Linearizing the Method of Conjugate Gradients

David Titley-Peloquin
Postdoctoral Fellow
Fondation STAE — IRIT-ENSEEIHT

Joint work with S. Gratton, P. Toint, and J. Tshimanga

Journée des Doctorants 2012

Research funded by

Fondation Sciences et Technologies pour l'Aéronautique et l'Espace

Projet ADTAO

www.fondation-stae.net

and

Natural Sciences and Engineering Research Council of Canada

www.nserc.ca

The method of Conjugate Gradients (CG) is widely used for solving large sparse systems of equations:

$$Ax = b, \qquad A \in \mathbb{R}^{n \times n} \text{ symmetric positive definite}$$

Starting from a given x_0, CG produces iterates $x_k \in \mathbb{R}^n, \quad k = 1, 2, \ldots$ designed to converge to $x^* = A^{-1}b$.

The method of Conjugate Gradients (CG) is widely used for solving large sparse systems of equations:

$$Ax = b, \qquad A \in \mathbb{R}^{n\times n} \text{ symmetric positive definite}$$

Starting from a given x_0, CG produces iterates $x_k \in \mathbb{R}^n, \quad k = 1, 2, \ldots$ designed to converge to $x^* = A^{-1}b$.

For fixed A, x_0, k,

$$x_k(b) : \mathbb{R}^n \to \mathbb{R}^n$$

is a nonlinear differentiable function of b.

In this talk:

$$J_k = \frac{\partial x_k}{\partial b} \in \mathbb{R}^{n\times n}$$

Outline

- Motivation / Possible Applications
- The Jacobian and 2-norm Condition Number of x_k
- Computing Matvecs with J_k and J_k^T
- Summary and Future Research Directions

Outline

- Motivation / Possible Applications
- The Jacobian and 2-norm Condition Number of x_k
- Computing Matvecs with J_k and J_k^T
- Summary and Future Research Directions

(A) Truncated CG Regression

- consider the statistical linear model

$$c = Kx + v, \qquad \text{where} \quad v \sim \mathcal{N}(0, \Sigma)$$

(A) Truncated CG Regression

- consider the statistical linear model

$$c = Kx + v, \qquad \text{where} \quad v \sim \mathcal{N}(0, \Sigma)$$

- the BLUE of x is

$$x^* = \arg\min_x \|c - Kx\|_{\Sigma^{-1}} = (K^T \Sigma^{-1} K)^{-1} K^T \Sigma^{-1} c$$

(A) Truncated CG Regression

- consider the statistical linear model

$$c = Kx + v, \qquad \text{where} \quad v \sim \mathcal{N}(0, \Sigma)$$

- the BLUE of x is

$$x^* = \arg\min_x \|c - Kx\|_{\Sigma^{-1}} = (K^T\Sigma^{-1}K)^{-1}K^T\Sigma^{-1}c$$

and $\qquad \operatorname{cov}\{x^* - x\} = (K^T\Sigma^{-1}K)^{-1}$

(A) Truncated CG Regression

- consider the statistical linear model

$$c = Kx + v, \qquad \text{where} \quad v \sim \mathcal{N}(0, \Sigma)$$

- the BLUE of x is

$$x^* = \arg\min_x \|c - Kx\|_{\Sigma^{-1}} = (K^T\Sigma^{-1}K)^{-1}K^T\Sigma^{-1}c$$

and $$\text{cov}\{x^* - x\} = (K^T\Sigma^{-1}K)^{-1}$$

- but in many applications, we compute x_k for $k \ll n$

What is $\text{cov}\{x_k - x\}$?

(A) Truncated CG Regression

- consider the statistical linear model

$$c = Kx + v, \qquad \text{where} \quad v \sim \mathcal{N}(0, \Sigma)$$

(A) Truncated CG Regression

- consider the statistical linear model

$$c = Kx + v, \qquad \text{where} \quad v \sim \mathcal{N}(0, \Sigma)$$

- taking a Taylor expansion about c,

$$x_k(c+v) \approx x_k(c) + J_k(c) \cdot v$$
$$x_k(c+v) - x \approx x_k(c) - x + J_k(c) \cdot v$$

