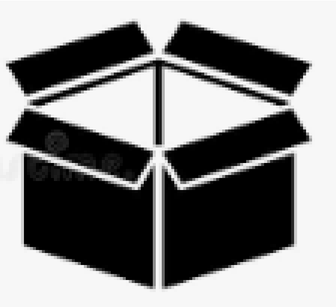
Newton-Type Methods

Exploring the Interplay Between Inner and Outer Iterations

Part II

Fred Roosta

School of Mathematics and Physics
University of Queensland



	$\mathbf{Hp} \approx -\mathbf{g}$	
	CG	**MINRES**
Sub-problems		
Problem class		
Metric / Rate		

	$\mathbf{Hp} \approx -\mathbf{g}$	
	CG	**MINRES**
Sub-problems	$\mathbf{p}_t = \underset{\mathbf{p} \in \mathcal{K}_t}{\arg\min} \ \frac{1}{2}\langle \mathbf{p}, \mathbf{Hp}\rangle - \langle \mathbf{p}, \mathbf{g}\rangle$	$\mathbf{p}_t = \underset{\mathbf{p} \in \mathcal{K}_t}{\arg\min} \ \Vert \mathbf{Hp} + \mathbf{g} \Vert^2$
Problem class		
Metric / Rate		

Krylov subspace:

$$\mathcal{K}_t(\mathbf{H}, \mathbf{g}) = \text{Span}\{\mathbf{g}, \mathbf{Hg}, \ldots, \mathbf{H}^{t-1}\mathbf{g}\}$$

	$\mathbf{Hp} \approx -\mathbf{g}$	
	CG	**MINRES**
Sub-problems	$\mathbf{p}_t = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \frac{1}{2}\langle \mathbf{p}, \mathbf{Hp} \rangle - \langle \mathbf{p}, \mathbf{g} \rangle$	$\mathbf{p}_t = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \Vert \mathbf{Hp} + \mathbf{g} \Vert^2$
Problem class	**Positive Definite**	**Symmetric**
Metric / Rate		

Krylov subspace:

$$\mathcal{K}_t(\mathbf{H}, \mathbf{g}) = \text{Span}\{\mathbf{g}, \mathbf{Hg}, \ldots, \mathbf{H}^{t-1}\mathbf{g}\}$$

	$\mathbf{Hp} \approx -\mathbf{g}$	
	CG	**MINRES**
Sub-problems	$\mathbf{p}_t = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \frac{1}{2} \langle \mathbf{p}, \mathbf{Hp} \rangle - \langle \mathbf{p}, \mathbf{g} \rangle$	$\mathbf{p}_t = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \\|\mathbf{Hp} + \mathbf{g}\\|^2$
Problem class	**Positive Definite**	**Symmetric**
Metric / Rate	$\\|\mathbf{Hp}_t + \mathbf{g}\\|$: **R-linear**	$\\|\mathbf{Hp}_t + \mathbf{g}\\|$: **Q-linear**

Krylov subspace:

$$\mathcal{K}_t(\mathbf{H}, \mathbf{g}) = \text{Span}\{\mathbf{g}, \mathbf{Hg}, \ldots, \mathbf{H}^{t-1}\mathbf{g}\}$$

	$\mathbf{Hp} \approx -\mathbf{g}$	
	CG	**MINRES**
Sub-problems	$\mathbf{p}_t = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \frac{1}{2}\langle \mathbf{p}, \mathbf{Hp}\rangle - \langle \mathbf{p}, \mathbf{g}\rangle$	$\mathbf{p}_t = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \|\mathbf{Hp} + \mathbf{g}\|^2$
Problem class	**Positive Definite**	**Symmetric**
Metric / Rate	$\|\mathbf{Hp}_t + \mathbf{g}\|$: **R-linear** $\|\mathbf{p}_t - \mathbf{p}^\star\|_{\mathbf{H}}$: **Q-Linear**	$\|\mathbf{Hp}_t + \mathbf{g}\|$: **Q-linear** $\|\mathbf{p}_t - \mathbf{p}^\star\|_{\mathbf{H}}$: **R-Linear (PD)**

Krylov subspace:

$$\mathcal{K}_t(\mathbf{H}, \mathbf{g}) = \text{Span}\{\mathbf{g}, \mathbf{Hg}, \ldots, \mathbf{H}^{t-1}\mathbf{g}\}$$

For simplicity, assume $\mathbf{p}_0 = \mathbf{0}$. Most, but not all, of what follows can be generalized to arbitrary initialization.

Both CG and MINRES are instances of **Projection Framework**:

Both CG and MINRES are instances of **Projection Framework**:

Projection Methods

- **Search Subspace**: $\mathcal{K}_t$ with $\dim(\mathcal{K}_t) = t$

Both CG and MINRES are instances of **Projection Framework**:

Projection Methods

- **Search Subspace**: $\mathcal{K}_t$ with $\dim(\mathcal{K}_t) = t$
- **Constraint Subspace**: $\mathcal{W}_t$ with $\dim(\mathcal{W}_t) = t$

Both CG and MINRES are instances of **Projection Framework**:

Projection Methods

- **Search Subspace**: $\mathcal{K}_t$ with $\dim(\mathcal{K}_t) = t$
- **Constraint Subspace**: $\mathcal{W}_t$ with $\dim(\mathcal{W}_t) = t$
- **Petrov-Galerkin Conditions**: Find $\mathbf{p}_t \in \mathcal{K}_t$ s.t. $\mathbf{r}_t \perp \mathcal{W}_t$

Both CG and MINRES are instances of **Projection Framework**:

Projection Methods

- **Search Subspace**: $\mathcal{K}_t$ with $\dim(\mathcal{K}_t) = t$
- **Constraint Subspace**: $\mathcal{W}_t$ with $\dim(\mathcal{W}_t) = t$
- **Petrov-Galerkin Conditions**: Find $\mathbf{p}_t \in \mathcal{K}_t$ s.t. $\mathbf{r}_t \perp \mathcal{W}_t$

- **Orthogonal Projection**: $\mathcal{W}_t = \mathcal{K}_t$

Both CG and MINRES are instances of **Projection Framework**:

Projection Methods

- **Search Subspace**: $\mathcal{K}_t$ with $\dim(\mathcal{K}_t) = t$
- **Constraint Subspace**: $\mathcal{W}_t$ with $\dim(\mathcal{W}_t) = t$
- **Petrov-Galerkin Conditions**: Find $\mathbf{p}_t \in \mathcal{K}_t$ s.t. $\mathbf{r}_t \perp \mathcal{W}_t$

- **Orthogonal Projection**: $\mathcal{W}_t = \mathcal{K}_t$
- **Oblique Projection**: $\mathcal{W}_t \neq \mathcal{K}_t$, e.g., $\mathcal{W}_t = \mathbf{H}\mathcal{K}_t$

Theorem

1. *The case where* $\mathbf{H} \succ \mathbf{0}$ *and* $\mathcal{W}_t = \mathcal{K}_t$ *is equivalent to*

Theorem

1. *The case where* $\mathbf{H} \succ \mathbf{0}$ *and* $\mathcal{W}_t = \mathcal{K}_t$ *is equivalent to*

$$\mathbf{p}_t = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \|\mathbf{p} - \mathbf{p}^\star\|_{\mathbf{H}}^2$$

Theorem

1. *The case where* $\mathbf{H} \succ \mathbf{0}$ *and* $\mathcal{W}_t = \mathcal{K}_t$ *is equivalent to*

$$\mathbf{p}_t = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \|\mathbf{p} - \mathbf{p}^\star\|_{\mathbf{H}}^2 = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \frac{1}{2} \langle \mathbf{p}, \mathbf{H}\mathbf{p} \rangle + \langle \mathbf{g}, \mathbf{p} \rangle .$$

Theorem

1. *The case where* $\mathbf{H} \succ \mathbf{0}$ *and* $\mathcal{W}_t = \mathcal{K}_t$ *is equivalent to*

$$\mathbf{p}_t = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \|\mathbf{p} - \mathbf{p}^\star\|_{\mathbf{H}}^2 = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \frac{1}{2} \langle \mathbf{p}, \mathbf{H}\mathbf{p} \rangle + \langle \mathbf{g}, \mathbf{p} \rangle .$$

2. *The case where* $\det(\mathbf{H}) \neq 0$ *and* $\mathcal{W}_t = \mathbf{H}\mathcal{K}_t$ *is equivalent to*

Theorem

1. *The case where* $\mathbf{H} \succ \mathbf{0}$ *and* $\mathcal{W}_t = \mathcal{K}_t$ *is equivalent to*

$$\mathbf{p}_t = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \|\mathbf{p} - \mathbf{p}^\star\|_{\mathbf{H}}^2 = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \frac{1}{2} \langle \mathbf{p}, \mathbf{H}\mathbf{p} \rangle + \langle \mathbf{g}, \mathbf{p} \rangle .$$

2. *The case where* $\det(\mathbf{H}) \neq 0$ *and* $\mathcal{W}_t = \mathbf{H}\mathcal{K}_t$ *is equivalent to*

$$\mathbf{p}_t = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \|\mathbf{H}\mathbf{p} + \mathbf{g}\|^2 .$$

Proof.

1. Let $\mathbf{V}_t \in \mathbb{R}^{d \times t}$ be such that $\text{Range}(\mathbf{V}_t) = \mathcal{K}_t$.

□

Proof.

1. Let $\mathbf{V}_t \in \mathbb{R}^{d \times t}$ be such that $\text{Range}(\mathbf{V}_t) = \mathcal{K}_t$. To solve

$$\mathbf{p}_t = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \frac{1}{2} \langle \mathbf{p}, \mathbf{H}\mathbf{p} \rangle + \langle \mathbf{g}, \mathbf{p} \rangle ,$$

□

Proof.

1. Let $\mathbf{V}_t \in \mathbb{R}^{d \times t}$ be such that $\text{Range}(\mathbf{V}_t) = \mathcal{K}_t$. To solve

$$\mathbf{p}_t = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \frac{1}{2} \langle \mathbf{p}, \mathbf{H}\mathbf{p} \rangle + \langle \mathbf{g}, \mathbf{p} \rangle,$$

we can equivalently solve

$$\mathbf{y}_t = \arg\min_{\mathbf{y} \in \mathbb{R}^t} \frac{1}{2} \langle \mathbf{V}_t\mathbf{y}, \mathbf{H}(\mathbf{V}_t\mathbf{y}) \rangle + \langle \mathbf{g}, \mathbf{V}_t\mathbf{y} \rangle,$$

and then set $\mathbf{p}_t = \mathbf{V}_t\mathbf{y}_t$.

□

Proof.

1. Let $\mathbf{V}_t \in \mathbb{R}^{d\times t}$ be such that $\text{Range}(\mathbf{V}_t) = \mathcal{K}_t$. To solve

$$\mathbf{p}_t = \arg\min_{\mathbf{p}\in\mathcal{K}_t} \frac{1}{2}\langle \mathbf{p}, \mathbf{H}\mathbf{p}\rangle + \langle \mathbf{g}, \mathbf{p}\rangle \,,$$

we can equivalently solve

$$\mathbf{y}_t = \arg\min_{\mathbf{y}\in\mathbb{R}^t} \frac{1}{2}\langle \mathbf{V}_t\mathbf{y}, \mathbf{H}\left(\mathbf{V}_t\mathbf{y}\right)\rangle + \langle \mathbf{g}, \mathbf{V}_t\mathbf{y}\rangle \,,$$

and then set $\mathbf{p}_t = \mathbf{V}_t\mathbf{y}_t$. Since $\mathbf{H} \succ \mathbf{0}$

□

Proof.

1. Let $\mathbf{V}_t \in \mathbb{R}^{d \times t}$ be such that $\text{Range}(\mathbf{V}_t) = \mathcal{K}_t$. To solve

$$\mathbf{p}_t = \underset{\mathbf{p} \in \mathcal{K}_t}{\arg\min} \frac{1}{2} \langle \mathbf{p}, \mathbf{H}\mathbf{p} \rangle + \langle \mathbf{g}, \mathbf{p} \rangle ,$$

we can equivalently solve

$$\mathbf{y}_t = \underset{\mathbf{y} \in \mathbb{R}^t}{\arg\min} \frac{1}{2} \langle \mathbf{V}_t \mathbf{y}, \mathbf{H} (\mathbf{V}_t \mathbf{y}) \rangle + \langle \mathbf{g}, \mathbf{V}_t \mathbf{y} \rangle ,$$

and then set $\mathbf{p}_t = \mathbf{V}_t \mathbf{y}_t$. Since $\mathbf{H} \succ \mathbf{0}$, it is necessary and sufficient for the optimal $\mathbf{y}_t$ to satisfy

$$\mathbf{V}_t^\intercal \mathbf{H} \mathbf{V}_t \mathbf{y}_t + \mathbf{V}_t^\intercal \mathbf{g} = \mathbf{0}$$

□

Proof.

1. Let $\mathbf{V}_t \in \mathbb{R}^{d \times t}$ be such that $\text{Range}(\mathbf{V}_t) = \mathcal{K}_t$. To solve

$$\mathbf{p}_t = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \frac{1}{2} \langle \mathbf{p}, \mathbf{H}\mathbf{p} \rangle + \langle \mathbf{g}, \mathbf{p} \rangle \,,$$

we can equivalently solve

$$\mathbf{y}_t = \arg\min_{\mathbf{y} \in \mathbb{R}^t} \frac{1}{2} \langle \mathbf{V}_t\mathbf{y}, \mathbf{H}\left(\mathbf{V}_t\mathbf{y}\right) \rangle + \langle \mathbf{g}, \mathbf{V}_t\mathbf{y} \rangle \,,$$

and then set $\mathbf{p}_t = \mathbf{V}_t\mathbf{y}_t$. Since $\mathbf{H} \succ \mathbf{0}$, it is necessary and sufficient for the optimal $\mathbf{y}_t$ to satisfy

$$\mathbf{V}_t^\intercal \mathbf{H} \mathbf{V}_t \mathbf{y}_t + \mathbf{V}_t^\intercal \mathbf{g} = \mathbf{0} \iff \mathbf{V}_t^\intercal \left(\mathbf{H}\mathbf{p}_t + \mathbf{g}\right) = \mathbf{0}$$

□

Proof.

1. Let $\mathbf{V}_t \in \mathbb{R}^{d\times t}$ be such that Range($\mathbf{V}_t$) $= \mathcal{K}_t$. To solve

$$\mathbf{p}_t = \underset{\mathbf{p}\in\mathcal{K}_t}{\arg\min}\ \frac{1}{2}\langle \mathbf{p}, \mathbf{H}\mathbf{p}\rangle + \langle \mathbf{g}, \mathbf{p}\rangle\,,$$

we can equivalently solve

$$\mathbf{y}_t = \underset{\mathbf{y}\in\mathbb{R}^t}{\arg\min}\ \frac{1}{2}\langle \mathbf{V}_t\mathbf{y}, \mathbf{H}\left(\mathbf{V}_t\mathbf{y}\right)\rangle + \langle \mathbf{g}, \mathbf{V}_t\mathbf{y}\rangle\,,$$

and then set $\mathbf{p}_t = \mathbf{V}_t\mathbf{y}_t$. Since $\mathbf{H} \succ \mathbf{0}$, it is necessary and sufficient for the optimal $\mathbf{y}_t$ to satisfy

$$\mathbf{V}_t^\intercal\mathbf{H}\mathbf{V}_t\mathbf{y}_t + \mathbf{V}_t^\intercal\mathbf{g} = \mathbf{0} \iff \mathbf{V}_t^\intercal\left(\mathbf{H}\mathbf{p}_t + \mathbf{g}\right) = \mathbf{0}$$

□

Proof.

