Adapting Craig's Method for Least-Squares Problems

Eric Hallman

University of California, Berkeley

Fall 2018

LSQR

Iterative algorithm for

$$\min_x \|Ax - b\|_2^2$$

- Developed in 1982 (Paige, Saunders)
- Properties of LSQR
  - Minimizes $\|r_k\| := \|b - Ax_k\|$ for every iterate x_k
  - Equivalent to CG on the normal equations $A^T Ax = A^T b$
  - $\|x_k\|$ monotonically increasing (update directions positively correlated)
  - $\|x_k - x_*\|$ monotonically decreasing
- Cost: $Av, A^T u$ plus $O(m + n)$ operations per iteration ($A \in \mathbb{R}^{m \times n}$)
- Can be adapted to solve the problem $\min_x \|Ax - b\|_2^2 + \lambda^2 \|x\|^2$

Outline

1. Previous Work
   - Golub-Kahan bidiagonalization and LSQR
   - LSLQ and estimating $\|x_k - x_*\|$
2. Our Work
   - Craig's method and minimizing $\|x_k - x_*\|$
   - Main results
   - Open problems and future directions

Golub-Kahan bidiagonalization

Given $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^n$,

$$\begin{aligned} b &= U_k(\beta_1 e_1) \\ AV_k &= U_{k+1} B_k \\ A^T U_k &= V_k L_k^T \end{aligned}$$

where

$$B_k = \begin{pmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \ddots & \ddots & \\ & & \beta_k & \alpha_k \\ & & & \beta_{k+1} \end{pmatrix}, \quad L_k = \begin{pmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \ddots & \ddots & \\ & & \beta_k & \alpha_k \end{pmatrix},$$

and U_k, V_k are orthogonal

Golub-Kahan bidiagonalization

Given $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^n$,

$$\begin{aligned} b &= U_k(\beta_1 e_1) \\ AV_k &= U_{k+1} B_k \\ A^T U_k &= V_k L_k^T \end{aligned}$$

Iterative bidiagonalization

$\beta_1 u_1 = b, \alpha_1 v_1 = A^T u_1$
for $k = 1, 2, \ldots,$ do

$$\begin{aligned} \beta_{k+1} u_{k+1} &= Av_k - \alpha_k u_k \\ \alpha_{k+1} v_{k+1} &= A^T u_{k+1} - \beta_{k+1} v_k \end{aligned}$$

- Cost: $Av_k, A^T u_k$ plus $3m + 3n$ flops
- Only the most recent u_k and v_k are stored

Golub-Kahan bidiagonalization

V_k and U_k span the Krylov subspaces:

$$\operatorname{span}(u_1, \ldots, u_k) = \operatorname{span}(b, (AA^T)b, \ldots, (AA^T)^{k-1}b),$$
$$\operatorname{span}(v_1, \ldots, v_k) = \operatorname{span}(A^Tb, (A^TA)A^Tb, \ldots, (A^TA)^{k-1}A^Tb)$$

Defining $x_k := V_k y_k$ and $r_k := b - Ax_k$, get

LSQR Subproblem

$$\min_{x_k} \|r_k\| = \min_{y_k} \|\beta_1 e_1 - B_k y_k\|$$

where B_k is $(k+1) \times k$.

LSQR Subproblem

$$
\begin{aligned}
&\min_{x_k} \|r_k\| && \\
=&\min_{y_k} \|b - AV_k y_k\| && x_k = V_k y_k \\
=&\min_{y_k} \|U_{k+1}(\beta_1 e_1 - B_k y_k)\| && b = \beta_1 u_1, AV_k = U_{k+1} B_k \\
=&\min_{y_k} \left\| \begin{pmatrix} f_k \\ \phi'_{k+1} \end{pmatrix} - \begin{pmatrix} R_k y_k \\ 0 \end{pmatrix} \right\| && Q_{k+1} \begin{pmatrix} B_k & \beta_1 e_1 \end{pmatrix} = \begin{pmatrix} R_k & f_k \\ 0 & \phi'_{k+1} \end{pmatrix} \\
=&|\phi'_{k+1}| && y_k = R_k^{-1} f_k
\end{aligned}
$$

Computation

$$x_k = V_k y_k = (V_k R_k^{-1}) f_k = D_k f_k = x_{k-1} + \phi_k d_k \qquad [d_k = D_k(:, k)]$$

Outline

1. Previous Work
   - Golub-Kahan bidiagonalization and LSQR
   - LSLQ and estimating $\|x_k - x_*\|$
2. Our Work
   - Craig's method and minimizing $\|x_k - x_*\|$
   - Main results
   - Open problems and future directions

Estimating $\|x_k - x_*\|$

How can we estimate $\|x_k - x_*\|$?