(A) Truncated CG Regression

- consider the statistical linear model

$$c = Kx + v, \qquad \text{where} \quad v \sim \mathcal{N}(0, \Sigma)$$

- taking a Taylor expansion about c,

$$x_k(c+v) \approx x_k(c) + J_k(c) \cdot v$$
$$x_k(c+v) - x \approx x_k(c) - x + J_k(c) \cdot v$$

- to first order,

$$\text{cov}\{x_k - x\} \approx \text{cov}\{J_k v\} = J_k \Sigma J_k^T$$

(A) Truncated CG Regression

- consider the statistical linear model

$$c = Kx + v, \qquad \text{where} \quad v \sim \mathcal{N}(0, \Sigma)$$

- taking a Taylor expansion about c,

$$x_k(c+v) \approx x_k(c) + J_k(c) \cdot v$$
$$x_k(c+v) - x \approx x_k(c) - x + J_k(c) \cdot v$$

- to first order,

$$\text{cov}\{x_k - x\} \approx \text{cov}\{J_k v\} = J_k \Sigma J_k^T$$

- important in data assimilation applications
  e.g. [Fischer and Courtier (1995), Moore et. al. (2012)]

(B) Sensitivity Analysis

- in itself considered by many to be a very important topic (see e.g. reference books of [Demmel, Golub and Van Loan, Higham])

(B) Sensitivity Analysis

- in itself considered by many to be a very important topic (see e.g. reference books of [Demmel, Golub and Van Loan, Higham])
- e.g. for the exact solution $x^* = A^{-1}b$,

$$J^* = \frac{\partial x^*}{\partial b} = A^{-1}, \qquad \|J^*\|_2 = \|A^{-1}\|_2 = \frac{1}{\lambda_{\min}(A)}$$

shows how perturbations in b affect the true solution x^*

(B) Sensitivity Analysis

- in itself considered by many to be a very important topic (see e.g. reference books of [Demmel, Golub and Van Loan, Higham])
- e.g. for the exact solution $x^* = A^{-1}b$,

$$J^* = \frac{\partial x^*}{\partial b} = A^{-1}, \qquad \|J^*\|_2 = \|A^{-1}\|_2 = \frac{1}{\lambda_{\text{min}}(A)}$$

shows how perturbations in b affect the true solution x^*

- but how robust is x_k $(k \ll n)$ to perturbations in b ?

$$\lim_{\epsilon \to 0} \sup_{\|\delta b\|_2 \leq \epsilon} \frac{\|x_k(b + \delta b) - x_k(b)\|_2}{\|\delta b\|_2} = \|J_k\|_2$$

(C) Other Potential Applications

- solve for several RHS vectors

$$Ax^{(i)} = b^{(i)} \qquad i = 1, 2, \ldots$$

- if we have computed $x_k(b^{(i)})$ and $J_k(b^{(i)})$ then to first order

$$x_k(b^{(j)}) \approx x_k(b^{(i)}) + J_k(b^{(i)}) \cdot (b^{(j)} - b^{(i)})$$

(C) Other Potential Applications

- solve for several RHS vectors

$$Ax^{(i)} = b^{(i)} \qquad i = 1, 2, \ldots$$

- if we have computed $x_k(b^{(i)})$ and $J_k(b^{(i)})$ then to first order

$$x_k(b^{(j)}) \approx x_k(b^{(i)}) + J_k(b^{(i)}) \cdot (b^{(j)} - b^{(i)})$$

- if $J_k \approx A^{-1}$
  - use it to precondition subsequent solves ?
  - use it to estimate the energy norm of the error ?

$$\|\epsilon_k\|_A^2 = \epsilon_k^T A \epsilon_k = r_k^T A^{-1} r_k \approx r_k^T J_k r_k$$

Outline

- Motivation / Possible Applications
- The Jacobian and 2-norm Condition Number of x_k
- Computing Matvecs with J_k and J_k^T
- Summary and Future Research Directions

Lanczos Algorithm

Starting from normalized $v_1 \in \mathbb{R}^n$, gives: $AV_k = V_k T_k + \beta_{k+1} v_{k+1} e_k^T$
where

$$V_k = [v_1, \ldots, v_k] \in \mathbb{R}^{n \times k}, \qquad V_k^T V_k = I_k, \qquad T_k \ \text{tridiagonal}$$