1. Let $\mathbf{V}_t \in \mathbb{R}^{d\times t}$ be such that $\text{Range}(\mathbf{V}_t) = \mathcal{K}_t$. To solve

$$\mathbf{p}_t = \arg\min_{\mathbf{p}\in\mathcal{K}_t} \frac{1}{2}\langle \mathbf{p}, \mathbf{H}\mathbf{p}\rangle + \langle \mathbf{g}, \mathbf{p}\rangle,$$

we can equivalently solve

$$\mathbf{y}_t = \arg\min_{\mathbf{y}\in\mathbb{R}^t} \frac{1}{2}\langle \mathbf{V}_t\mathbf{y}, \mathbf{H}(\mathbf{V}_t\mathbf{y})\rangle + \langle \mathbf{g}, \mathbf{V}_t\mathbf{y}\rangle,$$

and then set $\mathbf{p}_t = \mathbf{V}_t\mathbf{y}_t$. Since $\mathbf{H} \succ \mathbf{0}$, it is necessary and sufficient for the optimal $\mathbf{y}_t$ to satisfy

$$\mathbf{V}_t^{\intercal}\mathbf{H}\mathbf{V}_t\mathbf{y}_t + \mathbf{V}_t^{\intercal}\mathbf{g} = \mathbf{0} \iff \mathbf{V}_t^{\intercal}\underbrace{(\mathbf{H}\mathbf{p}_t + \mathbf{g})}_{\mathbf{r}_t} = \mathbf{0}$$

□

Proof.

1. Let $\mathbf{V}_t \in \mathbb{R}^{d\times t}$ be such that $\text{Range}(\mathbf{V}_t) = \mathcal{K}_t$. To solve

$$\mathbf{p}_t = \arg\min_{\mathbf{p}\in\mathcal{K}_t} \frac{1}{2}\langle \mathbf{p}, \mathbf{H}\mathbf{p}\rangle + \langle \mathbf{g}, \mathbf{p}\rangle,$$

we can equivalently solve

$$\mathbf{y}_t = \arg\min_{\mathbf{y}\in\mathbb{R}^t} \frac{1}{2}\langle \mathbf{V}_t\mathbf{y}, \mathbf{H}(\mathbf{V}_t\mathbf{y})\rangle + \langle \mathbf{g}, \mathbf{V}_t\mathbf{y}\rangle,$$

and then set $\mathbf{p}_t = \mathbf{V}_t\mathbf{y}_t$. Since $\mathbf{H} \succ \mathbf{0}$, it is necessary and sufficient for the optimal $\mathbf{y}_t$ to satisfy

$$\mathbf{V}_t^\intercal \mathbf{H}\mathbf{V}_t\mathbf{y}_t + \mathbf{V}_t^\intercal\mathbf{g} = \mathbf{0} \iff \mathbf{V}_t^\intercal \underbrace{(\mathbf{H}\mathbf{p}_t + \mathbf{g})}_{\mathbf{r}_t} = \mathbf{0} \iff \mathbf{r}_t \perp \mathcal{K}_t.$$

□

Proof.

1. Let $\mathbf{V}_t \in \mathbb{R}^{d \times t}$ be such that $\text{Range}(\mathbf{V}_t) = \mathcal{K}_t$. To solve

$$\mathbf{p}_t = \underset{\mathbf{p} \in \mathcal{K}_t}{\arg\min} \, \frac{1}{2} \langle \mathbf{p}, \mathbf{H}\mathbf{p} \rangle + \langle \mathbf{g}, \mathbf{p} \rangle \, ,$$

we can equivalently solve

$$\mathbf{y}_t = \underset{\mathbf{y} \in \mathbb{R}^t}{\arg\min} \, \frac{1}{2} \langle \mathbf{V}_t \mathbf{y}, \mathbf{H} \left(\mathbf{V}_t \mathbf{y} \right) \rangle + \langle \mathbf{g}, \mathbf{V}_t \mathbf{y} \rangle \, ,$$

and then set $\mathbf{p}_t = \mathbf{V}_t \mathbf{y}_t$. Since $\mathbf{H} \succ \mathbf{0}$, it is necessary and sufficient for the optimal $\mathbf{y}_t$ to satisfy

$$\mathbf{V}_t^{\intercal} \mathbf{H} \mathbf{V}_t \mathbf{y}_t + \mathbf{V}_t^{\intercal} \mathbf{g} = \mathbf{0} \iff \mathbf{V}_t^{\intercal} \underbrace{(\mathbf{H}\mathbf{p}_t + \mathbf{g})}_{\mathbf{r}_t} = \mathbf{0} \iff \mathbf{r}_t \perp \mathcal{K}_t.$$

2. Similarly as above.

□

Lanczos Process: When $\mathbf{H}$ is symmetric, the Lanczos process finds a basis $\mathbf{V}_t$ for $\mathcal{K}_t(\mathbf{H}, \mathbf{g})$, based on Gram-Schmidt procedure, such that

$$\mathbf{H}\mathbf{V}_t = \mathbf{V}_{t+1}\mathbf{T}_{t+1,t},$$

Lanczos Process: When $\mathbf{H}$ is symmetric, the Lanczos process finds a basis $\mathbf{V}_t$ for $\mathcal{K}_t(\mathbf{H}, \mathbf{g})$, based on Gram-Schmidt procedure, such that

$$\mathbf{H}\mathbf{V}_t = \mathbf{V}_{t+1}\mathbf{T}_{t+1,t},$$

where

$$\mathbf{V}_t = \left[\mathbf{v}_1 \mid \quad \mathbf{v}_2 \mid \ldots \quad \mid \mathbf{v}_t\right], \quad \text{with} \quad \mathbf{v}_1 = \frac{\mathbf{g}}{\|\mathbf{g}\|},$$

$$\mathbf{T}_{t+1,t} = \begin{bmatrix} \alpha_1 & \beta_2 & & & \\ \beta_2 & \alpha_2 & \beta_3 & & \\ & \beta_3 & \alpha_3 & \ddots & \\ & & \ddots & \ddots & \beta_t \\ & & & \beta_t & \alpha_t \\ \hdashline & & & & \beta_{t+1} \end{bmatrix} \triangleq \begin{bmatrix} \mathbf{T}_t \\ \beta_{t+1}\mathbf{e}_t^\intercal \end{bmatrix}.$$

Lanczos Process: When $\mathbf{H}$ is symmetric, the Lanczos process finds a basis $\mathbf{V}_t$ for $\mathcal{K}_t(\mathbf{H}, \mathbf{g})$, based on Gram-Schmidt procedure, such that

$$\mathbf{H}\mathbf{V}_t = \mathbf{V}_{t+1}\mathbf{T}_{t+1,t},$$

where

$$\mathbf{V}_t = \left[\mathbf{v}_1 \mid \quad \mathbf{v}_2 \mid \dots \quad \mid \mathbf{v}_t\right], \quad \text{with} \quad \mathbf{v}_1 = \frac{\mathbf{g}}{\|\mathbf{g}\|},$$

$$\mathbf{T}_{t+1,t} = \begin{bmatrix} \alpha_1 & \beta_2 & & & \\ \beta_2 & \alpha_2 & \beta_3 & & \\ & \beta_3 & \alpha_3 & \ddots & \\ & & \ddots & \ddots & \beta_t \\ & & & \beta_t & \alpha_t \\ \hline & & & & \beta_{t+1} \end{bmatrix} \triangleq \begin{bmatrix} \mathbf{T}_t \\ \beta_{t+1}\mathbf{e}_t^{\intercal} \end{bmatrix}.$$

So, $\mathbf{V}_t^{\intercal}\mathbf{H}\mathbf{V}_t = \mathbf{T}_t$ in tridiagonal.

Conjugate Gradient (CG): Assuming $\mathbf{H} \succ \mathbf{0}$,

$$\mathbf{p}_t = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \frac{1}{2} \langle \mathbf{p}, \mathbf{H}\mathbf{p} \rangle + \langle \mathbf{g}, \mathbf{p} \rangle$$

Conjugate Gradient (CG): Assuming $\mathbf{H} \succ \mathbf{0}$,

$$\mathbf{p}_t = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \frac{1}{2} \langle \mathbf{p}, \mathbf{H}\mathbf{p} \rangle + \langle \mathbf{g}, \mathbf{p} \rangle$$

If $\mathbf{V}_t$ is the basis from the Lanczos procedure, then we can set $\mathbf{p}_t = \mathbf{V}_t \mathbf{y}_t$ where $\mathbf{y}_t$ is the solution to:

Conjugate Gradient (CG): Assuming $\mathbf{H} \succ \mathbf{0}$,

$$\mathbf{p}_t = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \frac{1}{2} \langle \mathbf{p}, \mathbf{H}\mathbf{p} \rangle + \langle \mathbf{g}, \mathbf{p} \rangle$$

If $\mathbf{V}_t$ is the basis from the Lanczos procedure, then we can set $\mathbf{p}_t = \mathbf{V}_t \mathbf{y}_t$ where $\mathbf{y}_t$ is the solution to:

$$\mathbf{y}_t = \arg\min_{\mathbf{y} \in \mathbb{R}^t} \frac{1}{2} \langle \mathbf{V}_t \mathbf{y}, \mathbf{H}\mathbf{V}_t \mathbf{y} \rangle + \langle \mathbf{g}, \mathbf{V}_t \mathbf{y} \rangle$$

Conjugate Gradient (CG): Assuming $\mathbf{H} \succ \mathbf{0}$,

$$\mathbf{p}_t = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \frac{1}{2} \langle \mathbf{p}, \mathbf{H}\mathbf{p} \rangle + \langle \mathbf{g}, \mathbf{p} \rangle$$

If $\mathbf{V}_t$ is the basis from the Lanczos procedure, then we can set $\mathbf{p}_t = \mathbf{V}_t \mathbf{y}_t$ where $\mathbf{y}_t$ is the solution to:

$$\begin{aligned} \mathbf{y}_t &= \arg\min_{\mathbf{y} \in \mathbb{R}^t} \frac{1}{2} \langle \mathbf{V}_t \mathbf{y}, \mathbf{H}\mathbf{V}_t \mathbf{y} \rangle + \langle \mathbf{g}, \mathbf{V}_t \mathbf{y} \rangle \\ &= \arg\min_{\mathbf{y} \in \mathbb{R}^t} \frac{1}{2} \langle \mathbf{y}, \mathbf{T}_t \mathbf{y} \rangle + \langle \|\mathbf{g}\| \, \mathbf{e}_1, \mathbf{y} \rangle . \end{aligned}$$

Conjugate Gradient (CG): Assuming $\mathbf{H} \succ \mathbf{0}$,

$$\mathbf{p}_t = \underset{\mathbf{p} \in \mathcal{K}_t}{\arg\min} \frac{1}{2} \langle \mathbf{p}, \mathbf{H}\mathbf{p} \rangle + \langle \mathbf{g}, \mathbf{p} \rangle$$

If $\mathbf{V}_t$ is the basis from the Lanczos procedure, then we can set $\mathbf{p}_t = \mathbf{V}_t \mathbf{y}_t$ where $\mathbf{y}_t$ is the solution to:

$$\begin{aligned} \mathbf{y}_t &= \underset{\mathbf{y} \in \mathbb{R}^t}{\arg\min} \frac{1}{2} \langle \mathbf{V}_t \mathbf{y}, \mathbf{H}\mathbf{V}_t \mathbf{y} \rangle + \langle \mathbf{g}, \mathbf{V}_t \mathbf{y} \rangle \\ &= \underset{\mathbf{y} \in \mathbb{R}^t}{\arg\min} \frac{1}{2} \langle \mathbf{y}, \mathbf{T}_t \mathbf{y} \rangle + \langle \|\mathbf{g}\| \mathbf{e}_1, \mathbf{y} \rangle . \end{aligned}$$

Taking the derivative and setting it to zero, gives

Conjugate Gradient (CG): Assuming $\mathbf{H} \succ \mathbf{0}$,

$$\mathbf{p}_t = \arg\min_{\mathbf{p}\in\mathcal{K}_t} \frac{1}{2}\langle \mathbf{p}, \mathbf{H}\mathbf{p}\rangle + \langle \mathbf{g}, \mathbf{p}\rangle$$

If $\mathbf{V}_t$ is the basis from the Lanczos procedure, then we can set $\mathbf{p}_t = \mathbf{V}_t\mathbf{y}_t$ where $\mathbf{y}_t$ is the solution to:

$$\begin{aligned}\mathbf{y}_t &= \arg\min_{\mathbf{y}\in\mathbb{R}^t} \frac{1}{2}\langle \mathbf{V}_t\mathbf{y}, \mathbf{H}\mathbf{V}_t\mathbf{y}\rangle + \langle \mathbf{g}, \mathbf{V}_t\mathbf{y}\rangle \\ &= \arg\min_{\mathbf{y}\in\mathbb{R}^t} \frac{1}{2}\langle \mathbf{y}, \mathbf{T}_t\mathbf{y}\rangle + \langle \|\mathbf{g}\|\,\mathbf{e}_1, \mathbf{y}\rangle .\end{aligned}$$

Taking the derivative and setting it to zero, gives

$$\mathbf{T}_t\mathbf{y}_t = -\|\mathbf{g}\|\,\mathbf{e}_1.$$

Conjugate Gradient (CG): Assuming $\mathbf{H} \succ \mathbf{0}$,

$$\mathbf{p}_t = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \frac{1}{2} \langle \mathbf{p}, \mathbf{H}\mathbf{p} \rangle + \langle \mathbf{g}, \mathbf{p} \rangle$$

If $\mathbf{V}_t$ is the basis from the Lanczos procedure, then we can set $\mathbf{p}_t = \mathbf{V}_t \mathbf{y}_t$ where $\mathbf{y}_t$ is the solution to:

$$\begin{aligned}\mathbf{y}_t &= \arg\min_{\mathbf{y} \in \mathbb{R}^t} \frac{1}{2} \langle \mathbf{V}_t \mathbf{y}, \mathbf{H}\mathbf{V}_t \mathbf{y} \rangle + \langle \mathbf{g}, \mathbf{V}_t \mathbf{y} \rangle \\ &= \arg\min_{\mathbf{y} \in \mathbb{R}^t} \frac{1}{2} \langle \mathbf{y}, \mathbf{T}_t \mathbf{y} \rangle + \langle \|\mathbf{g}\| \, \mathbf{e}_1, \mathbf{y} \rangle .\end{aligned}$$

Taking the derivative and setting it to zero, gives

$$\mathbf{T}_t \mathbf{y}_t = -\|\mathbf{g}\| \, \mathbf{e}_1.$$

Note that $\mathbf{T}_t \succ \mathbf{0}$.

Conjugate Gradient (CG): Assuming $\mathbf{H} \succ \mathbf{0}$,

$$\mathbf{p}_t = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \frac{1}{2} \langle \mathbf{p}, \mathbf{H}\mathbf{p} \rangle + \langle \mathbf{g}, \mathbf{p} \rangle$$

If $\mathbf{V}_t$ is the basis from the Lanczos procedure, then we can set $\mathbf{p}_t = \mathbf{V}_t \mathbf{y}_t$ where $\mathbf{y}_t$ is the solution to:

$$\begin{aligned} \mathbf{y}_t &= \arg\min_{\mathbf{y} \in \mathbb{R}^t} \frac{1}{2} \langle \mathbf{V}_t \mathbf{y}, \mathbf{H}\mathbf{V}_t \mathbf{y} \rangle + \langle \mathbf{g}, \mathbf{V}_t \mathbf{y} \rangle \\ &= \arg\min_{\mathbf{y} \in \mathbb{R}^t} \frac{1}{2} \langle \mathbf{y}, \mathbf{T}_t \mathbf{y} \rangle + \langle \|\mathbf{g}\| \mathbf{e}_1, \mathbf{y} \rangle . \end{aligned}$$

Taking the derivative and setting it to zero, gives

$$\mathbf{T}_t \mathbf{y}_t = -\|\mathbf{g}\| \mathbf{e}_1 .$$

Note that $\mathbf{T}_t \succ \mathbf{0}$. Iterates of CG are generated in a way that conceptually amounts to solving this linear system using Cholesky factorization of $\mathbf{T}_t$.

