_0)} \frac{f_N - f_\star}{\frac{1}{2}\left\|x_0 - x_\star\right\|^2 - \frac{1}{2}\left\|x_N - \frac{g_N}{\eta} - x_\star\right\|^2} = \frac{\eta}{2-\eta} \quad \text{where } \eta = \frac{1}{1 + \sum_i h_i} = \prod_i (h_i - 1)$$

Composable Sequences

Definition

$\mathbf{h}$ is f-composable with rate η if

$$\sup_{(f,x_0)} \frac{f_N - f_\star}{\frac{1}{2}\left\|x_0 - x_\star\right\|^2} = \eta \quad \text{where } \eta = \frac{1}{1 + 2\sum_i h_i} = \prod_i (h_i - 1)^2$$

Definition

$\mathbf{h}$ is s-composable with rate η if

$$\sup_{(f,x_0)} \frac{f_N - f_\star}{\frac{1}{2}\left\|x_0 - x_\star\right\|^2 - \frac{1}{2}\left\|x_N - \frac{g_N}{\eta} - x_\star\right\|^2} = \frac{\eta}{2-\eta} \quad \text{where } \eta = \frac{1}{1 + \sum_i h_i} = \prod_i (h_i - 1)$$

Example: $[\,]$ is f-composable and s-composable with rate 1

Composing to get s-Composable Sequences

Theorem

Suppose

- $\mathbf{a}$ is s-composable with rate α
- $\mathbf{b}$ is s-composable with rate β

Composing to get s-Composable Sequences

Theorem

Suppose

- $\mathbf{a}$ is s-composable with rate α
- $\mathbf{b}$ is s-composable with rate β

Define the s-join

$$\mathbf{a} \bowtie \mathbf{b} := [\mathbf{a}, \mu, \mathbf{b}]$$

where $\mu = 1 + \frac{\sqrt{\alpha^2+6\alpha\beta+\beta^2}-\alpha-\beta}{2\alpha\beta}$.

Composing to get s-Composable Sequences

Theorem

Suppose

- $\mathbf{a}$ is s-composable with rate α
- $\mathbf{b}$ is s-composable with rate β

Define the s-join

$$\mathbf{a} \bowtie \mathbf{b} := [\mathbf{a}, \mu, \mathbf{b}]$$

where $\mu = 1 + \frac{\sqrt{\alpha^2+6\alpha\beta+\beta^2}-\alpha-\beta}{2\alpha\beta}$. Then $\mathbf{a} \bowtie \mathbf{b}$ is s-composable with rate

$$\alpha \bowtie \beta := \frac{2\alpha\beta}{\alpha + \beta + \sqrt{\alpha^2 + 6\alpha\beta + \beta^2}}$$

Composing to get s-Composable Sequences

Theorem

Suppose

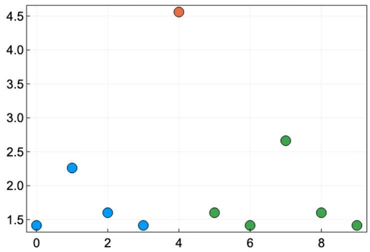
- $\mathbf{a}$ is s-composable with rate α
- $\mathbf{b}$ is s-composable with rate β

Define the s-join

$$\mathbf{a} \bowtie \mathbf{b} := [\mathbf{a}, \mu, \mathbf{b}]$$

where $\mu = 1 + \frac{\sqrt{\alpha^2+6\alpha\beta+\beta^2}-\alpha-\beta}{2\alpha\beta}$. Then $\mathbf{a} \bowtie \mathbf{b}$ is s-composable with rate

$$\alpha \bowtie \beta := \frac{2\alpha\beta}{\alpha + \beta + \sqrt{\alpha^2 + 6\alpha\beta + \beta^2}}$$



Composing to get s-Composable Sequences

Theorem

Suppose

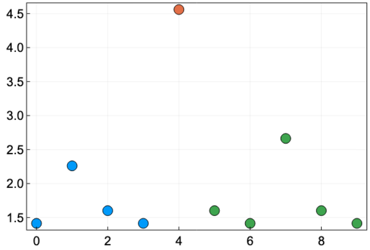
- $\mathbf{a}$ is s-composable with rate α
- $\mathbf{b}$ is s-composable with rate β

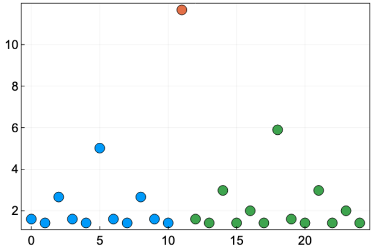
Define the s-join

$$\mathbf{a} \bowtie \mathbf{b} := [\mathbf{a}, \mu, \mathbf{b}]$$

where $\mu = 1 + \frac{\sqrt{\alpha^2+6\alpha\beta+\beta^2}-\alpha-\beta}{2\alpha\beta}$. Then $\mathbf{a} \bowtie \mathbf{b}$ is s-composable with rate

$$\alpha \bowtie \beta := \frac{2\alpha\beta}{\alpha + \beta + \sqrt{\alpha^2 + 6\alpha\beta + \beta^2}}$$





Composing to get f-Composable Sequences

Theorem

Suppose

- $\mathbf{a}$ is s-composable with rate α
- $\mathbf{b}$ is f-composable with rate β

Composing to get f-Composable Sequences

Theorem

Suppose

- $\mathbf{a}$ is s-composable with rate α
- $\mathbf{b}$ is f-composable with rate β

Define the f-join

$$\mathbf{a} \triangleright \mathbf{b} := [\mathbf{a}, \mu, \mathbf{b}]$$

where $\mu = 1 + \frac{\sqrt{\alpha^2+8\alpha\beta}-\alpha}{4\alpha\beta}$.