Lanczos Algorithm

Starting from normalized $v_1 \in \mathbb{R}^n$, gives: $AV_k = V_k T_k + \beta_{k+1} v_{k+1} e_k^T$
where

$$V_k = [v_1, \ldots, v_k] \in \mathbb{R}^{n \times k}, \qquad V_k^T V_k = I_k, \qquad T_k \text{ tridiagonal}$$

Relation to CG

$$x_k = x_0 + V_k y_k, \quad y_k = \|r_0\|_2 T_k^{-1} e_1$$
$$x_k - x_0 \in \text{range}(V_k) = \text{span}\{r_0, Ar_0, \ldots, A^{k-1} r_0\}$$

Lanczos Algorithm

Starting from normalized $v_1 \in \mathbb{R}^n$, gives: $AV_k = V_k T_k + \beta_{k+1} v_{k+1} e_k^T$
where

$$V_k = [v_1, \ldots, v_k] \in \mathbb{R}^{n \times k}, \qquad V_k^T V_k = I_k, \qquad T_k \ \text{tridiagonal}$$

Relation to CG

$$x_k = x_0 + V_k y_k, \quad y_k = \|r_0\|_2 T_k^{-1} e_1$$
$$x_k - x_0 \in \text{range}(V_k) = \text{span}\{r_0, Ar_0, \ldots, A^{k-1} r_0\}$$

Polynomials

$$x_k - x_0 = \zeta_{k-1}(A) r_0, \quad \epsilon_k = \rho_k(A)\epsilon_0, \quad \|\epsilon_k\|_A = \min_{\substack{\rho \in \Pi_k \\ \rho(0)=1}} \|\rho(A)\epsilon_0\|_A$$

Ingredients

- the Lanczos relation: $AV_k = V_k T_k + \beta_{k+1} v_{k+1} e_k^T$
- the polynomials:

$$x_k - x_0 = \zeta_{k-1}(A) r_0, \quad \epsilon_k = \rho_k(A)\epsilon_0, \quad \|\epsilon_k\|_A = \min_{\substack{\rho \in \Pi_k \\ \rho(0)=1}} \|\rho(A)\epsilon_0\|_A$$

Ingredients

- the Lanczos relation: $AV_k = V_k T_k + \beta_{k+1} v_{k+1} e_k^T$
- the polynomials:

$$x_k - x_0 = \zeta_{k-1}(A) r_0, \quad \epsilon_k = \rho_k(A)\epsilon_0, \quad \|\epsilon_k\|_A = \min_{\substack{\rho \in \Pi_k \\ \rho(0)=1}} \|\rho(A)\epsilon_0\|_A$$

Theorem [Gratton, T.-P., Toint, and Tshimanga (submitted, 2012)]

$$J_k = A^{-1}\left[I - \rho_k(A)\right] + 2 V_k T_k^{-1} V_k^T \rho_k(A)$$

Ingredients

- the Lanczos relation: $AV_k = V_k T_k + \beta_{k+1} v_{k+1} e_k^T$
- the polynomials:

$$x_k - x_0 = \zeta_{k-1}(A) r_0, \quad \epsilon_k = \rho_k(A)\epsilon_0, \quad \|\epsilon_k\|_A = \min_{\substack{\rho\in\Pi_k \\ \rho(0)=1}} \|\rho(A)\epsilon_0\|_A$$

Theorem [Gratton, T.-P., Toint, and Tshimanga (submitted, 2012)]

$$J_k = A^{-1}\left[I - \rho_k(A)\right] + 2V_k T_k^{-1} V_k^T \rho_k(A)$$

Equivalently,

$$J_k = 2V_k T_k^{-1} V_k^T + \left(I - 2V_k T_k^{-1} V_k^T A\right)\zeta_{k-1}(A)$$

$$J_k = A^{-1}\left[I - \rho_k(A)\right] + 2V_k T_k^{-1} V_k^T \rho_k(A)$$

Sketch of Proof

$$x_k \;=\; A^{-1}b - \epsilon_k \;=\; A^{-1}b - \rho_k(A)\epsilon_0 \;=\; A^{-1}b - \rho_k(A)\big(A^{-1}b - x_0\big)$$

$$J_k = A^{-1}\left[I - \rho_k(A)\right] + 2V_k T_k^{-1} V_k^T \rho_k(A)$$

Sketch of Proof

$$x_k \;=\; A^{-1}b - \epsilon_k \;=\; A^{-1}b - \rho_k(A)\epsilon_0 \;=\; A^{-1}b - \rho_k(A)\left(A^{-1}b - x_0\right)$$