- Suppose we have $\widetilde{\sigma} \leq \sigma_{\min}(A)$ (or regularization with $\lambda > 0$)
- Naive: $\|x_k - x_*\| = \|A^\dagger(Ax_k - b)\| \leq \|r_k\|/\widetilde{\sigma}$
- ...but $\|r_k\|$ may not converge to zero

Estimating $\|x_k - x_*\|$

Key idea (Estrin, Orban, and Saunders, 2017):

Theorem

Define $e_1 = [1, 0, \ldots, 0]^T$ *and*

$$\widetilde{R}_{k+1} = \begin{bmatrix} R_k & \theta_{k+1} e_k \\ & \omega \end{bmatrix} = \begin{bmatrix} \rho_1 & \theta_2 & & & \\ & \rho_2 & \theta_3 & & \\ & & \ddots & \ddots & \\ & & & \rho_k & \theta_{k+1} \\ & & & & \omega \end{bmatrix}$$

where ω *is chosen so that* $\sigma_{\min}(\widetilde{R}_{k+1}) \le \widetilde{\sigma} \le \sigma_{\min}(A)$. *Then*

$$\|x_*\|^2 \le \alpha_1^2 \beta_1^2 e_1^T \left(\widetilde{R}_{k+1}^T \widetilde{R}_{k+1} \right)^{-2} e_1 .$$

- By properties of LSQR, $\|x_k^{LSQR} - x_*\|^2 \le \|x_*\|^2 - \|x_k^{LSQR}\|^2$.
- The bound converges to zero when $\widetilde{\sigma} > 0$.

Estimating $\|x_k - x_*\|$: LSLQ

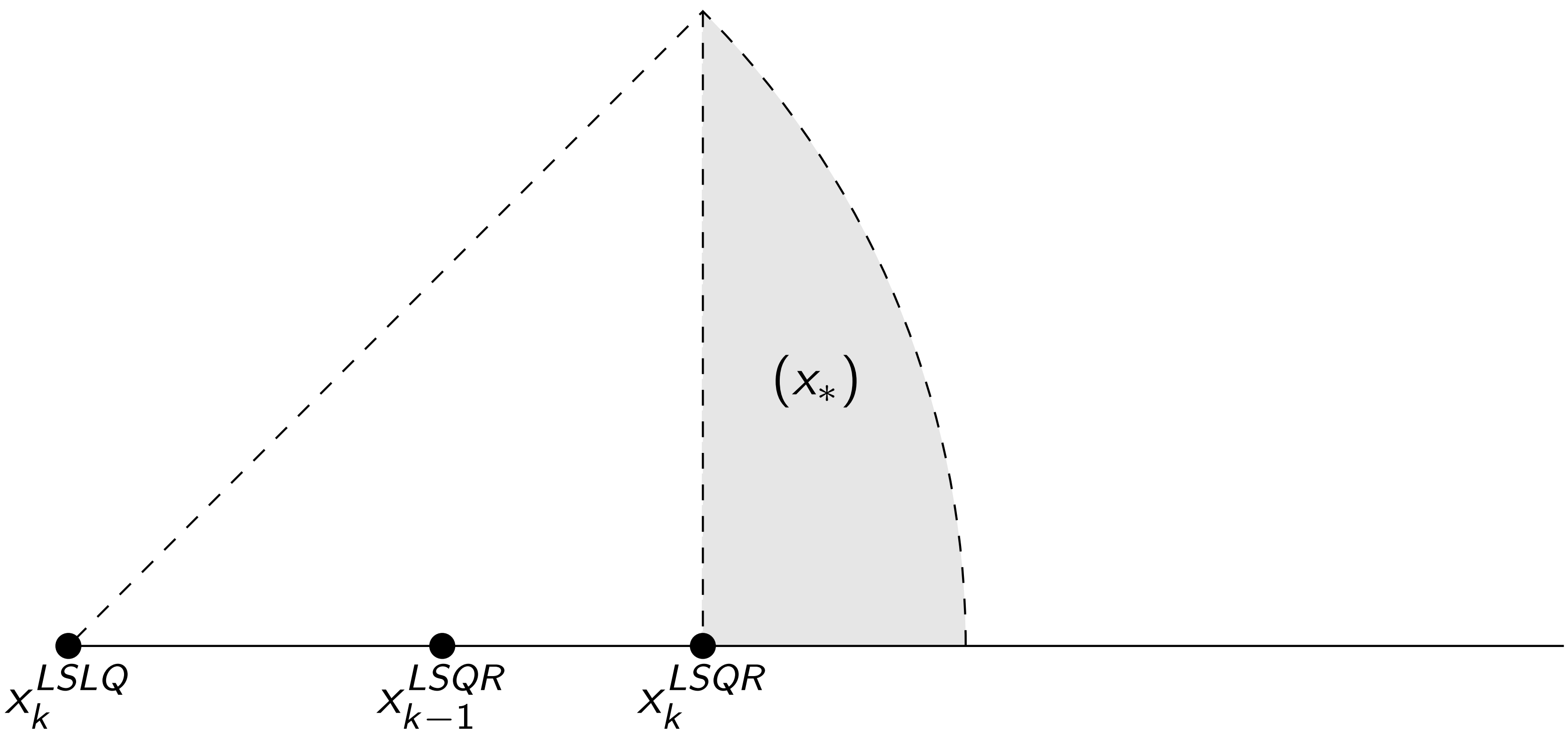
We can use the related algorithm LSLQ (Estrin, Orban, and Saunders, 2017) to tighten the bound.