Algorithm Conjugate Gradient

$\mathbf{r}_0 = \mathbf{d}_1 = -\mathbf{g}$, and $\mathbf{p}_0 = \mathbf{0}$
for $t = 1, 2, \ldots$ until $\|\mathbf{r}_{t-1}\| \leq \tau$ **do**
    $\alpha_t = \langle \mathbf{r}_{t-1}, \mathbf{r}_{t-1} \rangle \,/\, \langle \mathbf{d}_t, \mathbf{H}\mathbf{d}_t \rangle$
    $\mathbf{p}_t = \mathbf{p}_{t-1} + \alpha_t \mathbf{d}_t$
    $\mathbf{r}_t = \mathbf{r}_{t-1} - \alpha_t \mathbf{H}\mathbf{d}_t$
    $\beta_{t+1} = \langle \mathbf{r}_t, \mathbf{r}_t \rangle \,/\, \langle \mathbf{r}_{t-1}, \mathbf{r}_{t-1} \rangle$
    $\mathbf{d}_{t+1} = \mathbf{r}_t + \beta_{t+1} \mathbf{d}_t$
end for

Note: In practice $\mathbf{H}\mathbf{d}_t$ is computed once and reused in various lines, also $\|\mathbf{r}_t\|^2 = \langle \mathbf{r}_t, \mathbf{r}_t \rangle$ from each iteration is reused in the next iteration to check the termination criterion, and also to compute α_t and β_k in the next iteration.

Minimum Residual (MINRES): Assuming $\mathbf{H} = \mathbf{H}^\intercal$,

$$\mathbf{p}_t = \arg\min_{\mathbf{p} \in \mathcal{K}_t} \|\mathbf{H}\mathbf{p} + \mathbf{g}\|$$

Minimum Residual (MINRES): Assuming $\mathbf{H} = \mathbf{H}^\intercal$,

$$\mathbf{p}_t = \underset{\mathbf{p} \in \mathcal{K}_t}{\arg\min} \|\mathbf{H}\mathbf{p} + \mathbf{g}\|$$

If $\mathbf{V}_t$ is the basis from the Lanczos procedure, then we can set $\mathbf{p}_t = \mathbf{V}_t \mathbf{y}_t$ where $\mathbf{y}_t$ is the solution to:

Minimum Residual (MINRES): Assuming $\mathbf{H} = \mathbf{H}^\intercal$,

$$\mathbf{p}_t = \underset{\mathbf{p} \in \mathcal{K}_t}{\arg\min} \|\mathbf{H}\mathbf{p} + \mathbf{g}\|$$

If $\mathbf{V}_t$ is the basis from the Lanczos procedure, then we can set $\mathbf{p}_t = \mathbf{V}_t \mathbf{y}_t$ where $\mathbf{y}_t$ is the solution to:

$$\mathbf{y}_t = \underset{\mathbf{y} \in \mathbb{R}^t}{\arg\min} \|\mathbf{H}\mathbf{V}_t \mathbf{y} + \mathbf{g}\|$$

Minimum Residual (MINRES): Assuming $\mathbf{H} = \mathbf{H}^\intercal$,

$$\mathbf{p}_t = \underset{\mathbf{p} \in \mathcal{K}_t}{\arg\min} \|\mathbf{H}\mathbf{p} + \mathbf{g}\|$$

If $\mathbf{V}_t$ is the basis from the Lanczos procedure, then we can set $\mathbf{p}_t = \mathbf{V}_t \mathbf{y}_t$ where $\mathbf{y}_t$ is the solution to:

$$\begin{aligned}\mathbf{y}_t &= \underset{\mathbf{y} \in \mathbb{R}^t}{\arg\min} \|\mathbf{H}\mathbf{V}_t \mathbf{y} + \mathbf{g}\| \\ &= \underset{\mathbf{y} \in \mathbb{R}^t}{\arg\min} \|\mathbf{V}_{t+1} \mathbf{T}_{t+1,t} \mathbf{y} + \mathbf{g}\|\end{aligned}$$

Minimum Residual (MINRES): Assuming $\mathbf{H} = \mathbf{H}^\intercal$,

$$\mathbf{p}_t = \underset{\mathbf{p} \in \mathcal{K}_t}{\arg\min} \|\mathbf{H}\mathbf{p} + \mathbf{g}\|$$

If $\mathbf{V}_t$ is the basis from the Lanczos procedure, then we can set $\mathbf{p}_t = \mathbf{V}_t \mathbf{y}_t$ where $\mathbf{y}_t$ is the solution to:

$$\begin{aligned}
\mathbf{y}_t &= \underset{\mathbf{y} \in \mathbb{R}^t}{\arg\min} \|\mathbf{H}\mathbf{V}_t \mathbf{y} + \mathbf{g}\| \\
&= \underset{\mathbf{y} \in \mathbb{R}^t}{\arg\min} \|\mathbf{V}_{t+1} \mathbf{T}_{t+1,t} \mathbf{y} + \mathbf{g}\| \\
&= \underset{\mathbf{y} \in \mathbb{R}^t}{\arg\min} \left\| {\color{red}\mathbf{T}_{t+1,t}} \mathbf{y} + {\color{blue}\|\mathbf{g}\| \, \mathbf{e}_1} \right\|.
\end{aligned}$$

Minimum Residual (MINRES): Assuming $\mathbf{H} = \mathbf{H}^\intercal$,

$$\mathbf{p}_t = \underset{\mathbf{p} \in \mathcal{K}_t}{\arg\min} \left\| \mathbf{H}\mathbf{p} + \mathbf{g} \right\|$$

If $\mathbf{V}_t$ is the basis from the Lanczos procedure, then we can set $\mathbf{p}_t = \mathbf{V}_t \mathbf{y}_t$ where $\mathbf{y}_t$ is the solution to:

$$\begin{aligned} \mathbf{y}_t &= \underset{\mathbf{y} \in \mathbb{R}^t}{\arg\min} \left\| \mathbf{H}\mathbf{V}_t \mathbf{y} + \mathbf{g} \right\| \\ &= \underset{\mathbf{y} \in \mathbb{R}^t}{\arg\min} \left\| \mathbf{V}_{t+1} \mathbf{T}_{t+1,t} \mathbf{y} + \mathbf{g} \right\| \\ &= \underset{\mathbf{y} \in \mathbb{R}^t}{\arg\min} \left\| \mathbf{T}_{t+1,t} \mathbf{y} + \left\| \mathbf{g} \right\| \mathbf{e}_1 \right\|. \end{aligned}$$

MINRES iterates are generated in a way that conceptually amounts to solving this least squares using the reduced QR factorization of $\mathbf{T}_{t+1,t}$.

Letting $\mathbf{Q}_{t+1}\mathbf{T}_{t+1,t} = \begin{bmatrix} \mathbf{R}_t \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{(t+1)\times t}$ be the full QR decomposition of $\mathbf{T}_{t+1,t}$ using application of a series of 2×2 Householder reflections.

Letting $\mathbf{Q}_{t+1}\mathbf{T}_{t+1,t} = \begin{bmatrix} \mathbf{R}_t \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{(t+1)\times t}$ be the full QR decomposition of $\mathbf{T}_{t+1,t}$ using application of a series of 2×2 Householder reflections.

Denoting $\mathbf{Q}_{t+1} \|\mathbf{g}\| \, \mathbf{e}_1 = \begin{bmatrix} \mathbf{u}_t \\ \phi_t \end{bmatrix}$, we have

$$\mathbf{y}_t = \arg\min_{\mathbf{y}\in\mathbb{R}^t} \left\| \mathbf{T}_{t+1,t}\mathbf{y} + \|\mathbf{g}\| \, \mathbf{e}_1 \right\| = \arg\min_{\mathbf{y}\in\mathbb{R}^t} \left\| \begin{bmatrix} \mathbf{R}_t \\ \mathbf{0} \end{bmatrix} \mathbf{y} + \begin{bmatrix} \mathbf{u}_t \\ \phi_t \end{bmatrix} . \right\|$$

Letting $\mathbf{Q}_{t+1}\mathbf{T}_{t+1,t} = \begin{bmatrix} \mathbf{R}_t \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{(t+1)\times t}$ be the full QR decomposition of $\mathbf{T}_{t+1,t}$ using application of a series of 2×2 Householder reflections.

Denoting $\mathbf{Q}_{t+1} \|\mathbf{g}\| \, \mathbf{e}_1 = \begin{bmatrix} \mathbf{u}_t \\ \phi_t \end{bmatrix}$, we have

$$\mathbf{y}_t = \arg\min_{\mathbf{y}\in\mathbb{R}^t} \left\| \mathbf{T}_{t+1,t}\mathbf{y} + \|\mathbf{g}\| \, \mathbf{e}_1 \right\| = \arg\min_{\mathbf{y}\in\mathbb{R}^t} \left\| \begin{bmatrix} \mathbf{R}_t \\ \mathbf{0} \end{bmatrix} \mathbf{y} + \begin{bmatrix} \mathbf{u}_t \\ \phi_t \end{bmatrix} . \right\|$$

Let $\mathbf{D}_t = \mathbf{V}_t\mathbf{R}_t^{-1}$ and denote $\mathbf{V}_t = [\mathbf{V}_{t-1} \mid \mathbf{v}_t]$.

Letting $\mathbf{Q}_{t+1}\mathbf{T}_{t+1,t} = \begin{bmatrix} \mathbf{R}_t \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{(t+1)\times t}$ be the full QR decomposition of $\mathbf{T}_{t+1,t}$ using application of a series of 2×2 Householder reflections.

Denoting $\mathbf{Q}_{t+1} \|\mathbf{g}\| \mathbf{e}_1 = \begin{bmatrix} \mathbf{u}_t \\ \phi_t \end{bmatrix}$, we have

$$\mathbf{y}_t = \underset{\mathbf{y}\in\mathbb{R}^t}{\arg\min} \left\| \mathbf{T}_{t+1,t}\mathbf{y} + \|\mathbf{g}\| \mathbf{e}_1 \right\| = \underset{\mathbf{y}\in\mathbb{R}^t}{\arg\min} \left\| \begin{bmatrix} \mathbf{R}_t \\ \mathbf{0} \end{bmatrix} \mathbf{y} + \begin{bmatrix} \mathbf{u}_t \\ \phi_t \end{bmatrix} . \right\|$$

Let $\mathbf{D}_t = \mathbf{V}_t\mathbf{R}_t^{-1}$ and denote $\mathbf{V}_t = [\mathbf{V}_{t-1} \mid \mathbf{v}_t]$. We get

Letting $\mathbf{Q}_{t+1}\mathbf{T}_{t+1,t} = \begin{bmatrix} \mathbf{R}_t \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{(t+1)\times t}$ be the full QR decomposition of $\mathbf{T}_{t+1,t}$ using application of a series of 2×2 Householder reflections.

Denoting $\mathbf{Q}_{t+1} \|\mathbf{g}\| \mathbf{e}_1 = \begin{bmatrix} \mathbf{u}_t \\ \phi_t \end{bmatrix}$, we have

$$\mathbf{y}_t = \arg\min_{\mathbf{y}\in\mathbb{R}^t} \left\| \mathbf{T}_{t+1,t}\mathbf{y} + \|\mathbf{g}\| \mathbf{e}_1 \right\| = \arg\min_{\mathbf{y}\in\mathbb{R}^t} \left\| \begin{bmatrix} \mathbf{R}_t \\ \mathbf{0} \end{bmatrix} \mathbf{y} + \begin{bmatrix} \mathbf{u}_t \\ \phi_t \end{bmatrix} . \right\|$$

Let $\mathbf{D}_t = \mathbf{V}_t\mathbf{R}_t^{-1}$ and denote $\mathbf{V}_t = [\mathbf{V}_{t-1} \mid \mathbf{v}_t]$. We get

$$\mathbf{D}_t = \mathbf{V}_t\mathbf{R}_t^{-1}$$

Letting $\mathbf{Q}_{t+1}\mathbf{T}_{t+1,t} = \begin{bmatrix} \mathbf{R}_t \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{(t+1)\times t}$ be the full QR decomposition of $\mathbf{T}_{t+1,t}$ using application of a series of 2×2 Householder reflections.

Denoting $\mathbf{Q}_{t+1} \|\mathbf{g}\| \mathbf{e}_1 = \begin{bmatrix} \mathbf{u}_t \\ \phi_t \end{bmatrix}$, we have

$$\mathbf{y}_t = \underset{\mathbf{y}\in\mathbb{R}^t}{\arg\min} \left\| \mathbf{T}_{t+1,t}\mathbf{y} + \|\mathbf{g}\| \mathbf{e}_1 \right\| = \underset{\mathbf{y}\in\mathbb{R}^t}{\arg\min} \left\| \begin{bmatrix} \mathbf{R}_t \\ \mathbf{0} \end{bmatrix} \mathbf{y} + \begin{bmatrix} \mathbf{u}_t \\ \phi_t \end{bmatrix} . \right\|$$

Let $\mathbf{D}_t = \mathbf{V}_t\mathbf{R}_t^{-1}$ and denote $\mathbf{V}_t = [\mathbf{V}_{t-1} \mid \mathbf{v}_t]$. We get

$$\mathbf{D}_t = \mathbf{V}_t\mathbf{R}_t^{-1} = \begin{bmatrix} \mathbf{V}_{t-1} & \mathbf{v}_t \end{bmatrix} \begin{bmatrix} \mathbf{R}_{t-1}^{-1} & \star \\ \mathbf{0}^\intercal & \star \end{bmatrix}$$

Letting $\mathbf{Q}_{t+1}\mathbf{T}_{t+1,t} = \begin{bmatrix}\mathbf{R}_t \\ \mathbf{0}\end{bmatrix} \in \mathbb{R}^{(t+1)\times t}$ be the full QR decomposition of $\mathbf{T}_{t+1,t}$ using application of a series of 2×2 Householder reflections.

Denoting $\mathbf{Q}_{t+1} \|\mathbf{g}\| \mathbf{e}_1 = \begin{bmatrix}\mathbf{u}_t \\ \phi_t\end{bmatrix}$, we have

$$\mathbf{y}_t = \arg\min_{\mathbf{y}\in\mathbb{R}^t} \left\| \mathbf{T}_{t+1,t}\mathbf{y} + \|\mathbf{g}\| \mathbf{e}_1 \right\| = \arg\min_{\mathbf{y}\in\mathbb{R}^t} \left\| \begin{bmatrix}\mathbf{R}_t \\ \mathbf{0}\end{bmatrix} \mathbf{y} + \begin{bmatrix}\mathbf{u}_t \\ \phi_t\end{bmatrix} . \right\|$$

Let $\mathbf{D}_t = \mathbf{V}_t\mathbf{R}_t^{-1}$ and denote $\mathbf{V}_t = [\mathbf{V}_{t-1} \mid \mathbf{v}_t]$. We get

$$\begin{aligned}\mathbf{D}_t = \mathbf{V}_t\mathbf{R}_t^{-1} &= \begin{bmatrix}\mathbf{V}_{t-1} & \mathbf{v}_t\end{bmatrix} \begin{bmatrix}\mathbf{R}_{t-1}^{-1} & \star \\ \mathbf{0}^\intercal & \star\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{V}_{t-1}\mathbf{R}_{t-1}^{-1} & \mathbf{d}_t\end{bmatrix}\end{aligned}$$

Letting $\mathbf{Q}_{t+1}\mathbf{T}_{t+1,t} = \begin{bmatrix} \mathbf{R}_t \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{(t+1)\times t}$ be the full QR decomposition of $\mathbf{T}_{t+1,t}$ using application of a series of 2×2 Householder reflections.