$$\partial x_k = A^{-1}\left[I - \rho_k(A)\right] - \partial\rho_k(A)\epsilon_0$$

$$J_k = A^{-1}\left[I - \rho_k(A)\right] + 2V_k T_k^{-1} V_k^T \rho_k(A)$$

Sketch of Proof

$$x_k = A^{-1}b - \epsilon_k = A^{-1}b - \rho_k(A)\epsilon_0 = A^{-1}b - \rho_k(A)\left(A^{-1}b - x_0\right)$$

$$\partial x_k = A^{-1}\left[I - \rho_k(A)\right] - \partial\rho_k(A)\epsilon_0$$

Differentiating ρ_k

$$\rho_k(A) = I + \sum \tau_i A^i, \qquad t_k = \left[\tau_1, \ldots, \tau_k\right]^T$$

$$\partial\rho_k(A)\epsilon_0 = \sum \partial\tau_i A^i \epsilon_0 = \left[A\epsilon_0, \ldots, A^k\epsilon_0\right] \partial t_k$$

$$J_k = A^{-1}\left[I - \rho_k(A)\right] + 2V_k T_k^{-1} V_k^T \rho_k(A)$$

Sketch of Proof

$$x_k = A^{-1}b - \epsilon_k = A^{-1}b - \rho_k(A)\epsilon_0 = A^{-1}b - \rho_k(A)\left(A^{-1}b - x_0\right)$$

$$\partial x_k = A^{-1}\left[I - \rho_k(A)\right] - \partial\rho_k(A)\epsilon_0$$

Differentiating ρ_k

$$\rho_k(A) = I + \sum \tau_i A^i, \qquad t_k = \left[\tau_1, \ldots, \tau_k\right]^T$$

$$\partial\rho_k(A)\epsilon_0 = \sum \partial\tau_i A^i \epsilon_0 = \left[A\epsilon_0, \ldots, A^k\epsilon_0\right] \partial t_k$$

Coefficients

From the optimality of $\|\epsilon_k\|_A$ we obtain $t_k = -(L_k^T W^2 L_k)^{-1} L_k^T W^2 e$

where L_k is Vandermonde, W diagonal, and only W depends on b.

Some Consequenses of the Theorem

How close is J_k to A^{-1} ?

Some Consequenses of the Theorem

How close is J_k to A^{-1} ?

$$\frac{\|J_k - A^{-1}\|_2}{\|A^{-1}\|_2} \leq 3\|\rho_k(A)\|_2$$

Some Consequenses of the Theorem

How close is J_k to A^{-1} ?

$$\frac{\|J_k - A^{-1}\|_2}{\|A^{-1}\|_2} \leq 3\|\rho_k(A)\|_2$$

The roots of ρ_k (Ritz values) converge to some eigenvalues of A.
Let

$$A = Q\Lambda Q^T = Q \begin{bmatrix} \Lambda_1 & \\ & \Lambda_2 \end{bmatrix} Q^T, \qquad \rho_k(\Lambda_1) = 0$$

When CG converges, i.e. when $x_k = A^{-1}b$,

$$J_k = A^{-1} - Q \begin{bmatrix} 0 & \\ & \Lambda_2^{-1}\rho_k(\Lambda_2) \end{bmatrix} Q^T$$

Some Consequenses of the Theorem

How close is J_k to $V_k T_k^{-1} V_k^T$?

Some Consequenses of the Theorem

How close is J_k to $V_k T_k^{-1} V_k^T$?

$$J_k = \zeta_{k-1}(A) + 2 V_k T_k^{-1} V_k^T \rho_k(A)$$

Some Consequenses of the Theorem

How close is J_k to $V_k T_k^{-1} V_k^T$?

$$\begin{aligned} J_k &= \zeta_{k-1}(A) + 2 V_k T_k^{-1} V_k^T \rho_k(A) \\ &\approx \zeta_{k-1}(V_k T_k V_k^T) + 2 V_k T_k^{-1} V_k^T \rho_k(V_k T_k V_k^T) \\ &= \ldots \\ &= V_k T_k^{-1} V_k^T \end{aligned}$$

Some Consequenses of the Theorem

How close is J_k to $V_k T_k^{-1} V_k^T$?