Composing to get f-Composable Sequences

Theorem

Suppose

- $\mathbf{a}$ is s-composable with rate α
- $\mathbf{b}$ is f-composable with rate β

Define the f-join

$$\mathbf{a} \triangleright \mathbf{b} := [\mathbf{a}, \mu, \mathbf{b}]$$

where $\mu = 1 + \frac{\sqrt{\alpha^2+8\alpha\beta}-\alpha}{4\alpha\beta}$. Then $\mathbf{a} \triangleright \mathbf{b}$ is f-composable with rate

$$\alpha \triangleright \beta := \frac{2\alpha\beta}{\alpha + 4\beta + \sqrt{\alpha^2 + 8\alpha\beta}}$$

Composing to get f-Composable Sequences

Theorem

Suppose

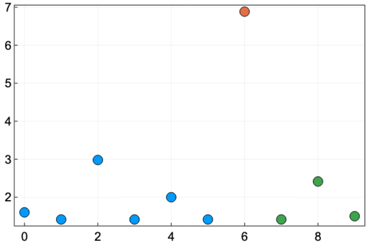
- $\mathbf{a}$ is s-composable with rate α
- $\mathbf{b}$ is f-composable with rate β

Define the f-join

$$\mathbf{a} \triangleright \mathbf{b} := [\mathbf{a}, \mu, \mathbf{b}]$$

where $\mu = 1 + \frac{\sqrt{\alpha^2+8\alpha\beta}-\alpha}{4\alpha\beta}$. Then $\mathbf{a} \triangleright \mathbf{b}$ is f-composable with rate

$$\alpha \triangleright \beta := \frac{2\alpha\beta}{\alpha + 4\beta + \sqrt{\alpha^2 + 8\alpha\beta}}$$



Composing to get f-Composable Sequences

Theorem

Suppose

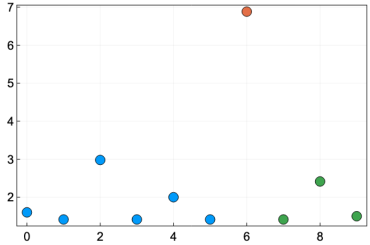
- $\mathbf{a}$ is s-composable with rate α
- $\mathbf{b}$ is f-composable with rate β

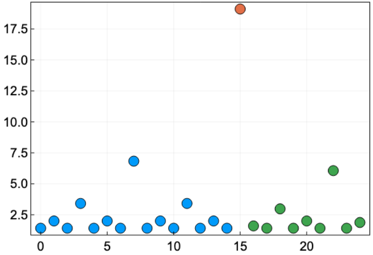
Define the f-join

$$\mathbf{a} \triangleright \mathbf{b} := [\mathbf{a}, \mu, \mathbf{b}]$$

where $\mu = 1 + \frac{\sqrt{\alpha^2+8\alpha\beta}-\alpha}{4\alpha\beta}$. Then $\mathbf{a} \triangleright \mathbf{b}$ is f-composable with rate

$$\alpha \triangleright \beta := \frac{2\alpha\beta}{\alpha + 4\beta + \sqrt{\alpha^2 + 8\alpha\beta}}$$





Composing to get f-Composable Sequences

Theorem

Suppose

- $\mathbf{a}$ is s-composable with rate α
- $\mathbf{b}$ is f-composable with rate β

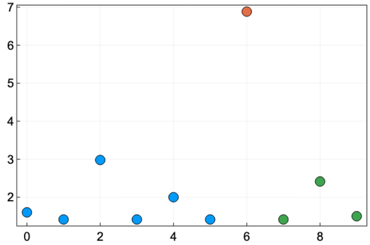
Define the f-join

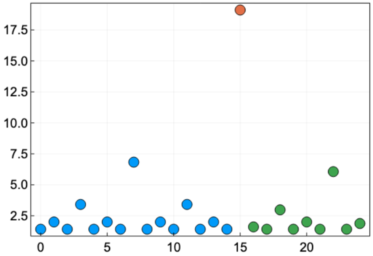
$$\mathbf{a} \triangleright \mathbf{b} := [\mathbf{a}, \mu, \mathbf{b}]$$

where $\mu = 1 + \frac{\sqrt{\alpha^2+8\alpha\beta}-\alpha}{4\alpha\beta}$. Then $\mathbf{a} \triangleright \mathbf{b}$ is f-composable with rate

$$\alpha \triangleright \beta := \frac{2\alpha\beta}{\alpha + 4\beta + \sqrt{\alpha^2 + 8\alpha\beta}}$$

$\mathbf{h}$ is basic if it can be constructed from $[\,]$, $\bowtie$, $\triangleright$





Applications: Silver Stepsizes

- The silver stepsize schedules of **[Altschuler, Parrilo '23a,b]**:

$$
\begin{aligned}
\mathbf{h}^{(0)} &= [\,] \\
\mathbf{h}^{(1)} &= [\sqrt{2}] \\
\mathbf{h}^{(2)} &= [\sqrt{2}, 2, \sqrt{2}] \\
\mathbf{h}^{(3)} &= [\sqrt{2}, 2, \sqrt{2}, 2+\sqrt{2}, \sqrt{2}, 2, \sqrt{2}] \\
&\dots
\end{aligned}
$$