Denoting $\mathbf{Q}_{t+1} \|\mathbf{g}\| \mathbf{e}_1 = \begin{bmatrix} \mathbf{u}_t \\ \phi_t \end{bmatrix}$, we have

$$\mathbf{y}_t = \underset{\mathbf{y}\in\mathbb{R}^t}{\arg\min} \left\| \mathbf{T}_{t+1,t}\mathbf{y} + \|\mathbf{g}\| \mathbf{e}_1 \right\| = \underset{\mathbf{y}\in\mathbb{R}^t}{\arg\min} \left\| \begin{bmatrix} \mathbf{R}_t \\ \mathbf{0} \end{bmatrix} \mathbf{y} + \begin{bmatrix} \mathbf{u}_t \\ \phi_t \end{bmatrix} . \right\|$$

Let $\mathbf{D}_t = \mathbf{V}_t\mathbf{R}_t^{-1}$ and denote $\mathbf{V}_t = [\mathbf{V}_{t-1} \mid \mathbf{v}_t]$. We get

$$\begin{aligned} \mathbf{D}_t = \mathbf{V}_t\mathbf{R}_t^{-1} &= \begin{bmatrix} \mathbf{V}_{t-1} & \mathbf{v}_t \end{bmatrix} \begin{bmatrix} \mathbf{R}_{t-1}^{-1} & \star \\ \mathbf{0}^\intercal & \star \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{V}_{t-1}\mathbf{R}_{t-1}^{-1} & \mathbf{d}_t \end{bmatrix} = \begin{bmatrix} \mathbf{D}_{t-1} & \mathbf{d}_t \end{bmatrix} . \end{aligned}$$

As a result, from $\mathbf{y}_t = -\mathbf{R}_t^{-1}\mathbf{u}_t$, we get

$$\mathbf{p}_t = \mathbf{V}_t \mathbf{y}_t$$

As a result, from $\mathbf{y}_t = -\mathbf{R}_t^{-1}\mathbf{u}_t$, we get

$$\mathbf{p}_t = \mathbf{V}_t\mathbf{y}_t = -\mathbf{V}_t\mathbf{R}_t^{-1}\mathbf{u}_t$$

As a result, from $\mathbf{y}_t = -\mathbf{R}_t^{-1}\mathbf{u}_t$, we get

$$\mathbf{p}_t = \mathbf{V}_t\mathbf{y}_t = -\mathbf{V}_t\mathbf{R}_t^{-1}\mathbf{u}_t = -\mathbf{D}_t\mathbf{u}_t$$

As a result, from $\mathbf{y}_t = -\mathbf{R}_t^{-1}\mathbf{u}_t$, we get

$$\begin{aligned}\mathbf{p}_t &= \mathbf{V}_t\mathbf{y}_t = -\mathbf{V}_t\mathbf{R}_t^{-1}\mathbf{u}_t = -\mathbf{D}_t\mathbf{u}_t \\ &= \begin{bmatrix}\mathbf{D}_{t-1} & \mathbf{d}_t\end{bmatrix}\begin{bmatrix}-\mathbf{u}_{t-1} \\ \tau_t\end{bmatrix}\end{aligned}$$

As a result, from $\mathbf{y}_t = -\mathbf{R}_t^{-1}\mathbf{u}_t$, we get

$$\begin{aligned}\mathbf{p}_t &= \mathbf{V}_t\mathbf{y}_t = -\mathbf{V}_t\mathbf{R}_t^{-1}\mathbf{u}_t = -\mathbf{D}_t\mathbf{u}_t \\ &= \begin{bmatrix}\mathbf{D}_{t-1} & \mathbf{d}_t\end{bmatrix}\begin{bmatrix}-\mathbf{u}_{t-1} \\ \tau_t\end{bmatrix} = \mathbf{p}_{t-1} + \tau_t\mathbf{d}_t\end{aligned}$$

As a result, from $\mathbf{y}_t = -\mathbf{R}_t^{-1}\mathbf{u}_t$, we get

$$\mathbf{p}_t = \mathbf{V}_t\mathbf{y}_t = -\mathbf{V}_t\mathbf{R}_t^{-1}\mathbf{u}_t = -\mathbf{D}_t\mathbf{u}_t$$
$$= \begin{bmatrix}\mathbf{D}_{t-1} & \mathbf{d}_t\end{bmatrix}\begin{bmatrix}-\mathbf{u}_{t-1} \\ \tau_t\end{bmatrix} = \mathbf{p}_{t-1} + \tau_t\mathbf{d}_t$$

and from $\mathbf{V}_t = \mathbf{D}_t\mathbf{R}_t$ and the fact that only the diagonal, the super-diagonal, and the second super-diagonal elements of $\mathbf{R}_t$ can be non-zero, we get

$$\mathbf{d}_t = \left(\mathbf{v}_t - \epsilon_t\mathbf{d}_{t-2} - \delta_t^{(2)}\mathbf{d}_{t-1}\right)/\gamma^{(2)}.$$

Algorithm 1 $\textbf{MINRES}(\mathbf{H}, \mathbf{g}, \eta)$

Input: Hessian $\mathbf{H}$, gradient $\mathbf{g}$, and inexactness tolerance $\eta > 0$

$\phi_0 = \tilde{\beta}_1 = \|\mathbf{g}\|$, $\mathbf{r}_0 = -\mathbf{g}$, $\mathbf{v}_1 = \mathbf{r}_0/\phi_0$, $\mathbf{v}_0 = \mathbf{s}_0 = \mathbf{w}_0 = \mathbf{w}_{-1} = \mathbf{0}$,

$s_0 = 0$, $c_0 = -1$, $\delta_1 = \tau_0 = 0$, $t = 1$, $\mathrm{D}_{\mathrm{type}} = $ 'SOL',

while True **do**

  $\mathbf{q}_t = \mathbf{H}\mathbf{v}_t$, $\tilde{\alpha}_t = \mathbf{v}_t^\intercal \mathbf{q}_t$, $\mathbf{q}_t = \mathbf{q}_t - \tilde{\beta}_t \mathbf{v}_{t-1}$, $\mathbf{q}_t = \mathbf{q}_t - \tilde{\alpha}_t \mathbf{v}_t$, $\tilde{\beta}_{t+1} = \|\mathbf{q}_t\|$

  $$\begin{bmatrix} \delta_t^{[2]} & \epsilon_{t+1} \\ \gamma_t & \delta_{t+1} \end{bmatrix} = \begin{bmatrix} c_{t-1} & s_{t-1} \\ s_{t-1} & -c_{t-1} \end{bmatrix} \begin{bmatrix} \delta_t & 0 \\ \tilde{\alpha}_t & \tilde{\beta}_{t+1} \end{bmatrix}$$

  if $c_{t-1}\gamma_t \geq 0$ **then**

    $\mathrm{D}_{\mathrm{type}} = $ 'NPC'

    return $\mathbf{r}_{t-1}$, $\mathrm{D}_{\mathrm{type}}$.

  end if

  if $\phi_{t-1}\sqrt{\gamma_t^2 + \delta_{t+1}^2} \leq \eta\sqrt{\phi_0^2 - \phi_{t-1}^2}$ **then**

    $\mathrm{D}_{\mathrm{type}} = $ 'SOL'

    return $\mathbf{s}_{t-1}$, $\mathrm{D}_{\mathrm{type}}$

  end if

  $\gamma_t^{[2]} = \sqrt{\gamma_t^2 + \tilde{\beta}_{t+1}^2}$,

  if $\gamma_t^{[2]} \neq 0$ **then**

    $c_t = \gamma_t/\gamma_t^{[2]}$, $s_t = \tilde{\beta}_{t+1}/\gamma_t^{[2]}$, $\tau_t = c_t\phi_{t-1}$, $\phi_t = s_t\phi_{t-1}$,

    $\mathbf{w}_t = \left(\mathbf{v}_t - \delta_t^{[2]}\mathbf{w}_{t-1} - \epsilon_t \mathbf{w}_{t-2}\right)/\gamma_t^{[2]}$, $\mathbf{s}_t = \mathbf{s}_{t-1} + \tau_t \mathbf{w}_t$

    if $\tilde{\beta}_{t+1} \neq 0$ **then**

      $\mathbf{v}_{t+1} = \mathbf{q}_t/\tilde{\beta}_{t+1}$, $\mathbf{r}_t = s_t^2 \mathbf{r}_{t-1} - \phi_t c_t \mathbf{v}_{t+1}$,

    end if

  else

    $c_t = 0$, $s_t = 1$, $\tau_t = 0$, $\phi_t = \phi_{t-1}$, $\mathbf{r}_t = \mathbf{r}_{t-1}$, $\mathbf{s}_t = \mathbf{s}_{t-1}$,

  end if

  $t \leftarrow t + 1$,

end while

CG VERSUS MINRES: AN EMPIRICAL COMPARISON*

DAVID CHIN-LUNG FONG† AND MICHAEL SAUNDERS‡

Abstract. For iterative solution of symmetric systems $Ax = b$, the conjugate gradient method (CG) is commonly used when A is positive definite, while the minimum residual method (MINRES) is typically reserved for indefinite systems. We investigate the sequence of approximate solutions x_k generated by each method and suggest that even if A is positive definite, MINRES may be preferable to CG if iterations are to be terminated early. In particular, we show for MINRES that the solution norms $\|x_k\|$ are monotonically increasing when A is positive definite (as was already known for CG), and the solution errors $\|x^* - x_k\|$ are monotonically decreasing. We also show that the backward errors for the MINRES iterates x_k are monotonically decreasing.

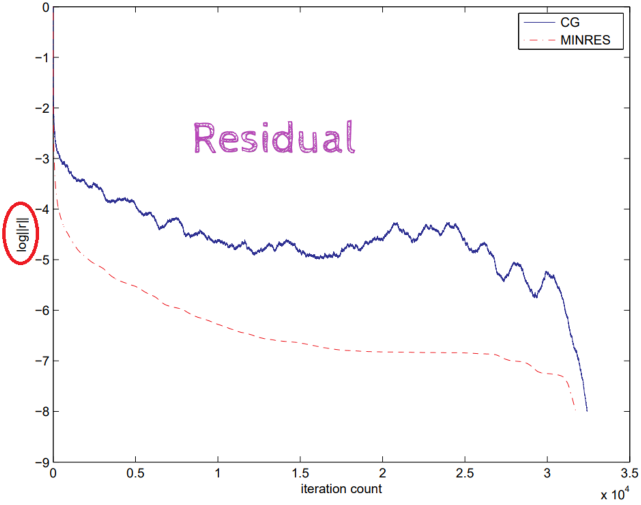
Key words. conjugate gradient method, minimum residual method, iterative method, sparse matrix, linear equations, CG, CR, MINRES, Krylov subspace method, trust-region method

1. Introduction. The conjugate gradient method (CG) [11] and the minimum residual method (MINRES) [18] are both Krylov subspace methods for the iterative solution of symmetric linear equations $Ax = b$. CG is commonly used when the matrix A is positive definite, while MINRES is generally reserved for indefinite systems [27, p85]. We reexamine this wisdom from the point of view of early termination on positive-definite systems.

We assume that the system $Ax = b$ is real with A symmetric positive definite (spd) and of dimension $n \times n$. The Lanczos process [13] with starting vector b may be used to generate the $n \times k$ matrix $V_k \equiv \begin{pmatrix} v_1 & v_2 & \dots & v_k \end{pmatrix}$ and the $(k+1) \times k$

(Fong and Saunders, 2012)

MINRES terminates faster!



(Fong and Saunders, 2012)

Q: Why ubiquitous reliance on CG?

	CG	MINRES

Q: Why ubiquitous reliance on CG?

	CG	MINRES
Simplicity	✓	✗

Q: Why ubiquitous reliance on CG?

	CG	MINRES
Simplicity	✓	✗
Coverage in Textbook	✓	✗

Q: Why ubiquitous reliance on CG?

	CG	MINRES
Simplicity	✓	✗
Coverage in Textbook	✓	✗
Software Libraries	✓	✗

Q: Why ubiquitous reliance on CG?

	CG	MINRES
Simplicity	✓	✗
Coverage in Textbook	✓	✗
Software Libraries	✓	✗
Theoretical Properties	✓	?

Q: Why ubiquitous reliance on CG?

	CG	MINRES
Simplicity	✓	✗
Coverage in Textbook	✓	✗
Software Libraries	✓	✗
Theoretical Properties	✓	?
Numerical Properties	✓	✗

CG has optimal convergence rate ($\mathbf{H} \succ \mathbf{0}$):

CG has optimal convergence rate ($\mathbf{H} \succ \mathbf{0}$):
letting $\mathbf{p} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g}) = \mathrm{Span}\{\mathbf{g}, \mathbf{H}\mathbf{g}, \ldots, \mathbf{H}^{t-1}\mathbf{g}\}$, and noting that $\mathbf{H}\mathbf{p}^\star = -\mathbf{g}$, we have

,

CG has optimal convergence rate ($\mathbf{H} \succ \mathbf{0}$):
letting $\mathbf{p} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g}) = \mathrm{Span}\{\mathbf{g}, \mathbf{H}\mathbf{g}, \ldots, \mathbf{H}^{t-1}\mathbf{g}\}$, and noting that $\mathbf{H}\mathbf{p}^\star = -\mathbf{g}$, we have

$$\mathbf{p} = -p_{t-1}(\mathbf{H})\mathbf{g} \qquad ,$$

CG has optimal convergence rate ($\mathbf{H} \succ \mathbf{0}$):
letting $\mathbf{p} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g}) = \text{Span}\{\mathbf{g}, \mathbf{Hg}, \ldots, \mathbf{H}^{t-1}\mathbf{g}\}$, and noting that $\mathbf{Hp}^\star = -\mathbf{g}$, we have

$$\mathbf{p} = -p_{t-1}(\mathbf{H})\mathbf{g} \Longrightarrow \mathbf{p} - \mathbf{p}^\star = r_t(\mathbf{H})\mathbf{p}^\star,$$

CG has optimal convergence rate ($\mathbf{H} \succ \mathbf{0}$):
letting $\mathbf{p} \in \mathcal{K}_t\left(\mathbf{H}, \mathbf{g}\right) = \text{Span}\{\mathbf{g}, \mathbf{Hg}, \ldots, \mathbf{H}^{t-1}\mathbf{g}\}$, and noting that $\mathbf{Hp}^\star = -\mathbf{g}$, we have

$$\mathbf{p} = -p_{t-1}(\mathbf{H})\mathbf{g} \Longrightarrow \mathbf{p} - \mathbf{p}^\star = r_t(\mathbf{H})\mathbf{p}^\star,$$

where $r_t(\mathbf{H}) \triangleq \mathbf{I} - \mathbf{H}p_{t-1}(\mathbf{H})$ is a residual polynomial of degree t.