$$
\begin{aligned}
J_k &= \zeta_{k-1}(A) + 2 V_k T_k^{-1} V_k^T \rho_k(A) \\
&\approx \zeta_{k-1}(V_k T_k V_k^T) + 2 V_k T_k^{-1} V_k^T \rho_k(V_k T_k V_k^T) \\
&= \ldots \\
&= V_k T_k^{-1} V_k^T
\end{aligned}
$$

Normwise:

$$
\| T_k^{-1} e_1 \|_2 \;\leq\; \| J_k \|_2 \;\leq\; 2 \| T_k^{-1} \|_2 + \left(1 + 2 \| A \|_2 \| T_k^{-1} \|_2\right) \| \zeta_{k-1}(A) \|_2
$$

Some Consequenses of the Theorem

How close is J_k to $V_k T_k^{-1} V_k^T$?

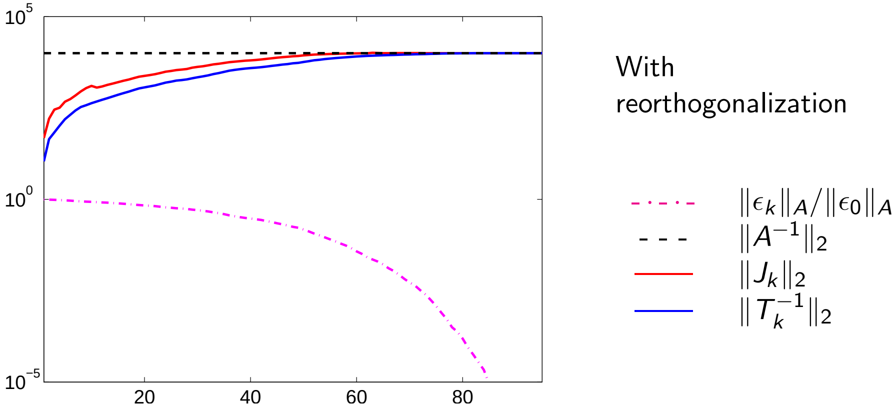
$$\begin{aligned} J_k &= \zeta_{k-1}(A) + 2 V_k T_k^{-1} V_k^T \rho_k(A) \\ &\approx \zeta_{k-1}(V_k T_k V_k^T) + 2 V_k T_k^{-1} V_k^T \rho_k(V_k T_k V_k^T) \\ &= \dots \\ &= V_k T_k^{-1} V_k^T \end{aligned}$$

Normwise:

$$\|T_k^{-1} e_1\|_2 \leq \|J_k\|_2 \leq 2\|T_k^{-1}\|_2 + \left(1 + 2\|A\|_2 \|T_k^{-1}\|_2\right) \|\zeta_{k-1}(A)\|_2$$

Example 1

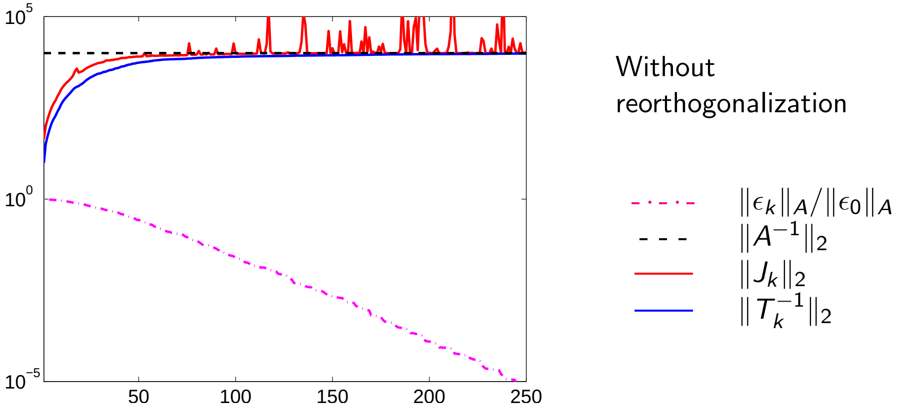
$$A = Q\Lambda Q^T, \qquad \Lambda = \text{logspace}(-4, 0, n)$$
$$b = \text{randn}(n)$$
$$x_0 = 0$$