Applications: Silver Stepsizes

- The silver stepsize schedules of **[Altschuler, Parrilo '23a,b]**:

$$
\begin{aligned}
\mathbf{h}^{(0)} &= [\,] && = [\,] \\
\mathbf{h}^{(1)} &= [\sqrt{2}] && = \mathbf{h}^{(0)} \bowtie \mathbf{h}^{(0)} \\
\mathbf{h}^{(2)} &= [\sqrt{2}, 2, \sqrt{2}] && = \mathbf{h}^{(1)} \bowtie \mathbf{h}^{(1)} \\
\mathbf{h}^{(3)} &= [\sqrt{2}, 2, \sqrt{2}, 2+\sqrt{2}, \sqrt{2}, 2, \sqrt{2}] && = \mathbf{h}^{(2)} \bowtie \mathbf{h}^{(2)} \\
&\dots
\end{aligned}
$$

Applications: Silver Stepsizes

- The silver stepsize schedules of [Altschuler, Parrilo '23a,b]:

$$
\begin{aligned}
\mathbf{h}^{(0)} &= [\,] && = [\,] \\
\mathbf{h}^{(1)} &= \left[\sqrt{2}\right] && = \mathbf{h}^{(0)} \bowtie \mathbf{h}^{(0)} \\
\mathbf{h}^{(2)} &= \left[\sqrt{2}, 2, \sqrt{2}\right] && = \mathbf{h}^{(1)} \bowtie \mathbf{h}^{(1)} \\
\mathbf{h}^{(3)} &= \left[\sqrt{2}, 2, \sqrt{2}, 2+\sqrt{2}, \sqrt{2}, 2, \sqrt{2}\right] && = \mathbf{h}^{(2)} \bowtie \mathbf{h}^{(2)} \\
&\dots
\end{aligned}
$$

- The sequence $\mathbf{h}^{(k)}$ is basic s-composable with rate $\eta^{(k)} = (1+\sqrt{2})^{-k}$

$$
\sup_{(f,x_0)} \frac{f_{2^k-1} - f_\star}{\frac{1}{2}\left\|x_0 - x_\star\right\|^2} = \frac{\eta^{(k)}}{2-\eta^{(k)}} \approx \frac{1}{2(1+\sqrt{2})^k}
$$

Applications: Numerically Optimal Schedules

- Numerically optimal stepsize schedules of **[Das Gupta, Van Parys, Ryu '23]**:

$$
\begin{aligned}
\mathbf{h}^{(0)} &= [\,] \\
\mathbf{h}^{(1)} &= [3/2] \\
\mathbf{h}^{(2)} &= [1.4142, 1.8767] \\
\mathbf{h}^{(3)} &= [1.4142, 2.4142, 1.5] \\
\mathbf{h}^{(4)} &= [1.4142, 1.6012, 3.0051, 1.5] \\
\mathbf{h}^{(5)} &= [1.4142, 2, 1.4142, 3.5576, 1.5] \\
&\dots
\end{aligned}
$$

Applications: Numerically Optimal Schedules

- Numerically optimal stepsize schedules of **[Das Gupta, Van Parys, Ryu '23]**:

$$
\begin{aligned}
\mathbf{h}^{(0)} &= [\,] && = [\,] \\
\mathbf{h}^{(1)} &= [3/2] && = [\,] \triangleright \mathbf{h}^{(0)} \\
\mathbf{h}^{(2)} &= [1.4142, 1.8767] && = ([\,] \bowtie [\,]) \triangleright \mathbf{h}^{(0)} \\
\mathbf{h}^{(3)} &= [1.4142, 2.4142, 1.5] && = ([\,] \bowtie [\,]) \triangleright \mathbf{h}^{(1)} \\
\mathbf{h}^{(4)} &= [1.4142, 1.6012, 3.0051, 1.5] && = (([\,] \bowtie [\,]) \bowtie [\,]) \triangleright \mathbf{h}^{(1)} \\
\mathbf{h}^{(5)} &= [1.4142, 2, 1.4142, 3.5576, 1.5] && = (([\,] \bowtie [\,]) \bowtie ([\,] \bowtie [\,])) \triangleright \mathbf{h}^{(1)} \\
&\dots
\end{aligned}
$$

Applications: Numerically Optimal Schedules

- Numerically optimal stepsize schedules of **[Das Gupta, Van Parys, Ryu '23]**:

$$
\begin{aligned}
\mathbf{h}^{(0)} &= [\,] && = [\,] \\
\mathbf{h}^{(1)} &= [3/2] && = [\,] \triangleright \mathbf{h}^{(0)} \\
\mathbf{h}^{(2)} &= [1.4142, 1.8767] && = ([\,] \bowtie [\,]) \triangleright \mathbf{h}^{(0)} \\
\mathbf{h}^{(3)} &= [1.4142, 2.4142, 1.5] && = ([\,] \bowtie [\,]) \triangleright \mathbf{h}^{(1)} \\
\mathbf{h}^{(4)} &= [1.4142, 1.6012, 3.0051, 1.5] && = (([\,] \bowtie [\,]) \bowtie [\,]) \triangleright \mathbf{h}^{(1)} \\
\mathbf{h}^{(5)} &= [1.4142, 2, 1.4142, 3.5576, 1.5] && = (([\,] \bowtie [\,]) \bowtie ([\,] \bowtie [\,])) \triangleright \mathbf{h}^{(1)} \\
&\dots
\end{aligned}
$$