CG has optimal convergence rate ($\mathbf{H} \succ \mathbf{0}$):
letting $\mathbf{p} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g}) = \text{Span}\{\mathbf{g}, \mathbf{H}\mathbf{g}, \ldots, \mathbf{H}^{t-1}\mathbf{g}\}$, and noting that $\mathbf{H}\mathbf{p}^\star = -\mathbf{g}$, we have

$$\mathbf{p} = -p_{t-1}(\mathbf{H})\mathbf{g} \Longrightarrow \mathbf{p} - \mathbf{p}^\star = r_t(\mathbf{H})\mathbf{p}^\star,$$

where $r_t(\mathbf{H}) \triangleq \mathbf{I} - \mathbf{H}p_{t-1}(\mathbf{H})$ is a residual polynomial of degree t. Recall that if $\lambda \in \text{spec}(\mathbf{H})$, then $r_t(\lambda) \in \text{spec}(r_t(\mathbf{H}))$ (Spectral Mapping Theorem for Matrix Polynomials).

CG has optimal convergence rate ($\mathbf{H} \succ \mathbf{0}$):
letting $\mathbf{p} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g}) = \text{Span}\{\mathbf{g}, \mathbf{Hg}, \ldots, \mathbf{H}^{t-1}\mathbf{g}\}$, and noting that $\mathbf{Hp}^\star = -\mathbf{g}$, we have

$$\mathbf{p} = -p_{t-1}(\mathbf{H})\mathbf{g} \Longrightarrow \mathbf{p} - \mathbf{p}^\star = r_t(\mathbf{H})\mathbf{p}^\star,$$

where $r_t(\mathbf{H}) \triangleq \mathbf{I} - \mathbf{H}p_{t-1}(\mathbf{H})$ is a residual polynomial of degree t. Recall that if $\lambda \in \text{spec}(\mathbf{H})$, then $r_t(\lambda) \in \text{spec}(r_t(\mathbf{H}))$ (Spectral Mapping Theorem for Matrix Polynomials).

For CG, we had $\mathbf{p}_t = \underset{\mathbf{p} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g})}{\arg\min} \|\mathbf{p} - \mathbf{p}^\star\|_{\mathbf{H}}$, which implies

CG has optimal convergence rate ($\mathbf{H} \succ \mathbf{0}$):
letting $\mathbf{p} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g}) = \mathrm{Span}\{\mathbf{g}, \mathbf{Hg}, \ldots, \mathbf{H}^{t-1}\mathbf{g}\}$, and noting that $\mathbf{Hp}^\star = -\mathbf{g}$, we have

$$\mathbf{p} = -p_{t-1}(\mathbf{H})\mathbf{g} \Longrightarrow \mathbf{p} - \mathbf{p}^\star = r_t(\mathbf{H})\mathbf{p}^\star,$$

where $r_t(\mathbf{H}) \triangleq \mathbf{I} - \mathbf{H}p_{t-1}(\mathbf{H})$ is a residual polynomial of degree t. Recall that if $\lambda \in \mathrm{spec}(\mathbf{H})$, then $r_t(\lambda) \in \mathrm{spec}(r_t(\mathbf{H}))$ (Spectral Mapping Theorem for Matrix Polynomials).

For CG, we had $\mathbf{p}_t = \underset{\mathbf{p} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g})}{\arg\min} \|\mathbf{p} - \mathbf{p}^\star\|_{\mathbf{H}}$, which implies

$$\|\mathbf{p}_t - \mathbf{p}^\star\|_{\mathbf{H}}$$

CG has optimal convergence rate ($\mathbf{H} \succ \mathbf{0}$):
letting $\mathbf{p} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g}) = \text{Span}\{\mathbf{g}, \mathbf{Hg}, \ldots, \mathbf{H}^{t-1}\mathbf{g}\}$, and noting that $\mathbf{Hp}^\star = -\mathbf{g}$, we have

$$\mathbf{p} = -p_{t-1}(\mathbf{H})\mathbf{g} \Longrightarrow \mathbf{p} - \mathbf{p}^\star = r_t(\mathbf{H})\mathbf{p}^\star,$$

where $r_t(\mathbf{H}) \triangleq \mathbf{I} - \mathbf{H}p_{t-1}(\mathbf{H})$ is a residual polynomial of degree t. Recall that if $\lambda \in \text{spec}(\mathbf{H})$, then $r_t(\lambda) \in \text{spec}(r_t(\mathbf{H}))$ (Spectral Mapping Theorem for Matrix Polynomials).

For CG, we had $\mathbf{p}_t = \underset{\mathbf{p} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g})}{\arg\min} \|\mathbf{p} - \mathbf{p}^\star\|_{\mathbf{H}}$, which implies

$$\|\mathbf{p}_t - \mathbf{p}^\star\|_{\mathbf{H}} = \min_{\mathbf{p} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g})} \|\mathbf{p} - \mathbf{p}^\star\|_{\mathbf{H}}$$

CG has optimal convergence rate ($\mathbf{H} \succ \mathbf{0}$):
letting $\mathbf{p} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g}) = \text{Span}\{\mathbf{g}, \mathbf{Hg}, \ldots, \mathbf{H}^{t-1}\mathbf{g}\}$, and noting that $\mathbf{Hp}^\star = -\mathbf{g}$, we have

$$\mathbf{p} = -p_{t-1}(\mathbf{H})\mathbf{g} \Longrightarrow \mathbf{p} - \mathbf{p}^\star = r_t(\mathbf{H})\mathbf{p}^\star,$$

where $r_t(\mathbf{H}) \triangleq \mathbf{I} - \mathbf{H}p_{t-1}(\mathbf{H})$ is a residual polynomial of degree t. Recall that if $\lambda \in \text{spec}(\mathbf{H})$, then $r_t(\lambda) \in \text{spec}(r_t(\mathbf{H}))$ (Spectral Mapping Theorem for Matrix Polynomials).

For CG, we had $\mathbf{p}_t = \underset{\mathbf{p} \in \mathcal{K}_t(\mathbf{H},\mathbf{g})}{\arg\min} \|\mathbf{p} - \mathbf{p}^\star\|_{\mathbf{H}}$, which implies

$$\|\mathbf{p}_t - \mathbf{p}^\star\|_{\mathbf{H}} = \min_{\mathbf{p} \in \mathcal{K}_t(\mathbf{H},\mathbf{g})} \|\mathbf{p} - \mathbf{p}^\star\|_{\mathbf{H}} = \min_{r_t \in \Pi_t} \|r_t(\mathbf{H})\mathbf{p}^\star\|_{\mathbf{H}}$$

CG has optimal convergence rate ($\mathbf{H} \succ \mathbf{0}$):
letting $\mathbf{p} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g}) = \text{Span}\{\mathbf{g}, \mathbf{H}\mathbf{g}, \ldots, \mathbf{H}^{t-1}\mathbf{g}\}$, and noting that $\mathbf{H}\mathbf{p}^\star = -\mathbf{g}$, we have

$$\mathbf{p} = -p_{t-1}(\mathbf{H})\mathbf{g} \Longrightarrow \mathbf{p} - \mathbf{p}^\star = r_t(\mathbf{H})\mathbf{p}^\star,$$

where $r_t(\mathbf{H}) \triangleq \mathbf{I} - \mathbf{H}p_{t-1}(\mathbf{H})$ is a residual polynomial of degree t. Recall that if $\lambda \in \text{spec}(\mathbf{H})$, then $r_t(\lambda) \in \text{spec}(r_t(\mathbf{H}))$ (Spectral Mapping Theorem for Matrix Polynomials).

For CG, we had $\mathbf{p}_t = \underset{\mathbf{p} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g})}{\arg\min} \|\mathbf{p} - \mathbf{p}^\star\|_{\mathbf{H}}$, which implies

$$\begin{aligned} \|\mathbf{p}_t - \mathbf{p}^\star\|_{\mathbf{H}} &= \min_{\mathbf{p} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g})} \|\mathbf{p} - \mathbf{p}^\star\|_{\mathbf{H}} = \min_{r_t \in \Pi_t} \|r_t(\mathbf{H})\mathbf{p}^\star\|_{\mathbf{H}} \\ &\leq \|\mathbf{p}^\star\|_{\mathbf{H}} \min_{r_t \in \Pi_t} \|r_t(\mathbf{H})\| \end{aligned}$$

CG has optimal convergence rate ($\mathbf{H} \succ \mathbf{0}$):
letting $\mathbf{p} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g}) = \text{Span}\{\mathbf{g}, \mathbf{H}\mathbf{g}, \ldots, \mathbf{H}^{t-1}\mathbf{g}\}$, and noting that $\mathbf{H}\mathbf{p}^\star = -\mathbf{g}$, we have

$$\mathbf{p} = -p_{t-1}(\mathbf{H})\mathbf{g} \Longrightarrow \mathbf{p} - \mathbf{p}^\star = r_t(\mathbf{H})\mathbf{p}^\star,$$

where $r_t(\mathbf{H}) \triangleq \mathbf{I} - \mathbf{H}p_{t-1}(\mathbf{H})$ is a residual polynomial of degree t. Recall that if $\lambda \in \text{spec}(\mathbf{H})$, then $r_t(\lambda) \in \text{spec}(r_t(\mathbf{H}))$ (Spectral Mapping Theorem for Matrix Polynomials).

For CG, we had $\mathbf{p}_t = \underset{\mathbf{p} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g})}{\arg\min} \|\mathbf{p} - \mathbf{p}^\star\|_{\mathbf{H}}$, which implies

$$\begin{aligned}
\|\mathbf{p}_t - \mathbf{p}^\star\|_{\mathbf{H}} &= \min_{\mathbf{p} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g})} \|\mathbf{p} - \mathbf{p}^\star\|_{\mathbf{H}} = \min_{r_t \in \Pi_t} \|r_t(\mathbf{H})\mathbf{p}^\star\|_{\mathbf{H}} \\
&\leq \|\mathbf{p}^\star\|_{\mathbf{H}} \min_{r_t \in \Pi_t} \|r_t(\mathbf{H})\| \\
&\leq \|\mathbf{p}^\star\|_{\mathbf{H}} \min_{r_t \in \Pi_t} \max_{\lambda \in \text{spec}(\mathbf{H})} |r_t(\lambda)|.
\end{aligned}$$

CG has optimal convergence rate ($\mathbf{H} \succ \mathbf{0}$):
letting $\mathbf{p} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g}) = \text{Span}\{\mathbf{g}, \mathbf{H}\mathbf{g}, \ldots, \mathbf{H}^{t-1}\mathbf{g}\}$, and noting that $\mathbf{H}\mathbf{p}^\star = -\mathbf{g}$, we have

$$\mathbf{p} = -p_{t-1}(\mathbf{H})\mathbf{g} \Longrightarrow \mathbf{p} - \mathbf{p}^\star = r_t(\mathbf{H})\mathbf{p}^\star,$$

where $r_t(\mathbf{H}) \triangleq \mathbf{I} - \mathbf{H}p_{t-1}(\mathbf{H})$ is a residual polynomial of degree t. Recall that if $\lambda \in \text{spec}(\mathbf{H})$, then $r_t(\lambda) \in \text{spec}(r_t(\mathbf{H}))$ (Spectral Mapping Theorem for Matrix Polynomials).

For CG, we had $\mathbf{p}_t = \underset{\mathbf{p} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g})}{\arg\min} \|\mathbf{p} - \mathbf{p}^\star\|_{\mathbf{H}}$, which implies

$$\begin{aligned}
\|\mathbf{p}_t - \mathbf{p}^\star\|_{\mathbf{H}} &= \min_{\mathbf{p} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g})} \|\mathbf{p} - \mathbf{p}^\star\|_{\mathbf{H}} = \min_{r_t \in \Pi_t} \|r_t(\mathbf{H})\mathbf{p}^\star\|_{\mathbf{H}} \\
&\leq \|\mathbf{p}^\star\|_{\mathbf{H}} \min_{r_t \in \Pi_t} \|r_t(\mathbf{H})\| \\
&\leq \|\mathbf{p}^\star\|_{\mathbf{H}} \min_{r_t \in \Pi_t} \max_{\lambda \in \text{spec}(\mathbf{H})} |r_t(\lambda)|.
\end{aligned}$$

So, using properties of Chebyshev polynomials, we get

$$\|\mathbf{p}_t - \mathbf{p}^\star\|_{\mathbf{H}} \leq 2 \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^t \|\mathbf{p}^\star\|_{\mathbf{H}}.$$

One can show a similar bound for MINRES when $\mathbf{H} \succeq \mathbf{0}$:

$$\|\mathbf{r}_t\| \leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^t \|\mathbf{g}\|.$$

Lemma (Liu and Roosta, 2022a)

Let

$$\mathbf{H} = \begin{bmatrix} \mathbf{U} & \mathbf{U}_\perp & \mathbf{U}_n \end{bmatrix} \begin{bmatrix} \Lambda & & \\ & \Lambda_\perp & \\ & & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{U} & \mathbf{U}_\perp & \mathbf{U}_n \end{bmatrix}^\intercal .$$

Lemma (Liu and Roosta, 2022a)

Let

$$\mathbf{H} = \begin{bmatrix} \mathbf{U} & \mathbf{U}_{\perp} & \mathbf{U}_{n} \end{bmatrix} \begin{bmatrix} \Lambda & & \\ & \Lambda_{\perp} & \\ & & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{U} & \mathbf{U}_{\perp} & \mathbf{U}_{n} \end{bmatrix}^{\intercal}.$$

For any $0 < \theta < 1$

Lemma (Liu and Roosta, 2022a)

Let

$$\mathbf{H} = \begin{bmatrix} \mathbf{U} & \mathbf{U}_\perp & \mathbf{U}_n \end{bmatrix} \begin{bmatrix} \Lambda & & \\ & \Lambda_\perp & \\ & & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{U} & \mathbf{U}_\perp & \mathbf{U}_n \end{bmatrix}^\intercal .$$

For any $0 < \theta < 1$, *after at most*

$$t \geq 0.25\sqrt{\max\{\kappa^+, \kappa^-\}} \log\left(4/\theta\right),$$

iterations of MINRES

Lemma (Liu and Roosta, 2022a)

Let

$$\mathbf{H} = \begin{bmatrix} \mathbf{U} & \mathbf{U}_{\perp} & \mathbf{U}_n \end{bmatrix} \begin{bmatrix} \Lambda & & \\ & \Lambda_{\perp} & \\ & & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{U} & \mathbf{U}_{\perp} & \mathbf{U}_n \end{bmatrix}^{\intercal}.$$

For any $0 < \theta < 1$, *after at most*

$$t \geq 0.25 \sqrt{\max\{\kappa^{+}, \kappa^{-}\}} \log\left(4/\theta\right),$$

iterations of MINRES, we have

$$\left\|\mathbf{U}_n \mathbf{U}_n^{\intercal} \mathbf{g}\right\|^2 \leq \left\|\mathbf{r}_t\right\|^2 \leq \left\|\mathbf{U}_n \mathbf{U}_n^{\intercal} \mathbf{g}\right\|^2 + \theta \left\|\mathbf{U}\mathbf{U}^{\intercal} \mathbf{g}\right\|^2,$$

Lemma (Liu and Roosta, 2022a)

Let

$$\mathbf{H} = \begin{bmatrix} \mathbf{U} & \mathbf{U}_\perp & \mathbf{U}_n \end{bmatrix} \begin{bmatrix} \Lambda & & \\ & \Lambda_\perp & \\ & & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{U} & \mathbf{U}_\perp & \mathbf{U}_n \end{bmatrix}^\intercal .$$

For any $0 < \theta < 1$, *after at most*

$$t \geq 0.25\sqrt{\max\{\kappa^+, \kappa^-\}} \log\left(4/\theta\right),$$

iterations of MINRES, we have

$$\left\|\mathbf{U}_n\mathbf{U}_n^\intercal\mathbf{g}\right\|^2 \leq \left\|\mathbf{r}_t\right\|^2 \leq \left\|\mathbf{U}_n\mathbf{U}_n^\intercal\mathbf{g}\right\|^2 + \theta\left\|\mathbf{U}\mathbf{U}^\intercal\mathbf{g}\right\|^2,$$

and in particular,

$$\left\|\mathbf{U}\mathbf{U}^\intercal\mathbf{r}_t\right\|^2 \leq \theta\left\|\mathbf{U}\mathbf{U}^\intercal\mathbf{g}\right\|^2.$$

Lemma (Liu and Roosta, 2022a)

Let

$$\mathbf{H} = \begin{bmatrix}\mathbf{U} & \mathbf{U}_\perp & \mathbf{U}_n\end{bmatrix} \begin{bmatrix}\Lambda & & \\ & \Lambda_\perp & \\ & & \mathbf{0}\end{bmatrix} \begin{bmatrix}\mathbf{U} & \mathbf{U}_\perp & \mathbf{U}_n\end{bmatrix}^\intercal .$$

For any $0 < \theta < 1$, *after at most*

$$t \geq 0.25\sqrt{\max\{\kappa^+, \kappa^-\}} \log\left(4/\theta\right),$$

iterations of MINRES, we have

$$\left\|\mathbf{U}_n\mathbf{U}_n^\intercal\mathbf{g}\right\|^2 \leq \left\|\mathbf{r}_t\right\|^2 \leq \left\|\mathbf{U}_n\mathbf{U}_n^\intercal\mathbf{g}\right\|^2 + \theta\left\|\mathbf{U}\mathbf{U}^\intercal\mathbf{g}\right\|^2,$$

and in particular,

$$\left\|\mathbf{U}\mathbf{U}^\intercal\mathbf{r}_t\right\|^2 \leq \theta\left\|\mathbf{U}\mathbf{U}^\intercal\mathbf{g}\right\|^2.$$

Available general convergence results for indefinite problems imply rates depending on κ^+ and κ^- as opposed to $\sqrt{\kappa^+}$ and $\sqrt{\kappa^-}$.