With reorthogonalization

Example 1

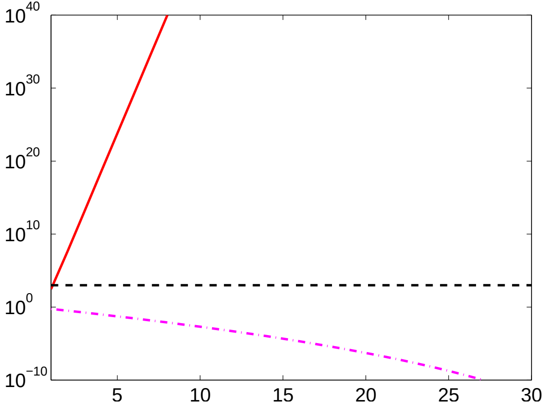
$$A = Q\Lambda Q^T, \qquad \Lambda = \text{logspace}(-4, 0, n)$$
$$b = \text{randn}(n)$$
$$x_0 = 0$$



Without reorthogonalization

Example 2

$$A = \mathrm{diag}(0.001, \ldots, 0.01, 100, \ldots, 1000)$$
$$b = [1, \ldots, 1, 0, \ldots 0]^T$$
$$x_0 = 0$$



By design

$\|\rho_k(A)\|_2 \approx 10^{5k}$

$\|\epsilon_k\|_A / \|\epsilon_0\|_A$
$\|A^{-1}\|_2$
$\|J_k\|_2$

Outline

- Motivation / Possible Applications
- The Jacobian and 2-norm Condition Number of x_k
- Computing Matvecs with J_k and J_k^T
- Summary and Future Research Directions

Motivation

- obviously not practical to explicitly compute and store $J_k \in \mathbb{R}^{n \times n}$
- computing $J_k v$ is useful for estimating norms:
  e.g. with $v \sim \mathcal{N}(0, I)$, $\|J_k\|_2 \geq \frac{\|J_k v\|_2}{\|v\|_2}$, $\mathbb{E}\{\|J_k v\|_2^2\} = \|J_k\|_F^2$
- [Moore et. al. (2012)] interested in

$$v^T \mathrm{cov}\{x_k - x\} v \approx (v^T J_k) \Sigma (J_k^T v)$$

Motivation

- obviously not practical to explicitly compute and store $J_k \in \mathbb{R}^{n \times n}$
- computing $J_k v$ is useful for estimating norms:
  e.g. with $v \sim \mathcal{N}(0, I)$, $\quad \|J_k\|_2 \geq \frac{\|J_k v\|_2}{\|v\|_2}, \quad \mathbb{E}\{\|J_k v\|_2^2\} = \|J_k\|_F^2$
- [Moore et. al. (2012)] interested in

$$v^T \mathrm{cov}\{x_k - x\} v \approx (v^T J_k) \Sigma (J_k^T v)$$

- straightforward to differentiate the algorithm line by line:
  e.g. $x_k = x_{k-1} + \alpha_k p_k \quad \Rightarrow \quad \partial x_k = \partial x_{k-1} + \alpha_k \partial p_k + p_k \partial \alpha_k$

Motivation

- obviously not practical to explicitly compute and store $J_k \in \mathbb{R}^{n \times n}$
- computing $J_k v$ is useful for estimating norms:
  e.g. with $v \sim \mathcal{N}(0, I)$, $\|J_k\|_2 \geq \frac{\|J_k v\|_2}{\|v\|_2}$, $\mathbb{E}\{\|J_k v\|_2^2\} = \|J_k\|_F^2$
- [Moore et. al. (2012)] interested in

$$v^T \text{cov}\{x_k - x\} v \approx (v^T J_k) \Sigma (J_k^T v)$$

- straightforward to differentiate the algorithm line by line:
  e.g. $x_k = x_{k-1} + \alpha_k p_k \quad \Rightarrow \quad \partial x_k = \partial x_{k-1} + \alpha_k \partial p_k + p_k \partial \alpha_k$
- main cost: one extra matvec with A per iteration

$$q_k = A p_k \quad \Rightarrow \quad \partial q_k = A \partial p_k$$

Automatic Differentiation

- express step k of CG as $z_k = \begin{bmatrix} x_k \\ r_k \\ p_k \end{bmatrix} = F(z_{k-1})$
- then

$$J_k = \frac{\partial x_k}{\partial b} = [I_n, 0, 0] \frac{\partial z_k}{\partial b} = [I_n, 0, 0] \left(\frac{\partial z_k}{\partial z_{k-1}} \right) \ldots \left(\frac{\partial z_1}{\partial z_0} \right) \left(\frac{\partial z_0}{\partial b} \right)$$