  - All $\mathbf{h}^{(n)}$ for $n = 1, \dots, 25$ are basic f-composable

Applications: Numerically Optimal Schedules

- Numerically optimal stepsize schedules of **[Das Gupta, Van Parys, Ryu '23]**:

$$
\begin{aligned}
\mathbf{h}^{(0)} &= [\,] && = [\,] \\
\mathbf{h}^{(1)} &= [3/2] && = [\,] \triangleright \mathbf{h}^{(0)} \\
\mathbf{h}^{(2)} &= [1.4142, 1.8767] && = ([\,] \bowtie [\,]) \triangleright \mathbf{h}^{(0)} \\
\mathbf{h}^{(3)} &= [1.4142, 2.4142, 1.5] && = ([\,] \bowtie [\,]) \triangleright \mathbf{h}^{(1)} \\
\mathbf{h}^{(4)} &= [1.4142, 1.6012, 3.0051, 1.5] && = (([\,] \bowtie [\,]) \bowtie [\,]) \triangleright \mathbf{h}^{(1)} \\
\mathbf{h}^{(5)} &= [1.4142, 2, 1.4142, 3.5576, 1.5] && = (([\,] \bowtie [\,]) \bowtie ([\,] \bowtie [\,])) \triangleright \mathbf{h}^{(1)} \\
&\dots
\end{aligned}
$$

  - All $\mathbf{h}^{(n)}$ for $n = 1, \dots, 25$ are basic f-composable
  - Most $\mathbf{h}^{(n)}$ actually $\approx 1\%$ *suboptimal* for $n = 6, \dots, 25$

Applications: Numerically Optimal Schedules

- Numerically optimal stepsize schedules of [Das Gupta, Van Parys, Ryu '23]:

$$
\begin{aligned}
\mathbf{h}^{(0)} &= [\,] && = [\,] \\
\mathbf{h}^{(1)} &= [3/2] && = [\,] \triangleright \mathbf{h}^{(0)} \\
\mathbf{h}^{(2)} &= [1.4142, 1.8767] && = ([\,] \bowtie [\,]) \triangleright \mathbf{h}^{(0)} \\
\mathbf{h}^{(3)} &= [1.4142, 2.4142, 1.5] && = ([\,] \bowtie [\,]) \triangleright \mathbf{h}^{(1)} \\
\mathbf{h}^{(4)} &= [1.4142, 1.6012, 3.0051, 1.5] && = (([\,] \bowtie [\,]) \bowtie [\,]) \triangleright \mathbf{h}^{(1)} \\
\mathbf{h}^{(5)} &= [1.4142, 2, 1.4142, 3.5576, 1.5] && = (([\,] \bowtie [\,]) \bowtie ([\,] \bowtie [\,])) \triangleright \mathbf{h}^{(1)} \\
&\dots
\end{aligned}
$$

  - All $\mathbf{h}^{(n)}$ for $n = 1, \dots, 25$ are basic f-composable
  - Most $\mathbf{h}^{(n)}$ actually $\approx 1\%$ *suboptimal* for $n = 6, \dots, 25$
- **Conjecture**: the local minimizers of

$$
\min_{\mathbf{h}} \sup_{(f, x_0)} \frac{f_N - f_\star}{\frac{1}{2}\left\|x_0 - x_\star\right\|^2}
$$

are basic f-composable schedules

Applications: Dynamic Short Stepsizes

- Dynamic short stepsize schedules:

$$
\begin{aligned}
\mathbf{h}^{(0)} &= [\,] \\
\mathbf{h}^{(1)} &= [1.4142] \\
\mathbf{h}^{(2)} &= [1.4142, 1.6012] \\
\mathbf{h}^{(3)} &= [1.4142, 1.6012, 1.7022] \\
\mathbf{h}^{(4)} &= [1.4142, 1.6012, 1.7022, 1.7642] \\
&\dots
\end{aligned}
$$

Applications: Dynamic Short Stepsizes

- Dynamic short stepsize schedules:

$$
\begin{aligned}
\mathbf{h}^{(0)} &= [\,] && = [\,] \\
\mathbf{h}^{(1)} &= [1.4142] && = \mathbf{h}^{(0)} \bowtie [\,] \\
\mathbf{h}^{(2)} &= [1.4142, 1.6012] && = \mathbf{h}^{(1)} \bowtie [\,] \\
\mathbf{h}^{(3)} &= [1.4142, 1.6012, 1.7022] && = \mathbf{h}^{(2)} \bowtie [\,] \\
\mathbf{h}^{(4)} &= [1.4142, 1.6012, 1.7022, 1.7642] && = \mathbf{h}^{(3)} \bowtie [\,] \\
& \dots
\end{aligned}
$$

Applications: Dynamic Short Stepsizes

- Dynamic short stepsize schedules:

$$
\begin{aligned}
\mathbf{h}^{(0)} &= [\,] && = [\,] \\
\mathbf{h}^{(1)} &= [1.4142] && = \mathbf{h}^{(0)} \bowtie [\,] \\
\mathbf{h}^{(2)} &= [1.4142, 1.6012] && = \mathbf{h}^{(1)} \bowtie [\,] \\
\mathbf{h}^{(3)} &= [1.4142, 1.6012, 1.7022] && = \mathbf{h}^{(2)} \bowtie [\,] \\
\mathbf{h}^{(4)} &= [1.4142, 1.6012, 1.7022, 1.7642] && = \mathbf{h}^{(3)} \bowtie [\,] \\
&\dots
\end{aligned}
$$

- Achieves optimal rate up to lower order terms among short step schedules:

$$
\sup_{(f, x_0)} \frac{f_N - f_\star}{\frac{1}{2}\|x_0 - x_\star\|^2} \approx \frac{1}{4N}
$$

Applications: Dynamic Short Stepsizes

- Dynamic short stepsize schedules:

$$
\begin{aligned}
\mathbf{h}^{(0)} &= [\,] && = [\,] \\
\mathbf{h}^{(1)} &= [1.4142] && = \mathbf{h}^{(0)} \bowtie [\,] \\
\mathbf{h}^{(2)} &= [1.4142, 1.6012] && = \mathbf{h}^{(1)} \bowtie [\,] \\
\mathbf{h}^{(3)} &= [1.4142, 1.6012, 1.7022] && = \mathbf{h}^{(2)} \bowtie [\,] \\
\mathbf{h}^{(4)} &= [1.4142, 1.6012, 1.7022, 1.7642] && = \mathbf{h}^{(3)} \bowtie [\,] \\
&\dots
\end{aligned}
$$