- LSLQ computes $x_k^{LSLQ} = V_k y_k$, where

$$y_k = \arg\min_y \|y\| \quad : \quad [R_{k-1},\ \theta_k e_{k-1}]y = f_{k-1}$$

- Orthogonal update directions
- $\|x_k^{LSLQ} - x_*\|$ monotonically decreasing
- $y_k^{LSLQ}, y_{k-1}^{LSQR}$ and y_k^{LSQR} all solve $[R_{k-1},\ \theta_k e_{k-1}]y = f_{k-1}\ldots$ collinear!

Estimating $\|x_k - x_*\|$: LSLQ



- Old bound: $\|x_k^{LSQR} - x_*\|^2 \leq \|x_*\|^2 - \|x_k^{LSQR}\|^2$
- New formulation: $\|x_k^{LSQR} - x_*\|^2 \leq (\|x_* - x_k^{LSLQ}\|^2) - \|x_k^{LSLQ} - x_k^{LSQR}\|^2$

LSLQ Subproblem

$$[R_{k-1},\ \theta_k e_{k-1}]y_k = f_{k-1}$$

$$[\overline{R}_{k-1}^T,\ 0]\overline{Q}_k^T y_k = f_{k-1} \qquad\qquad \overline{Q}_k \begin{bmatrix} R_{k-1}^T \\ \theta_k e_{k-1}^T \end{bmatrix} = \begin{bmatrix} \overline{R}_{k-1} \\ 0 \end{bmatrix}$$

$$y_k = \overline{Q}_k \begin{bmatrix} I \\ 0 \end{bmatrix} \overline{R}_{k-1}^{-T} f_{k-1}$$

$$x_k = V_k \overline{Q}_k \begin{bmatrix} I \\ 0 \end{bmatrix} \overline{f}_{k-1} \qquad\qquad \overline{f}_{k-1} = \overline{R}_{k-1}^{-T} f_{k-1}$$

$$x_k = \overline{V}_{k-1} \overline{f}_{k-1} \qquad\qquad [\overline{v}_{k-1},\ \overline{v}'_k] = [\overline{v}'_{k-1},\ v_k] \begin{bmatrix} \overline{c}_{k-1} & -\overline{s}_{k-1} \\ \overline{s}_{k-1} & \overline{c}_{k-1} \end{bmatrix}$$

$$x_k = x_{k-1} + \overline{\phi}_{k-1} \overline{v}_{k-1}$$

- $\|x_k^{LSLQ}\|$ (and also $\|x_k^{LSLQ} - x_k^{LSQR}\|$) cost O(1) to compute
- Computation of x_k^{LSLQ} not strictly necessary