One of the nice properties of CG is that it allows for ready access to NPC direction:

CG

$$\min_{\mathbf{p} \in \mathcal{K}_t(\mathbf{H}_k, \mathbf{g})} \langle \mathbf{p}, \mathbf{g} \rangle + \langle \mathbf{p}, \mathbf{Hp} \rangle /2$$

One of the nice properties of CG is that it allows for ready access to NPC direction:

CG

$$\min_{\mathbf{p} \in \mathcal{K}_t(\mathbf{H}_k, \mathbf{g})} \langle \mathbf{p}, \mathbf{g} \rangle + \langle \mathbf{p}, \mathbf{H}\mathbf{p} \rangle / 2$$

One of the nice properties of CG is that it allows for ready access to NPC direction:

CG

$$\min_{\mathbf{p} \in \mathcal{K}_t(\mathbf{H}_k, \mathbf{g})} \langle \mathbf{p}, \mathbf{g} \rangle + \langle \mathbf{p}, \mathbf{H}\mathbf{p} \rangle / 2$$

CG's NPC Condition

$$\langle \mathbf{d}_t, \mathbf{H}\mathbf{d}_t \rangle \leq 0$$

One of the nice properties of CG is that it allows for ready access to NPC direction:

CG

$$\min_{\mathbf{p} \in \mathcal{K}_t(\mathbf{H}_k, \mathbf{g})} \langle \mathbf{p}, \mathbf{g} \rangle + \langle \mathbf{p}, \mathbf{H}\mathbf{p} \rangle / 2$$

CG's NPC Condition

$$\langle \mathbf{d}_t, \mathbf{H}\mathbf{d}_t \rangle \leq 0$$

The NPC condition can be checked almost for free in CG since we always compute $\langle \mathbf{d}_t, \mathbf{H}\mathbf{d}_t \rangle$ in every iteration to find CG's step size α_t.

Minimum Residual

$$\min_{\mathbf{p}\in\mathcal{K}_t(\mathbf{H},\mathbf{g})} \left\|\mathbf{H}\mathbf{p}+\mathbf{g}\right\|^2$$

Minimum Residual

$$\min_{\mathbf{p}\in\mathcal{K}_t(\mathbf{H},\mathbf{g})} \langle \mathbf{p}, \mathbf{H}\mathbf{g}\rangle + \left\langle \mathbf{p}, \mathbf{H}^2\mathbf{p} \right\rangle /2$$

Minimum Residual

$$\min_{\mathbf{p}\in\mathcal{K}_t(\mathbf{H},\mathbf{g})} \langle \mathbf{p}, \mathbf{H}\mathbf{g} \rangle + \underbrace{\left\langle \mathbf{p}, \mathbf{H}^2\mathbf{p} \right\rangle}_{\frown} /2$$

Minimum Residual

$$\min_{\mathbf{p}\in\mathcal{K}_t(\mathbf{H},\mathbf{g})} \langle \mathbf{p}, \mathbf{H}\mathbf{g}\rangle + \underbrace{\left\langle \mathbf{p}, \mathbf{H}^2\mathbf{p}\right\rangle}_{\frown} /2$$

By construction, at every iteration of MINRES, we have

$$\mathbf{p}_t = \underset{\mathbf{p} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g})}{\arg\min} \, \|\mathbf{H}\mathbf{p} + \mathbf{g}\|^2$$

By construction, at every iteration of MINRES, we have

$$\mathbf{p}_t = \underset{\mathbf{p}\in\mathcal{K}_t(\mathbf{H},\mathbf{g})}{\arg\min} \left\|\mathbf{H}\mathbf{p} + \mathbf{g}\right\|^2 \iff \mathbf{r}_t = \mathbf{H}\mathbf{p}_t + \mathbf{g} \perp \mathbf{H}\mathcal{K}_t$$

By construction, at every iteration of MINRES, we have

$$\mathbf{p}_t = \underset{\mathbf{p} \in \mathcal{K}_t(\mathbf{H},\mathbf{g})}{\arg\min} \left\|\mathbf{H}\mathbf{p} + \mathbf{g}\right\|^2 \iff \mathbf{r}_t = \mathbf{H}\mathbf{p}_t + \mathbf{g} \perp \mathbf{H}\mathcal{K}_t$$

It is easy to show that $\mathbf{r}_i \in \mathcal{K}_t(\mathbf{H}, \mathbf{g})$ for any $i \leq t - 1$.

By construction, at every iteration of MINRES, we have

$$\mathbf{p}_t = \underset{\mathbf{p}\in\mathcal{K}_t(\mathbf{H},\mathbf{g})}{\arg\min} \|\mathbf{H}\mathbf{p}+\mathbf{g}\|^2 \iff \mathbf{r}_t = \mathbf{H}\mathbf{p}_t + \mathbf{g} \perp \mathbf{H}\mathcal{K}_t$$

It is easy to show that $\mathbf{r}_i \in \mathcal{K}_t(\mathbf{H},\mathbf{g})$ for any $i \leq t-1$. This implies that

$$\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_j \rangle = 0, \quad i \neq j.$$

By construction, at every iteration of MINRES, we have

$$\mathbf{p}_t = \underset{\mathbf{p}\in\mathcal{K}_t(\mathbf{H},\mathbf{g})}{\arg\min} \left\|\mathbf{H}\mathbf{p} + \mathbf{g}\right\|^2 \iff \mathbf{r}_t = \mathbf{H}\mathbf{p}_t + \mathbf{g} \perp \mathbf{H}\mathcal{K}_t$$

It is easy to show that $\mathbf{r}_i \in \mathcal{K}_t(\mathbf{H},\mathbf{g})$ for any $i \leq t-1$. This implies that

$$\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_j \rangle = 0, \quad i \neq j.$$

One can also show that as long as $\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_i \rangle \neq 0$, for all $0 \leq i \leq t-1$, then $\text{Span}\{\mathbf{r}_0, \mathbf{r}_1, \ldots, \mathbf{r}_{t-1}\} = \mathcal{K}_t(\mathbf{H},\mathbf{g})$.

By construction, at every iteration of MINRES, we have

$$\mathbf{p}_t = \underset{\mathbf{p}\in\mathcal{K}_t(\mathbf{H},\mathbf{g})}{\arg\min} \|\mathbf{H}\mathbf{p} + \mathbf{g}\|^2 \iff \mathbf{r}_t = \mathbf{H}\mathbf{p}_t + \mathbf{g} \perp \mathbf{H}\mathcal{K}_t$$

It is easy to show that $\mathbf{r}_i \in \mathcal{K}_t(\mathbf{H}, \mathbf{g})$ for any $i \leq t-1$. This implies that

$$\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_j \rangle = 0, \quad i \neq j.$$

One can also show that as long as $\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_i \rangle \neq 0$, for all $0 \leq i \leq t-1$, then $\text{Span}\{\mathbf{r}_0, \mathbf{r}_1, \ldots, \mathbf{r}_{t-1}\} = \mathcal{K}_t(\mathbf{H}, \mathbf{g})$. In particular, suppose $\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_i \rangle > 0$, for all $0 \leq i \leq t-1$ and let $\mathbf{0} \neq \mathbf{v} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g})$.

By construction, at every iteration of MINRES, we have

$$\mathbf{p}_t = \underset{\mathbf{p} \in \mathcal{K}_t(\mathbf{H},\mathbf{g})}{\arg\min} \left\|\mathbf{H}\mathbf{p} + \mathbf{g}\right\|^2 \iff \mathbf{r}_t = \mathbf{H}\mathbf{p}_t + \mathbf{g} \perp \mathbf{H}\mathcal{K}_t$$

It is easy to show that $\mathbf{r}_i \in \mathcal{K}_t(\mathbf{H}, \mathbf{g})$ for any $i \leq t-1$. This implies that

$$\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_j \rangle = 0, \quad i \neq j.$$

One can also show that as long as $\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_i \rangle \neq 0$, for all $0 \leq i \leq t-1$, then $\text{Span}\{\mathbf{r}_0, \mathbf{r}_1, \ldots, \mathbf{r}_{t-1}\} = \mathcal{K}_t(\mathbf{H}, \mathbf{g})$. In particular, suppose $\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_i \rangle > 0$, for all $0 \leq i \leq t-1$ and let $\mathbf{0} \neq \mathbf{v} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g})$. We can write $\mathbf{v} = [\mathbf{r}_0 \mid \mathbf{r}_1 \mid \ldots \mid \mathbf{r}_{t-1}]\mathbf{c}$ for some $\mathbf{0} \neq \mathbf{c} \in \mathbb{R}^t$.

By construction, at every iteration of MINRES, we have

$$\mathbf{p}_t = \underset{\mathbf{p} \in \mathcal{K}_t(\mathbf{H},\mathbf{g})}{\arg\min} \left\|\mathbf{H}\mathbf{p} + \mathbf{g}\right\|^2 \iff \mathbf{r}_t = \mathbf{H}\mathbf{p}_t + \mathbf{g} \perp \mathbf{H}\mathcal{K}_t$$

It is easy to show that $\mathbf{r}_i \in \mathcal{K}_t(\mathbf{H}, \mathbf{g})$ for any $i \leq t-1$. This implies that

$$\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_j \rangle = 0, \quad i \neq j.$$

One can also show that as long as $\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_i \rangle \neq 0$, for all $0 \leq i \leq t-1$, then $\text{Span}\{\mathbf{r}_0, \mathbf{r}_1, \ldots, \mathbf{r}_{t-1}\} = \mathcal{K}_t(\mathbf{H}, \mathbf{g})$. In particular, suppose $\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_i \rangle > 0$, for all $0 \leq i \leq t-1$ and let $\mathbf{0} \neq \mathbf{v} \in \mathcal{K}_t(\mathbf{H}, \mathbf{g})$. We can write $\mathbf{v} = [\mathbf{r}_0 \mid \mathbf{r}_1 \mid \ldots \mid \mathbf{r}_{t-1}]\mathbf{c}$ for some $\mathbf{0} \neq \mathbf{c} \in \mathbb{R}^t$. We have

$$\langle \mathbf{v}, \mathbf{H}\mathbf{v} \rangle = \sum_{i=1}^{t} c_i^2 \langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_i \rangle > 0.$$

By construction, at every iteration of MINRES, we have

$$\mathbf{p}_t = \underset{\mathbf{p}\in\mathcal{K}_t(\mathbf{H},\mathbf{g})}{\arg\min} \|\mathbf{H}\mathbf{p}+\mathbf{g}\|^2 \iff \mathbf{r}_t = \mathbf{H}\mathbf{p}_t + \mathbf{g} \perp \mathbf{H}\mathcal{K}_t$$

It is easy to show that $\mathbf{r}_i \in \mathcal{K}_t(\mathbf{H},\mathbf{g})$ for any $i \leq t-1$. This implies that

$$\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_j \rangle = 0, \quad i \neq j.$$

One can also show that as long as $\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_i \rangle \neq 0$, for all $0 \leq i \leq t-1$, then $\text{Span}\{\mathbf{r}_0, \mathbf{r}_1, \ldots, \mathbf{r}_{t-1}\} = \mathcal{K}_t(\mathbf{H},\mathbf{g})$. In particular, suppose $\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_i \rangle > 0$, for all $0 \leq i \leq t-1$ and let $\mathbf{0} \neq \mathbf{v} \in \mathcal{K}_t(\mathbf{H},\mathbf{g})$. We can write $\mathbf{v} = [\mathbf{r}_0 \mid \mathbf{r}_1 \mid \ldots \mid \mathbf{r}_{t-1}]\mathbf{c}$ for some $\mathbf{0} \neq \mathbf{c} \in \mathbb{R}^t$. We have

$$\langle \mathbf{v}, \mathbf{H}\mathbf{v} \rangle = \sum_{i=1}^{t} c_i^2 \langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_i \rangle > 0.$$

In other words, as long as $\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_i \rangle > 0$, for all $0 \leq i \leq t-1$, we have $\langle \mathbf{v}, \mathbf{H}\mathbf{v} \rangle > 0$ for any $\mathbf{0} \neq \mathbf{v} \in \mathcal{K}_t(\mathbf{H},\mathbf{g})$.

By construction, at every iteration of MINRES, we have

$$\mathbf{p}_t = \underset{\mathbf{p}\in\mathcal{K}_t(\mathbf{H},\mathbf{g})}{\arg\min} \left\|\mathbf{H}\mathbf{p} + \mathbf{g}\right\|^2 \iff \mathbf{r}_t = \mathbf{H}\mathbf{p}_t + \mathbf{g} \perp \mathbf{H}\mathcal{K}_t$$

It is easy to show that $\mathbf{r}_i \in \mathcal{K}_t(\mathbf{H},\mathbf{g})$ for any $i \leq t-1$. This implies that

$$\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_j \rangle = 0, \quad i \neq j.$$

One can also show that as long as $\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_i \rangle \neq 0$, for all $0 \leq i \leq t-1$, then $\text{Span}\{\mathbf{r}_0, \mathbf{r}_1, \ldots, \mathbf{r}_{t-1}\} = \mathcal{K}_t(\mathbf{H},\mathbf{g})$. In particular, suppose $\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_i \rangle > 0$, for all $0 \leq i \leq t-1$ and let $\mathbf{0} \neq \mathbf{v} \in \mathcal{K}_t(\mathbf{H},\mathbf{g})$. We can write $\mathbf{v} = [\mathbf{r}_0 \mid \mathbf{r}_1 \mid \ldots \mid \mathbf{r}_{t-1}]\mathbf{c}$ for some $\mathbf{0} \neq \mathbf{c} \in \mathbb{R}^t$. We have

$$\langle \mathbf{v}, \mathbf{H}\mathbf{v} \rangle = \sum_{i=1}^{t} c_i^2 \langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_i \rangle > 0.$$

In other words, as long as $\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_i \rangle > 0$, for all $0 \leq i \leq t-1$, we have $\langle \mathbf{v}, \mathbf{H}\mathbf{v} \rangle > 0$ for any $\mathbf{0} \neq \mathbf{v} \in \mathcal{K}_t(\mathbf{H},\mathbf{g})$. Conversely, $\langle \mathbf{r}_{t-1}, \mathbf{H}\mathbf{r}_{t-1} \rangle \leq 0$, then $\langle \mathbf{v}, \mathbf{H}\mathbf{v} \rangle > 0$ for some $\mathbf{0} \neq \mathbf{v} \in \mathcal{K}_t(\mathbf{H},\mathbf{g})$, namely $\mathbf{v} = \mathbf{r}_{t-1}$.