- Achieves optimal rate up to lower order terms among short step schedules:

$$
\sup_{(f, x_0)} \frac{f_N - f_\star}{\frac{1}{2}\left\|x_0 - x_\star\right\|^2} \approx \frac{1}{4N}
$$

- Compare with **[Teboulle, Vaisbourd '22] [Rotaru, Glineur, Patrinos '24]**

Applications: Optimal Basic Schedules

Optimal basic f-composable schedule of length N can be computed by DP

Optimal basic f-composable schedule of length N can be computed by DP

$$\mathbf{h}_{\mathsf{f}}^{(n)} = \min_{m=0,\ldots,n-1} \mathbf{h}_{\mathsf{s}}^{(m)} \triangleright \mathbf{h}_{\mathsf{f}}^{(n-m-1)}$$

Optimal basic f-composable schedule of length N can be computed by DP

$$\mathbf{h}_{\mathsf{f}}^{(n)} = \min_{m=0,\ldots,n-1} \mathbf{h}_{\mathsf{s}}^{(m)} \triangleright \mathbf{h}_{\mathsf{f}}^{(n-m-1)}$$

Theorem

The optimal basic f-composable schedule of length N has guarantee:

$$\sup_{(f,x_0)} \frac{f_N - f_\star}{\frac{1}{2}\left\|x_0 - x_\star\right\|^2} \leq \frac{0.42311 + o(1)}{(N+1)^{\log_2(1+\sqrt{2})}}$$

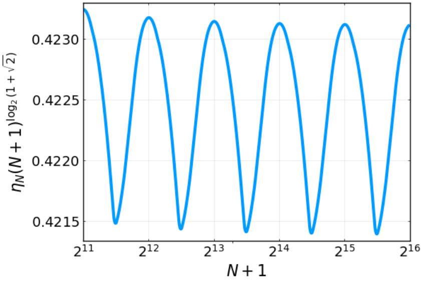
Optimal basic f-composable schedule of length N can be computed by DP

$$\mathbf{h}_{\mathsf{f}}^{(n)} = \min_{m=0,\dots,n-1} \mathbf{h}_{\mathsf{s}}^{(m)} \triangleright \mathbf{h}_{\mathsf{f}}^{(n-m-1)}$$

Theorem

The optimal basic f-composable schedule of length N has guarantee:

$$\sup_{(f,x_0)} \frac{f_N - f_\star}{\frac{1}{2}\|x_0 - x_\star\|^2} \leq \frac{0.42311 + o(1)}{(N+1)^{\log_2(1+\sqrt{2})}}$$



Summary and Pointers

Summary and Pointers

- **Composable stepsize schedules**: s-composable, f-composable

Summary and Pointers

- **Composable stepsize schedules**: s-composable, f-composable
- **Join operations**: $\bowtie$ and $\triangleright$ for producing new s/f-composable schedules

Summary and Pointers

- **Composable stepsize schedules**: s-composable, f-composable
- **Join operations**: $\bowtie$ and $\triangleright$ for producing new s/f-composable schedules
  - Recovers: Silver stepsizes, numerically optimal schedules, dynamic short stepsize schedules

Summary and Pointers

- **Composable stepsize schedules**: s-composable, f-composable
- **Join operations**: $\bowtie$ and $\triangleright$ for producing new s/f-composable schedules
  - Recovers: Silver stepsizes, numerically optimal schedules, dynamic short stepsize schedules
  - Optimal basic schedules

Summary and Pointers

- **Composable stepsize schedules**: s-composable, f-composable
- **Join operations**: $\bowtie$ and $\triangleright$ for producing new s/f-composable schedules
  - Recovers: Silver stepsizes, numerically optimal schedules, dynamic short stepsize schedules
  - Optimal basic schedules
- g-composable schedules and the g-join operation $\triangleleft$

$$\sup_{(f,x_0)} \frac{\frac{1}{2}\left\|g_N\right\|^2}{f_0 - f_\star}$$

Summary and Pointers

- **Composable stepsize schedules**: s-composable, f-composable
- **Join operations**: $\bowtie$ and $\triangleright$ for producing new s/f-composable schedules
  - Recovers: Silver stepsizes, numerically optimal schedules, dynamic short stepsize schedules
  - Optimal basic schedules
- g-composable schedules and the g-join operation $\triangleleft$

$$\sup_{(f,x_0)} \frac{\frac{1}{2}\|g_N\|^2}{f_0 - f_\star}$$

- H-duality: $\mathbf{h}$ is basic f-composable if and only if $\mathrm{rev}(\mathbf{h})$ is basic g-composable

Summary and Pointers

- **Composable stepsize schedules**: s-composable, f-composable
- **Join operations**: $\bowtie$ and $\triangleright$ for producing new s/f-composable schedules
  - Recovers: Silver stepsizes, numerically optimal schedules, dynamic short stepsize schedules
  - Optimal basic schedules
- g-composable schedules and the g-join operation $\triangleleft$

$$\sup_{(f,x_0)} \frac{\frac{1}{2}\|g_N\|^2}{f_0 - f_\star}$$

- H-duality: $\mathbf{h}$ is basic f-composable if and only if rev$(\mathbf{h})$ is basic g-composable

Thank you for listening!
Read more: `arXiv:2410.16249`
Questions?

References I

Altschuler, J. (2018). Greed, hedging, and acceleration in convex optimization. Master's thesis, Massachusetts Institute of Technology.