Automatic Differentiation

- express step k of CG as $\quad z_k = \begin{bmatrix} x_k \\ r_k \\ p_k \end{bmatrix} = F(z_{k-1})$
- then

$$J_k = \frac{\partial x_k}{\partial b} = [I_n, 0, 0] \frac{\partial z_k}{\partial b} = [I_n, 0, 0] \left(\frac{\partial z_k}{\partial z_{k-1}} \right) \dots \left(\frac{\partial z_1}{\partial z_0} \right) \left(\frac{\partial z_0}{\partial b} \right)$$

- "forward" or "direct" mode:

$$J_k v = [I_n, 0, 0] \left(\frac{\partial z_k}{\partial z_{k-1}} \right) \dots \left(\frac{\partial z_1}{\partial z_0} \right) \left(\frac{\partial z_0}{\partial b} \right) v$$

- "reverse" or "adjoint" mode:

$$J_k^T v = \left(\frac{\partial z_0}{\partial b} \right)^T \left(\frac{\partial z_1}{\partial z_0} \right)^T \dots \left(\frac{\partial z_k}{\partial z_{k-1}} \right)^T [I_n, 0, 0]^T v$$

Outline

- Motivation / Possible Applications
- The Jacobian and 2-norm Condition Number of x_k
- Computing Matvecs with J_k and J_k^T
- Summary and Future Research Directions

Summary

- given an expression for $J_k = \frac{\partial x_k}{\partial b} \in \mathbb{R}^{n \times n}$
- showed how J_k relates to A^{-1} and $V_k T_k^{-1} V_k^T$

Summary

- given an expression for $J_k = \frac{\partial x_k}{\partial b} \in \mathbb{R}^{n \times n}$
- showed how J_k relates to A^{-1} and $V_k T_k^{-1} V_k^T$

Open Problems

- main challenge: how to compute $J_k v$ and $J_k^T v$ efficiently?
  - automatic differentiation requires k matvecs with A
  - $J_k = 2V_k T_k^{-1} V_k^T + \left(I - 2V_k T_k^{-1} V_k^T A\right) \zeta_{k-1}(A)$

Summary

- given an expression for $J_k = \dfrac{\partial x_k}{\partial b} \in \mathbb{R}^{n \times n}$
- showed how J_k relates to A^{-1} and $V_k T_k^{-1} V_k^T$

Open Problems

- main challenge: how to compute $J_k v$ and $J_k^T v$ efficiently?
  - automatic differentiation requires k matvecs with A
  - $J_k = 2V_k T_k^{-1} V_k^T + \left(I - 2V_k T_k^{-1} V_k^T A\right)\zeta_{k-1}(A)$
  - in the above, $A \approx V_k T_k V_k^T$ gives $J_k \approx V_k T_k^{-1} V_k^T$
  - replace $\zeta_{k-1}(\cdot)$ by a lower-degree polynomial ?

Summary

- given an expression for $J_k = \dfrac{\partial x_k}{\partial b} \in \mathbb{R}^{n \times n}$
- showed how J_k relates to A^{-1} and $V_k T_k^{-1} V_k^T$

Open Problems

- main challenge: how to compute $J_k v$ and $J_k^T v$ efficiently?
  - automatic differentiation requires k matvecs with A
  - $J_k = 2V_k T_k^{-1} V_k^T + \left(I - 2V_k T_k^{-1} V_k^T A\right) \zeta_{k-1}(A)$
  - in the above, $A \approx V_k T_k V_k^T$ gives $J_k \approx V_k T_k^{-1} V_k^T$
  - replace $\zeta_{k-1}(\cdot)$ by a lower-degree polynomial ?
- extend to other polynomial-based iterative methods
- extend to other derivatives $\dfrac{\partial x_k}{\partial x_0}, \dfrac{\partial \|\epsilon_k\|}{\partial r_0}, \dfrac{\partial x_k}{\partial D}$, etc...

References

S. Gratton, D. Titley-Peloquin, P. Toint, and J. Tshimanga.
Linearizing the Method of Conjugate Gradients.
Submitted for publication.

(And hopefully more to come on this topic...)

Thank you for your attention!