MINRES' NPC Condition

$$\langle \mathbf{r}_{t-1}, \mathbf{H}\mathbf{r}_{t-1} \rangle \le 0$$

MINRES' NPC Condition

$$\langle \mathbf{r}_{t-1}, \mathbf{H}\mathbf{r}_{t-1} \rangle \leq 0$$

The NPC condition can be readily checked as

$$\langle \mathbf{r}_{t-1}, \mathbf{H}\mathbf{r}_{t-1} \rangle = \spadesuit_{t-1} \times \clubsuit_{t}$$

Theorem (Liu and Roosta, 2022b)

Suppose $\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_i \rangle > 0, \quad 0 \leq i \leq t-1.$

Theorem (Liu and Roosta, 2022b)

Suppose $\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_i \rangle > 0, \quad 0 \leq i \leq t-1.$

- $\mathbf{T}_i \succ \mathbf{0}, \quad 1 \leq i \leq t$

Theorem (Liu and Roosta, 2022b)

Suppose $\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_i \rangle > 0, \quad 0 \leq i \leq t-1.$

- $\mathbf{T}_i \succ \mathbf{0}, \quad 1 \leq i \leq t$
- *If* $t = |\Lambda(\mathbf{H})|$, *then* $\mathbf{H} \succeq \mathbf{0}$ *(Pick the r.h.s vector uniformly at random from unit sphere)*

Theorem (Liu and Roosta, 2022b)

Suppose $\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_i \rangle > 0, \quad 0 \le i \le t-1.$

- $\mathbf{T}_i \succ \mathbf{0}, \quad 1 \le i \le t$
- *If* $t = |\Lambda(\mathbf{H})|$, *then* $\mathbf{H} \succeq \mathbf{0}$ *(Pick the r.h.s vector uniformly at random from unit sphere)*
- $\langle \mathbf{p}_i, \mathbf{g} \rangle < -\langle \mathbf{p}_i, \mathbf{H}\mathbf{p}_i \rangle < 0, \quad 1 \le i \le t$

Theorem (Liu and Roosta, 2022b)

Suppose $\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_i \rangle > 0, \quad 0 \le i \le t-1.$

- $\mathbf{T}_i \succ \mathbf{0}, \quad 1 \le i \le t$
- *If* $t = |\Lambda(\mathbf{H})|$, *then* $\mathbf{H} \succeq \mathbf{0}$ *(Pick the r.h.s vector uniformly at random from unit sphere)*
- $\langle \mathbf{p}_i, \mathbf{g} \rangle < -\langle \mathbf{p}_i, \mathbf{H}\mathbf{p}_i \rangle < 0, \quad 1 \le i \le t$
- $\langle \mathbf{g}, \mathbf{p}_i \rangle + \langle \mathbf{p}_i, \mathbf{H}\mathbf{p}_i \rangle / 2 \downarrow, \quad 0 \le i \le t$

Theorem (Liu and Roosta, 2022b)

Suppose $\langle \mathbf{r}_i, \mathbf{H}\mathbf{r}_i \rangle > 0, \quad 0 \le i \le t-1.$

- $\mathbf{T}_i \succ \mathbf{0}, \quad 1 \le i \le t$
- *If* $t = |\Lambda(\mathbf{H})|$, *then* $\mathbf{H} \succeq \mathbf{0}$ *(Pick the r.h.s vector uniformly at random from unit sphere)*
- $\langle \mathbf{p}_i, \mathbf{g} \rangle < -\langle \mathbf{p}_i, \mathbf{H}\mathbf{p}_i \rangle < 0, \quad 1 \le i \le t$
- $\langle \mathbf{g}, \mathbf{p}_i \rangle + \langle \mathbf{p}_i, \mathbf{H}\mathbf{p}_i \rangle / 2 \downarrow, \quad 0 \le i \le t$
- $\|\mathbf{p}_i\| \uparrow, \quad 0 \le i \le t$

	CG	MINRES

	CG	MINRES
NPC Condition	$\langle \mathbf{d}_t, \mathbf{H}\mathbf{d}_t \rangle \leq 0$	$\langle \mathbf{r}_{t-1}, \mathbf{H}\mathbf{r}_{t-1} \rangle \leq 0$

	CG	MINRES
NPC Condition	$\langle \mathbf{d}_t, \mathbf{H}\mathbf{d}_t \rangle \leq 0$	$\langle \mathbf{r}_{t-1}, \mathbf{H}\mathbf{r}_{t-1} \rangle \leq 0$
1st-order descent	$\langle \mathbf{p}_t, \mathbf{g} \rangle < 0$	$\langle \mathbf{p}_t, \mathbf{g} \rangle < 0$

	CG	MINRES
NPC Condition	$\langle \mathbf{d}_t, \mathbf{H}\mathbf{d}_t \rangle \leq 0$	$\langle \mathbf{r}_{t-1}, \mathbf{H}\mathbf{r}_{t-1} \rangle \leq 0$
1st-order descent	$\langle \mathbf{p}_t, \mathbf{g} \rangle < 0$	$\langle \mathbf{p}_t, \mathbf{g} \rangle < 0$
2nd-order descent	$\langle \mathbf{p}_t, \mathbf{g} \rangle + \frac{1}{2} \langle \mathbf{p}_t, \mathbf{H}\mathbf{p}_t \rangle < 0 \quad (\downarrow)$	$\langle \mathbf{p}_t, \mathbf{g} \rangle + \frac{1}{2} \langle \mathbf{p}_t, \mathbf{H}\mathbf{p}_t \rangle < 0 \quad (\downarrow)$

	CG	MINRES
NPC Condition	$\langle \mathbf{d}_t, \mathbf{H}\mathbf{d}_t \rangle \leq 0$	$\langle \mathbf{r}_{t-1}, \mathbf{H}\mathbf{r}_{t-1} \rangle \leq 0$
1st-order descent	$\langle \mathbf{p}_t, \mathbf{g} \rangle < 0$	$\langle \mathbf{p}_t, \mathbf{g} \rangle < 0$
2nd-order descent	$\langle \mathbf{p}_t, \mathbf{g} \rangle + \frac{1}{2} \langle \mathbf{p}_t, \mathbf{H}\mathbf{p}_t \rangle < 0 \quad (\downarrow)$	$\langle \mathbf{p}_t, \mathbf{g} \rangle + \frac{1}{2} \langle \mathbf{p}_t, \mathbf{H}\mathbf{p}_t \rangle < 0 \quad (\downarrow)$
"2.5th"-order descent	$\langle \mathbf{p}_t, \mathbf{g} \rangle + \langle \mathbf{p}_t, \mathbf{H}\mathbf{p}_t \rangle = 0$	$\langle \mathbf{p}_t, \mathbf{g} \rangle + \langle \mathbf{p}_t, \mathbf{H}\mathbf{p}_t \rangle < 0$

	CG	MINRES
NPC Condition	$\langle \mathbf{d}_t, \mathbf{H}\mathbf{d}_t \rangle \leq 0$	$\langle \mathbf{r}_{t-1}, \mathbf{H}\mathbf{r}_{t-1} \rangle \leq 0$
1st-order descent	$\langle \mathbf{p}_t, \mathbf{g} \rangle < 0$	$\langle \mathbf{p}_t, \mathbf{g} \rangle < 0$
2nd-order descent	$\langle \mathbf{p}_t, \mathbf{g} \rangle + \frac{1}{2} \langle \mathbf{p}_t, \mathbf{H}\mathbf{p}_t \rangle < 0 \quad (\downarrow)$	$\langle \mathbf{p}_t, \mathbf{g} \rangle + \frac{1}{2} \langle \mathbf{p}_t, \mathbf{H}\mathbf{p}_t \rangle < 0 \quad (\downarrow)$
"2.5th"-order descent	$\langle \mathbf{p}_t, \mathbf{g} \rangle + \langle \mathbf{p}_t, \mathbf{H}\mathbf{p}_t \rangle = 0$	$\langle \mathbf{p}_t, \mathbf{g} \rangle + \langle \mathbf{p}_t, \mathbf{H}\mathbf{p}_t \rangle < 0$
1st-order descent for $\\|\mathbf{g}\\|^2$	$\langle \mathbf{p}_t, \mathbf{H}\mathbf{g} \rangle = -\\|\mathbf{g}\\|^2 < 0$	$\langle \mathbf{p}_t, \mathbf{H}\mathbf{g} \rangle = -\\|\mathbf{H}\mathbf{p}_t\\|^2 < 0$

	CG	MINRES
NPC Condition	$\langle \mathbf{d}_t, \mathbf{H}\mathbf{d}_t \rangle \leq 0$	$\langle \mathbf{r}_{t-1}, \mathbf{H}\mathbf{r}_{t-1} \rangle \leq 0$
1st-order descent	$\langle \mathbf{p}_t, \mathbf{g} \rangle < 0$	$\langle \mathbf{p}_t, \mathbf{g} \rangle < 0$
2nd-order descent	$\langle \mathbf{p}_t, \mathbf{g} \rangle + \frac{1}{2} \langle \mathbf{p}_t, \mathbf{H}\mathbf{p}_t \rangle < 0 \quad (\downarrow)$	$\langle \mathbf{p}_t, \mathbf{g} \rangle + \frac{1}{2} \langle \mathbf{p}_t, \mathbf{H}\mathbf{p}_t \rangle < 0 \quad (\downarrow)$
"2.5th"-order descent	$\langle \mathbf{p}_t, \mathbf{g} \rangle + \langle \mathbf{p}_t, \mathbf{H}\mathbf{p}_t \rangle = 0$	$\langle \mathbf{p}_t, \mathbf{g} \rangle + \langle \mathbf{p}_t, \mathbf{H}\mathbf{p}_t \rangle < 0$
1st-order descent for $\lVert \mathbf{g} \rVert^2$	$\langle \mathbf{p}_t, \mathbf{H}\mathbf{g} \rangle = - \lVert \mathbf{g} \rVert^2 < 0$	$\langle \mathbf{p}_t, \mathbf{H}\mathbf{g} \rangle = - \lVert \mathbf{H}\mathbf{p}_t \rVert^2 < 0$
Norm of iterates	$\lVert \mathbf{p}_t \rVert \uparrow$	$\lVert \mathbf{p}_t \rVert \uparrow$

	CG	MINRES
NPC Condition	$\langle \mathbf{d}_t, \mathbf{H}\mathbf{d}_t \rangle \leq 0$	$\langle \mathbf{r}_{t-1}, \mathbf{H}\mathbf{r}_{t-1} \rangle \leq 0$
1st-order descent	$\langle \mathbf{p}_t, \mathbf{g} \rangle < 0$	$\langle \mathbf{p}_t, \mathbf{g} \rangle < 0$
2nd-order descent	$\langle \mathbf{p}_t, \mathbf{g} \rangle + \frac{1}{2} \langle \mathbf{p}_t, \mathbf{H}\mathbf{p}_t \rangle < 0 \quad (\downarrow)$	$\langle \mathbf{p}_t, \mathbf{g} \rangle + \frac{1}{2} \langle \mathbf{p}_t, \mathbf{H}\mathbf{p}_t \rangle < 0 \quad (\downarrow)$
"2.5th"-order descent	$\langle \mathbf{p}_t, \mathbf{g} \rangle + \langle \mathbf{p}_t, \mathbf{H}\mathbf{p}_t \rangle = 0$	$\langle \mathbf{p}_t, \mathbf{g} \rangle + \langle \mathbf{p}_t, \mathbf{H}\mathbf{p}_t \rangle < 0$
1st-order descent for $\|\mathbf{g}\|^2$	$\langle \mathbf{p}_t, \mathbf{H}\mathbf{g} \rangle = -\|\mathbf{g}\|^2 < 0$	$\langle \mathbf{p}_t, \mathbf{H}\mathbf{g} \rangle = -\|\mathbf{H}\mathbf{p}_t\|^2 < 0$
Norm of iterates	$\|\mathbf{p}_t\| \uparrow$	$\|\mathbf{p}_t\| \uparrow$
1st-order descent (NPC)	$\langle \mathbf{d}_t, \mathbf{g} \rangle = -\|\mathbf{r}_t\|^2 < 0$	$\langle \mathbf{r}_{t-1}, \mathbf{g} \rangle = -\|\mathbf{r}_{t-1}\|^2 < 0$

	CG	MINRES
NPC Condition	$\langle \mathbf{d}_t, \mathbf{H}\mathbf{d}_t \rangle \leq 0$	$\langle \mathbf{r}_{t-1}, \mathbf{H}\mathbf{r}_{t-1} \rangle \leq 0$
1st-order descent	$\langle \mathbf{p}_t, \mathbf{g} \rangle < 0$	$\langle \mathbf{p}_t, \mathbf{g} \rangle < 0$
2nd-order descent	$\langle \mathbf{p}_t, \mathbf{g} \rangle + \frac{1}{2} \langle \mathbf{p}_t, \mathbf{H}\mathbf{p}_t \rangle < 0 \quad (\downarrow)$	$\langle \mathbf{p}_t, \mathbf{g} \rangle + \frac{1}{2} \langle \mathbf{p}_t, \mathbf{H}\mathbf{p}_t \rangle < 0 \quad (\downarrow)$
"2.5th"-order descent	$\langle \mathbf{p}_t, \mathbf{g} \rangle + \langle \mathbf{p}_t, \mathbf{H}\mathbf{p}_t \rangle = 0$	$\langle \mathbf{p}_t, \mathbf{g} \rangle + \langle \mathbf{p}_t, \mathbf{H}\mathbf{p}_t \rangle < 0$
1st-order descent for $\lVert \mathbf{g} \rVert^2$	$\langle \mathbf{p}_t, \mathbf{H}\mathbf{g} \rangle = - \lVert \mathbf{g} \rVert^2 < 0$	$\langle \mathbf{p}_t, \mathbf{H}\mathbf{g} \rangle = - \lVert \mathbf{H}\mathbf{p}_t \rVert^2 < 0$
Norm of iterates	$\lVert \mathbf{p}_t \rVert \uparrow$	$\lVert \mathbf{p}_t \rVert \uparrow$
1st-order descent (NPC)	$\langle \mathbf{d}_t, \mathbf{g} \rangle = - \lVert \mathbf{r}_t \rVert^2 < 0$	$\langle \mathbf{r}_{t-1}, \mathbf{g} \rangle = - \lVert \mathbf{r}_{t-1} \rVert^2 < 0$
1st-order non-ascent for $\lVert \mathbf{g} \rVert^2$ (NPC)	$\langle \mathbf{d}_t, \mathbf{H}\mathbf{g} \rangle = 0$	$\langle \mathbf{r}_{t-1}, \mathbf{H}\mathbf{g} \rangle = 0$

	CG	MINRES
Simplicity	✓	✗
Coverage in Textbook	✓	✗
Software Libraries	✓	✗
Theoretical Properties	✓	?
Numerical Properties	✓	✗

	CG	MINRES
Simplicity	✓	✗
Coverage in Textbook	✓	✗
Software Libraries	✓	✗
Theoretical Properties	✓	✓
Numerical Properties	✓	✗

MINRES is a much more general purpose than CG, and hence it is a more complicated.