Altschuler, J. M. and Parrilo, P. A. (2023a). Acceleration by stepsize hedging i: Multi-step descent and the silver stepsize schedule.

Altschuler, J. M. and Parrilo, P. A. (2023b). Acceleration by stepsize hedging ii: Silver stepsize schedule for smooth convex optimization.

Bok, J. and Altschuler, J. M. (2024). Accelerating proximal gradient descent via silver stepsizes. *arXiv preprint arXiv:2412.05497*.

Bok, J. and Altschuler, J. M. (2025). Optimized methods for composite optimization: a reduction perspective. *arXiv preprint arXiv:2506.23756*.

Daccache, A. (2019). Performance estimation of the gradient method with fixed arbitrary step sizes. Master's thesis, Université Catholique de Louvain.

Drori, Y. and Teboulle, M. (2012). Performance of first-order methods for smooth convex minimization: a novel approach. *Mathematical Programming*, 145:451–482.

Eloi, D. (2022). Worst-case functions for the gradient method with fixed variable step sizes. Master's thesis, Université Catholique de Louvain.

References II

Grimmer, B., Shu, K., and Wang, A. L. (2024a). Accelerated objective gap and gradient norm convergence for gradient descent via long steps.

Grimmer, B., Shu, K., and Wang, A. L. (2024b). Composing optimized stepsize schedules for gradient descent. *arXiv preprint arXiv:2410.16249*.

Gupta, S. D., Parys, B. P. V., and Ryu, E. (2023). Branch-and-bound performance estimation programming: A unified methodology for constructing optimal optimization methods. *Mathematical Programming*.

Kim, J. (2024). A proof of the exact convergence rate of gradient descent. *arXiv preprint arXiv:2412.04427*.

Rotaru, T., Glineur, F., and Patrinos, P. (2024). Exact worst-case convergence rates of gradient descent: a complete analysis for all constant stepsizes over nonconvex and convex functions.

Taylor, A., Hendrickx, J., and Glineur, F. (2017). Smooth strongly convex interpolation and exact worst-case performance of first-order methods. *Mathematical Programming*, 161:307–345.

Teboulle, M. and Vaisbourd, Y. (2022). An elementary approach to tight worst case complexity analysis of gradient based methods. *Math. Program.*, 201(1–2):63–96.

References III

Wang, B., Ma, S., Yang, J., and Zhou, D. (2024). Relaxed proximal point algorithm: Tight complexity bounds and acceleration without momentum.

Zhang, Z. and Jiang, R. (2024). Accelerated gradient descent by concatenation of stepsize schedules.

Zhang, Z., Lee, J. D., Du, S. S., and Chen, Y. (2024). Anytime acceleration of gradient descent. *arXiv preprint arXiv:2411.17668*.

Appendix:
Overview of Composition Proofs

Key lemma

$\mathbf{h}$ is f-composable iff $f_N - f_\star \leq \frac{\eta}{2}\|x_0 - x_\star\|^2$ and tight for Huber and quadratic

Key lemma

$\boxed{\mathbf{h} \text{ is f-composable iff } f_N - f_\star \leq \frac{\eta}{2}\left\|x_0 - x_\star\right\|^2 \text{ and tight for Huber and quadratic}}$

Lemma

If $\mathbf{h}$ is f-composable with rate η, then for any z and any $t \in [0,1]$, it holds that

$$f_N - f_\star \leq \frac{\eta}{2}\left\|x_0 - (1-t)x_\star - tz^+\right\|^2 + t\left(f(z)^+ - f_\star\right)$$

where $z^+ = z - \nabla f(z)$ and $f(z)^+ = f(z) - \frac{1}{2}\left\|\nabla f(z)\right\|^2$

$\mathbf{h}$ is f-composable iff $f_N - f_\star \leq \frac{\eta}{2}\left\|x_0 - x_\star\right\|^2$ and tight for Huber and quadratic

Lemma

If $\mathbf{h}$ is f-composable with rate η, then for any z and any $t \in [0,1]$, it holds that

$$f_N - f_\star \leq \frac{\eta}{2}\left\|x_0 - (1-t)x_\star - tz^+\right\|^2 + t\left(f(z)^+ - f_\star\right)$$

where $z^+ = z - \nabla f(z)$ and $f(z)^+ = f(z) - \frac{1}{2}\left\|\nabla f(z)\right\|^2$

Proof strategy.

Key lemma

$\boxed{\mathbf{h} \text{ is f-composable iff } f_N - f_\star \leq \frac{\eta}{2} \|x_0 - x_\star\|^2 \text{ and tight for Huber and quadratic}}$

Lemma

If $\mathbf{h}$ is f-composable with rate η, then for any z and any $t \in [0,1]$, it holds that

$$f_N - f_\star \leq \frac{\eta}{2} \left\|x_0 - (1-t)x_\star - tz^+\right\|^2 + t\left(f(z)^+ - f_\star\right)$$

where $z^+ = z - \nabla f(z)$ and $f(z)^+ = f(z) - \frac{1}{2}\|\nabla f(z)\|^2$

Proof strategy.

- PEP certificate: $f_\star - f_N + \frac{\eta}{2}\|x_0 - x_\star\|^2 = \sum_{i,j} \lambda_{i,j} Q_{i,j} + (\succeq 0)$

$\mathbf{h}$ is f-composable iff $f_N - f_\star \le \frac{\eta}{2} \|x_0 - x_\star\|^2$ and tight for Huber and quadratic

Lemma

If $\mathbf{h}$ is f-composable with rate η, then for any z and any $t \in [0,1]$, it holds that

$$f_N - f_\star \le \frac{\eta}{2} \left\|x_0 - (1-t)x_\star - tz^+\right\|^2 + t\left(f(z)^+ - f_\star\right)$$

where $z^+ = z - \nabla f(z)$ and $f(z)^+ = f(z) - \frac{1}{2}\|\nabla f(z)\|^2$

Proof strategy.

- PEP certificate: $f_\star - f_N + \frac{\eta}{2}\|x_0 - x_\star\|^2 = \sum_{i,j} \lambda_{i,j} Q_{i,j} + (\succeq 0)$
- Certificate has some structure because Huber is tight

Key lemma

$\boxed{\mathbf{h} \text{ is f-composable iff } f_N - f_\star \leq \frac{\eta}{2} \|x_0 - x_\star\|^2 \text{ and tight for Huber and quadratic}}$

Lemma

If $\mathbf{h}$ is f-composable with rate η, then for any z and any $t \in [0, 1]$, it holds that

$$f_N - f_\star \leq \frac{\eta}{2} \left\|x_0 - (1-t)x_\star - tz^+\right\|^2 + t\left(f(z)^+ - f_\star\right)$$

where $z^+ = z - \nabla f(z)$ and $f(z)^+ = f(z) - \frac{1}{2}\|\nabla f(z)\|^2$

Proof strategy.

- PEP certificate: $f_\star - f_N + \frac{\eta}{2}\|x_0 - x_\star\|^2 = \sum_{i,j} \lambda_{i,j} Q_{i,j} + (\succeq 0)$
- Certificate has some structure because Huber is tight
- Replace $Q_{\star,j}$ with $(1-t)Q_{\star,j} + tQ_{z,j} \geq 0$ $\square$

F-join of two sequences

- $\mathbf{a}$ is s-convergent with rate α

F-join of two sequences

- $\mathbf{a}$ is s-convergent with rate α
- $\mathbf{b}$ is f-convergent with rate β

F-join of two sequences

- $\mathbf{a}$ is s-convergent with rate α
- $\mathbf{b}$ is f-convergent with rate β
- Construct $\mathbf{h} = [\mathbf{a}, \mu, \mathbf{b}]$

F-join of two sequences

- $\mathbf{a}$ is s-convergent with rate α
- $\mathbf{b}$ is f-convergent with rate β
- Construct $\mathbf{h} = [\mathbf{a}, \mu, \mathbf{b}]$

x_0 is given

F-join of two sequences

- $\mathbf{a}$ is s-convergent with rate α
- $\mathbf{b}$ is f-convergent with rate β
- Construct $\mathbf{h} = [\mathbf{a}, \mu, \mathbf{b}]$

x_0 is given

$$x_{i+1} = x_i - a_i g_i$$

F-join of two sequences

- $\mathbf{a}$ is s-convergent with rate α
- $\mathbf{b}$ is f-convergent with rate β
- Construct $\mathbf{h} = [\mathbf{a}, \mu, \mathbf{b}]$

$$\begin{aligned} &\boxed{x_0} \text{ is given} \\ &x_{i+1} = x_i - a_i g_i \\ &\boxed{x_M} = x_{M-1} - a_{M-1} g_{M-1} \end{aligned}$$

F-join of two sequences

- $\mathbf{a}$ is s-convergent with rate α
- $\mathbf{b}$ is f-convergent with rate β
- Construct $\mathbf{h} = [\mathbf{a}, \mu, \mathbf{b}]$

x_0 is given

$$x_{i+1} = x_i - a_i g_i$$

$$x_M = x_{M-1} - a_{M-1} g_{M-1}$$

$$y_0 = x_M - \mu g_M$$

F-join of two sequences

- $\mathbf{a}$ is s-convergent with rate α
- $\mathbf{b}$ is f-convergent with rate β
- Construct $\mathbf{h} = [\mathbf{a}, \mu, \mathbf{b}]$

$$\boxed{x_0} \text{ is given} \qquad \boxed{y_0 = x_M - \mu g_M}$$

$$x_{i+1} = x_i - a_i g_i \qquad y_{i+1} = y_i - b_i \ell_i$$

$$\boxed{x_M} = x_{M-1} - a_{M-1} g_{M-1}$$

F-join of two sequences

- $\mathbf{a}$ is s-convergent with rate α
- $\mathbf{b}$ is f-convergent with rate β
- Construct $\mathbf{h} = [\mathbf{a}, \mu, \mathbf{b}]$

$$
\begin{aligned}
&\boxed{x_0} \text{ is given} && \boxed{y_0 = x_M - \mu g_M} \\
&x_{i+1} = x_i - a_i g_i && y_{i+1} = y_i - b_i \ell_i \\
&\boxed{x_M} = x_{M-1} - a_{M-1} g_{M-1} && \boxed{y_N} = y_{N-1} - b_{N-1} \ell_{N-1}
\end{aligned}
$$

F-join of two sequences

- $\mathbf{a}$ is s-convergent with rate α
- $\mathbf{b}$ is f-convergent with rate β
- Construct $\mathbf{h} = [\mathbf{a}, \mu, \mathbf{b}]$

$$
\begin{aligned}
&x_0 \text{ is given} && y_0 = x_M - \mu g_M \\
&x_{i+1} = x_i - a_i g_i && y_{i+1} = y_i - b_i \ell_i \\
&x_M = x_{M-1} - a_{M-1} g_{M-1} && y_N = y_{N-1} - b_{N-1} \ell_{N-1}
\end{aligned}
$$

- **Strategy**:

F-join of two sequences

- $\mathbf{a}$ is s-convergent with rate α
- $\mathbf{b}$ is f-convergent with rate β
- Construct $\mathbf{h} = [\mathbf{a}, \mu, \mathbf{b}]$

$$\begin{aligned} &\boxed{x_0} \text{ is given} && \boxed{y_0 = x_M - \mu g_M} \\ &x_{i+1} = x_i - a_i g_i && y_{i+1} = y_i - b_i \ell_i \\ &\boxed{x_M} = x_{M-1} - a_{M-1} g_{M-1} && \boxed{y_N} = y_{N-1} - b_{N-1} \ell_{N-1} \end{aligned}$$

- **Strategy**:
  - Use s-composable definition for $\mathbf{a}$

F-join of two sequences

- $\mathbf{a}$ is s-convergent with rate α
- $\mathbf{b}$ is f-convergent with rate β
- Construct $\mathbf{h} = [\mathbf{a}, \mu, \mathbf{b}]$

$$
\begin{aligned}
&\boxed{x_0} \text{ is given} && \boxed{y_0 = x_M - \mu g_M} \\
&x_{i+1} = x_i - a_i g_i && y_{i+1} = y_i - b_i \ell_i \\
&\boxed{x_M} = x_{M-1} - a_{M-1} g_{M-1} && \boxed{y_N} = y_{N-1} - b_{N-1} \ell_{N-1}
\end{aligned}
$$

- **Strategy**:
  - Use s-composable definition for $\mathbf{a}$
  - Use upgraded f-composable inequality for $\mathbf{b}$ and compare with $z = x_M$

Proof sketch

- s-composability of $\mathbf{a}$:

$$f_M - f_\star \leq \frac{\alpha}{2-\alpha}\left(\frac{1}{2}\left\|x_0 - x_\star\right\|^2 - \frac{1}{2}\left\|x_M - \frac{g_M}{\alpha} - x_\star\right\|^2\right)$$

Proof sketch

- s-composability of $\mathbf{a}$:

$$f_M - f_\star \leq \frac{\alpha}{2-\alpha}\left(\frac{1}{2}\|x_0 - x_\star\|^2 - \frac{1}{2}\left\|x_M - \frac{g_M}{\alpha} - x_\star\right\|^2\right)$$

- Key lemma applied to $\mathbf{b}$ with $z = x_M$

$$f_N - f_\star \leq \frac{\beta}{2}\|(1-t)(x_M - x_\star) - (\mu - t)g_M\|^2 - \frac{t}{2}\|g_M\|^2 + t\left(f_M - f_\star\right)$$

Proof sketch

- s-composability of $\mathbf{a}$:

$$f_M - f_\star \leq \frac{\alpha}{2-\alpha}\left(\frac{1}{2}\|x_0 - x_\star\|^2 - \frac{1}{2}\left\|x_M - \frac{g_M}{\alpha} - x_\star\right\|^2\right)$$

- Key lemma applied to $\mathbf{b}$ with $z = x_M$

$$f_N - f_\star \leq \frac{\beta}{2}\|(1-t)(x_M - x_\star) - (\mu - t)g_M\|^2 - \frac{t}{2}\|g_M\|^2 + t\left(f_M - f_\star\right)$$

Pick $t = 1 + \alpha(1-\mu)$ so that the norm terms cancel out:

$$f_N - f_\star \leq \frac{\alpha^2\beta(\mu-1)^2}{2}\|x_0 - x_\star\|^2 + (\text{expression in } f_M - f_\star, x_M - x_\star, g_M)$$

Proof sketch

- s-composability of $\mathbf{a}$:

$$f_M - f_\star \leq \frac{\alpha}{2-\alpha}\left(\frac{1}{2}\|x_0 - x_\star\|^2 - \frac{1}{2}\left\|x_M - \frac{g_M}{\alpha} - x_\star\right\|^2\right)$$

- Key lemma applied to $\mathbf{b}$ with $z = x_M$

$$f_N - f_\star \leq \frac{\beta}{2}\|(1-t)(x_M - x_\star) - (\mu - t)g_M\|^2 - \frac{t}{2}\|g_M\|^2 + t\,(f_M - f_\star)$$

  Pick $t = 1 + \alpha(1-\mu)$ so that the norm terms cancel out:

$$f_N - f_\star \leq \frac{\alpha^2\beta(\mu-1)^2}{2}\|x_0 - x_\star\|^2 + (\text{expression in } f_M - f_\star, x_M - x_\star, g_M)$$

- Ask Mathematica to optimize rate:

Proof sketch

- s-composability of $\mathbf{a}$:

$$f_M - f_\star \le \frac{\alpha}{2-\alpha}\left(\frac{1}{2}\|x_0 - x_\star\|^2 - \frac{1}{2}\left\|x_M - \frac{g_M}{\alpha} - x_\star\right\|^2\right)$$

- Key lemma applied to $\mathbf{b}$ with $z = x_M$

$$f_N - f_\star \le \frac{\beta}{2}\|(1-t)(x_M - x_\star) - (\mu - t)g_M\|^2 - \frac{t}{2}\|g_M\|^2 + t\,(f_M - f_\star)$$

Pick $t = 1 + \alpha(1-\mu)$ so that the norm terms cancel out:

$$f_N - f_\star \le \frac{\alpha^2\beta(\mu-1)^2}{2}\|x_0 - x_\star\|^2 + (\text{expression in } f_M - f_\star, x_M - x_\star, g_M)$$

- Ask Mathematica to optimize rate:

$$\mu = 1 + \frac{\sqrt{\alpha^2 + 8\alpha\beta} - \alpha}{4\alpha\beta}$$

$$\text{new rate} = \frac{2\alpha\beta}{\alpha + 4\beta + \sqrt{\alpha^2 + 8\alpha\beta}} = \alpha \triangleright \beta. \qquad \square$$