MINRES is a much more general purpose than CG, and hence it is a more complicated. It turns out there is a much simpler alternative algorithm that for all intents and purposes as an optimization sub-problem solver is “identical” to MINRES.

MINRES is a much more general purpose than CG, and hence it is a more complicated. It turns out there is a much simpler alternative algorithm that for all intents and purposes as an optimization sub-problem solver is "identical" to MINRES. Recall CG iterations are formed by solving:

$$\mathbf{p}_t = \underset{\mathbf{p}\in\mathcal{K}_t}{\arg\min}\ \frac{1}{2}\langle \mathbf{p}, \mathbf{H}\mathbf{p}\rangle + \langle \mathbf{p}, \mathbf{g}\rangle .$$

MINRES is a much more general purpose than CG, and hence it is a more complicated. It turns out there is a much simpler alternative algorithm that for all intents and purposes as an optimization sub-problem solver is “identical” to MINRES. Recall CG iterations are formed by solving:

$$\mathbf{p}_t = \underset{\mathbf{p} \in \mathcal{K}_t}{\arg\min} \ \frac{1}{2} \langle \mathbf{p}, \mathbf{H}\mathbf{p} \rangle + \langle \mathbf{p}, \mathbf{g} \rangle .$$

If $\mathbf{H} \succ \mathbf{0}$, we have we can replace the Euclidean inner product by an $\mathbf{H}$-inner product and get

$$\mathbf{p}_t = \underset{\mathbf{p} \in \mathcal{K}_t}{\arg\min} \ \frac{1}{2} \langle \mathbf{p}, \mathbf{H}\mathbf{p} \rangle_{\mathbf{H}} + \langle \mathbf{p}, \mathbf{g} \rangle_{\mathbf{H}} = \underset{\mathbf{p} \in \mathcal{K}_t}{\arg\min} \ \frac{1}{2} \left\| \mathbf{H}\mathbf{p} + \mathbf{g} \right\|^2 .$$

Algorithm Conjugate Residual[1]

$\mathbf{r}_0 = \mathbf{d}_1 = -\mathbf{g}$, and $\mathbf{p}_0 = \mathbf{0}$
for $t = 1, 2, \ldots$ until $\|\mathbf{r}_{t-1}\| \leq \tau$ **do**
    $\alpha_t = \langle \mathbf{r}_{t-1}, \mathbf{r}_{t-1} \rangle_{\mathbf{H}} \,/\, \langle \mathbf{d}_t, \mathbf{H}\mathbf{d}_t \rangle_{\mathbf{H}}$
    $\mathbf{p}_t = \mathbf{p}_{t-1} + \alpha_t \mathbf{d}_t$
    $\mathbf{r}_t = \mathbf{r}_{t-1} - \alpha_t \mathbf{H}\mathbf{d}_t$
    $\beta_{t+1} = \langle \mathbf{r}_t, \mathbf{r}_t \rangle_{\mathbf{H}} \,/\, \langle \mathbf{r}_{t-1}, \mathbf{r}_{t-1} \rangle_{\mathbf{H}}$
    $\mathbf{d}_{t+1} = \mathbf{r}_t + \beta_{t+1} \mathbf{d}_t$
end for

[1] CR can be implemented to have one matrix-vector product per iteration, in which case it requires one more vector of storage and one more vector update than the CG.

Algorithm Conjugate Residual[1]

$\mathbf{r}_0 = \mathbf{d}_1 = -\mathbf{g}$, and $\mathbf{p}_0 = \mathbf{0}$
for $t = 1, 2, \ldots$ until $\|\mathbf{r}_{t-1}\| \leq \tau$ **do**
$\quad \alpha_t = \langle \mathbf{r}_{t-1}, \mathbf{H}\mathbf{r}_{t-1} \rangle \,/\, \|\mathbf{H}\mathbf{d}_t\|^2$
$\quad \mathbf{p}_t = \mathbf{p}_{t-1} + \alpha_t \mathbf{d}_t$
$\quad \mathbf{r}_t = \mathbf{r}_{t-1} - \alpha_t \mathbf{H}\mathbf{d}_t$
$\quad \beta_{t+1} = \langle \mathbf{r}_t, \mathbf{H}\mathbf{r}_t \rangle \,/\, \langle \mathbf{r}_{t-1}, \mathbf{H}\mathbf{r}_{t-1} \rangle$
$\quad \mathbf{d}_{t+1} = \mathbf{r}_t + \beta_{t+1} \mathbf{d}_t$
end for

[1] CR can be implemented to have one matrix-vector product per iteration, in which case it requires one more vector of storage and one more vector update than the CG.

Algorithm Conjugate Residual[1]

$\mathbf{r}_0 = \mathbf{d}_1 = -\mathbf{g}$, and $\mathbf{p}_0 = \mathbf{0}$
for $t = 1, 2, \ldots$ until $\|\mathbf{r}_{t-1}\| \leq \tau$ **do**
  $\alpha_t = \langle \mathbf{r}_{t-1}, \mathbf{H}\mathbf{r}_{t-1} \rangle \,/\, \|\mathbf{H}\mathbf{d}_t\|^2$
  $\mathbf{p}_t = \mathbf{p}_{t-1} + \alpha_t \mathbf{d}_t$
  $\mathbf{r}_t = \mathbf{r}_{t-1} - \alpha_t \mathbf{H}\mathbf{d}_t$
  $\beta_{t+1} = \langle \mathbf{r}_t, \mathbf{H}\mathbf{r}_t \rangle \,/\, \langle \mathbf{r}_{t-1}, \mathbf{H}\mathbf{r}_{t-1} \rangle$
  $\mathbf{d}_{t+1} = \mathbf{r}_t + \beta_{t+1} \mathbf{d}_t$
end for

MINRES is also simple!

Theorem (Lim, Liu, and Roosta, 2024)

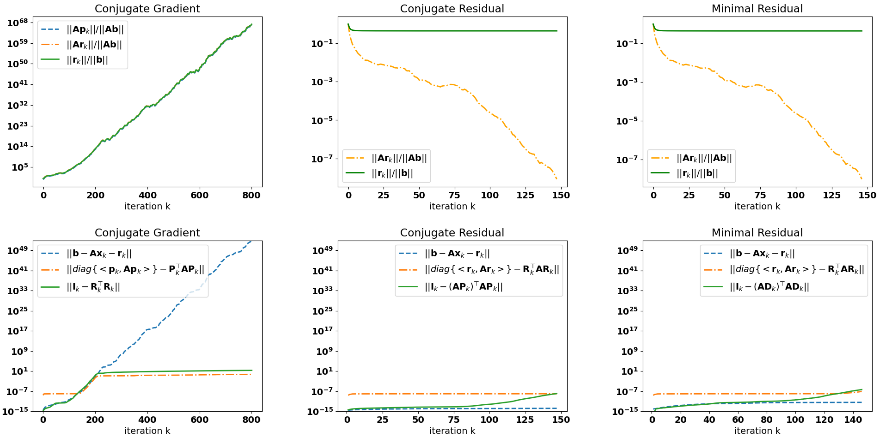
MINRES and CR are essentially the same for all $\mathbf{H}$ *(not just PD)!*

[1] CR can be implemented to have one matrix-vector product per iteration, in which case it requires one more vector of storage and one more vector update than the CG.

	CG	MINRES/CR
Simplicity	✓	✗
Coverage in Textbook	✓	✗
Software Libraries	✓	✗
Theoretical Properties	✓	✓
Numerical Properties	✓	✗

	CG	MINRES/CR
Simplicity	✓	✓
Coverage in Textbook	✓	✗
Software Libraries	✓	✗
Theoretical Properties	✓	✓
Numerical Properties	✓	✗

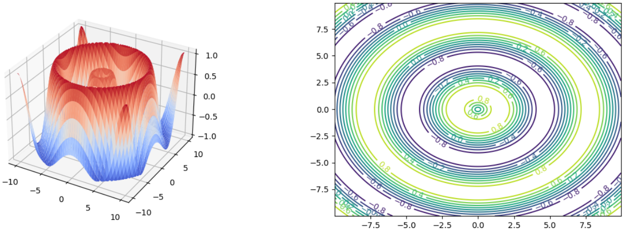
CG is unstable!



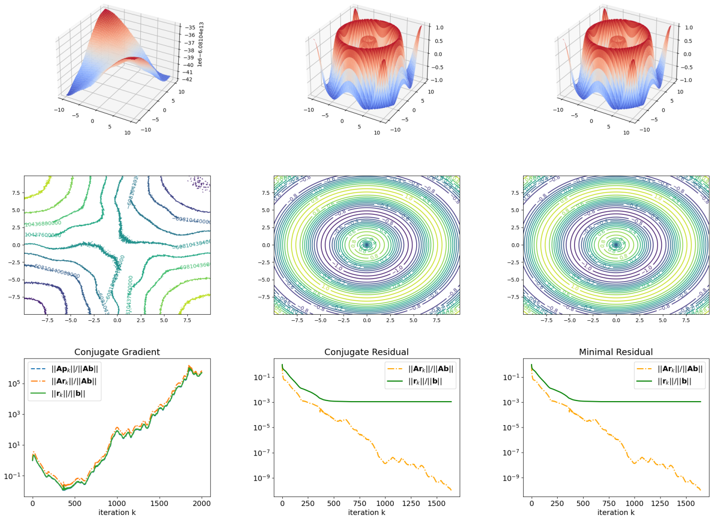
CG CR MINRES

(Lim, Liu, and Roosta, 2024)

CG is unstable



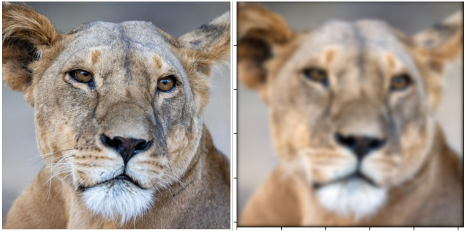
CG is unstable



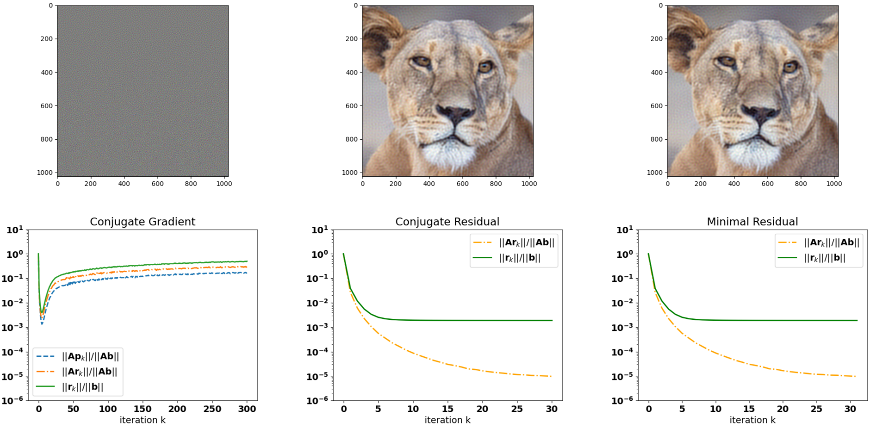
CG CR MINRES

(Lim, Liu, and Roosta, 2024)

CG is unstable



CG is unstable and/or its solutions can be useless!



CG CR MR

(Lim, Liu, and Roosta, 2024)

	CG	MINRES/CR
Simplicity	✓	✓
Coverage in Textbook	✓	✗
Software Libraries	✓	✗
Theoretical Properties	✓	✓
Numerical Properties		

	CG	MINRES/CR
Simplicity	✓	✓
Coverage in Textbook	✓	✗
Software Libraries	✓	✗
Theoretical Properties	✓	✓
Numerical Properties	✓	✗

CG Has No Intuitive Termination Condition

CG Has No Intuitive Termination Condition

$$\|\mathbf{r}\| \leq \eta$$

CG Has No Intuitive Termination Condition

$$\left\| \left(\mathbf{I} - \mathbf{H}\mathbf{H}^{\dagger} \right) \mathbf{g} \right\| \leq \ \|\mathbf{r}\| \leq \eta$$

CG Has No Intuitive Termination Condition

$$\|\mathbf{Hr}\| \leq \eta$$

CG Has No Intuitive Termination Condition

$$\|\mathbf{Hr}\| \leq \eta$$

Theorem (Lim, Liu, and Roosta, 2024)

With CG, when $\mathbf{g} \notin \text{Range}(\mathbf{H})$,

CG Has No Intuitive Termination Condition

$$\|\mathbf{Hr}\| \leq \eta$$

Theorem (Lim, Liu, and Roosta, 2024)

With CG, when $\mathbf{g} \notin \text{Range}(\mathbf{H})$, $\mathbf{p} \neq -\mathbf{H}^{\dagger}\mathbf{g} + (\mathbf{I} - \mathbf{H}^{\dagger}\mathbf{H})\mathbf{q}$, *for any* $\mathbf{q} \in \mathbb{R}^{d}$.

CG Has No Intuitive Termination Condition

$$\|\mathbf{Hr}\| \leq \eta$$

Theorem (Lim, Liu, and Roosta, 2024)

With CG, when $\mathbf{g} \notin \text{Range}(\mathbf{H})$, $\mathbf{p} \neq -\mathbf{H}^{\dagger}\mathbf{g} + (\mathbf{I} - \mathbf{H}^{\dagger}\mathbf{H})\mathbf{q}$, *for any* $\mathbf{q} \in \mathbb{R}^{d}$.

With MINRES, always,

CG Has No Intuitive Termination Condition

$$\|\mathbf{Hr}\| \leq \eta$$

Theorem (Lim, Liu, and Roosta, 2024)

With CG, when $\mathbf{g} \notin \text{Range}(\mathbf{H})$, $\mathbf{p} \neq -\mathbf{H}^{\dagger}\mathbf{g} + (\mathbf{I} - \mathbf{H}^{\dagger}\mathbf{H})\mathbf{q}$, *for any* $\mathbf{q} \in \mathbb{R}^d$.

With MINRES, always, $\mathbf{p} = -\mathbf{H}^{\dagger}\mathbf{g} + (\mathbf{I} - \mathbf{H}^{\dagger}\mathbf{H})\mathbf{q}$, for some $\mathbf{q} \in \mathbb{R}^d$.

	CG	MINRES/CR
Simplicity	✓	✓
Coverage in Textbook	✓	✗
Software Libraries	✓	✗
Theoretical Properties	✓	✓
Numerical Properties	✗	✓
Natural Termination	✗	✓

	CG	MINRES/CR
Simplicity	✓	✓
Coverage in Textbook	✓	✗
Software Libraries	✓	✗
Theoretical Properties	✓	✓
Numerical Properties	✗	✓
Natural Termination	✗	✓

Fong, David Chin-Lung and Michael Saunders (2012). "CG versus MINRES: An empirical comparison". In: *Sultan Qaboos University Journal for Science [SQUJS]* 17.1, pp. 44–62.

Liu, Yang and Fred Roosta (2022a). "A Newton-MR algorithm with complexity guarantees for nonconvex smooth unconstrained optimization". In: *arXiv preprint arXiv:2208.07095*.

— (2022b). "MINRES: From Negative Curvature Detection to Monotonicity Properties". In: *SIAM Journal on Optimization* 32.4, pp. 2636–2661.

Lim, Alexander, Yang Liu, and Fred Roosta (2024). "Conjugate Direction Methods Under Inconsistent Systems". In: *arXiv preprint arXiv:2401.11714*.