Outline

1. Previous Work
   - Golub-Kahan bidiagonalization and LSQR
   - LSLQ and estimating $\|x_k - x_*\|$
2. Our Work
   - Craig's method and minimizing $\|x_k - x_*\|$
   - Main results
   - Open problems and future directions

Minimizing $\|x_k - x_*\|$

What happens if we try to minimize $\|x_k - x_*\|$ directly?
One tempting possibility is Craig's method (1955):

- Solves $L_k y_k = \beta_1 e_1$ at each step
- Orthogonal update directions
- Minimizes $\|x_k - x_*\|$ on consistent systems
- . . . but does not converge on inconsistent ones

Minimizing $\|x_k - x_*\|$

What happens if we try to minimize $\|x_k - x_*\|$ directly?

- Earlier theorem:

$$
\begin{aligned}
\|x_*\|^2 &= b^T A(A^T A)^{-2} A^T b \\
&\leq \begin{bmatrix} f_k \\ \phi'_{k+1} \end{bmatrix}^T R'_{k+1}(\widetilde{R}_{k+1}^T \widetilde{R}_{k+1})^{-2}(R'_{k+1})^T \begin{bmatrix} f_k \\ \phi'_{k+1} \end{bmatrix},
\end{aligned}
$$

  where $R'_{k+1} = Q_{k+1} L_{k+1}$.

- Unfortunately,

$$
\begin{aligned}
\|x_k - x_*\|^2 &= r_k^T A(A^T A)^{-2} A^T r_k \\
&\not\leq \begin{bmatrix} f_k - R_k y_k \\ \phi'_{k+1} \end{bmatrix} R'_{k+1}(\widetilde{R}_{k+1}^T \widetilde{R}_{k+1})^{-2}(R'_{k+1})^T \begin{bmatrix} f_k - R_k y_k \\ \phi'_{k+1} \end{bmatrix}.
\end{aligned}
$$

But we can come close!

Minimizing $\|x_k - x_*\|$

Main idea: examine the top left block of $\left([V_{k+1},\ V^{\perp}]^T A^T A [V_{k+1},\ V^{\perp}]\right)^{-1}$.

- $A[V_{k+1},\ V^{\perp}] = [U_{k+2},\ U^{\perp}] \begin{bmatrix} B_{k+1} & e_{k+2}\tilde{v}^T \\ 0 & (U^{\perp})^T A V^{\perp} \end{bmatrix}$
- $\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & * \\ * & * \end{bmatrix}$
- Result:

$$V_{k+1}^T (A^T A)^{-1} V_{k+1} = (\widetilde{R}_{k+1}^T \widetilde{R}_{k+1})^{-1},$$

where

$$\widetilde{R}_{k+1} = \begin{bmatrix} R_k & \theta_{k+1} e_k \\ 0 & \rho'_{k+1}/c^*_{k+1} \end{bmatrix}$$

and $c_{k+1} \le c^*_{k+1} \le 1$.

Minimizing $\|x_k - x_*\|$

Theorem (Hallman, Gu 2018)

Pick $\tilde{c}_{k+1}$ *so that either* $\tilde{c}_{k+1} = 1$ *or* $\sigma_{\min}\left(\widetilde{R}_{k+1}\right) \leq \widetilde{\sigma}$. *Then*

$$\|P_A r_k\| \leq \left\| \begin{bmatrix} f_k - R_k y_k \\ \tilde{c}_{k+1}\phi'_{k+1} \end{bmatrix} \right\|,$$

where $P_A r_k$ *is the projection of* r_k *onto* $\text{span}(A)$.

In particular, $c^*_{k+1} = \|P_A r_k^{LSQR}\| / \|r_k^{LSQR}\|$.

Minimizing $\|x_k - x_*\|$

Apply the same ideas to $\left([V_{k+1},\ V^\perp]^T A^T A [V_{k+1},\ V^\perp]\right)^{-2}$.

- Result:

$$V_{k+1}^T (A^T A)^{-2} V_{k+1} = (\widetilde{R}_{k+1}^T \widetilde{R}_{k+1})^{-1} \begin{bmatrix} I & 0 \\ 0 & \xi_{k+1}^{-2} \end{bmatrix} (\widetilde{R}_{k+1}^T \widetilde{R}_{k+1})^{-1},$$

where ξ_{k+1} is somewhat difficult to bound effectively.

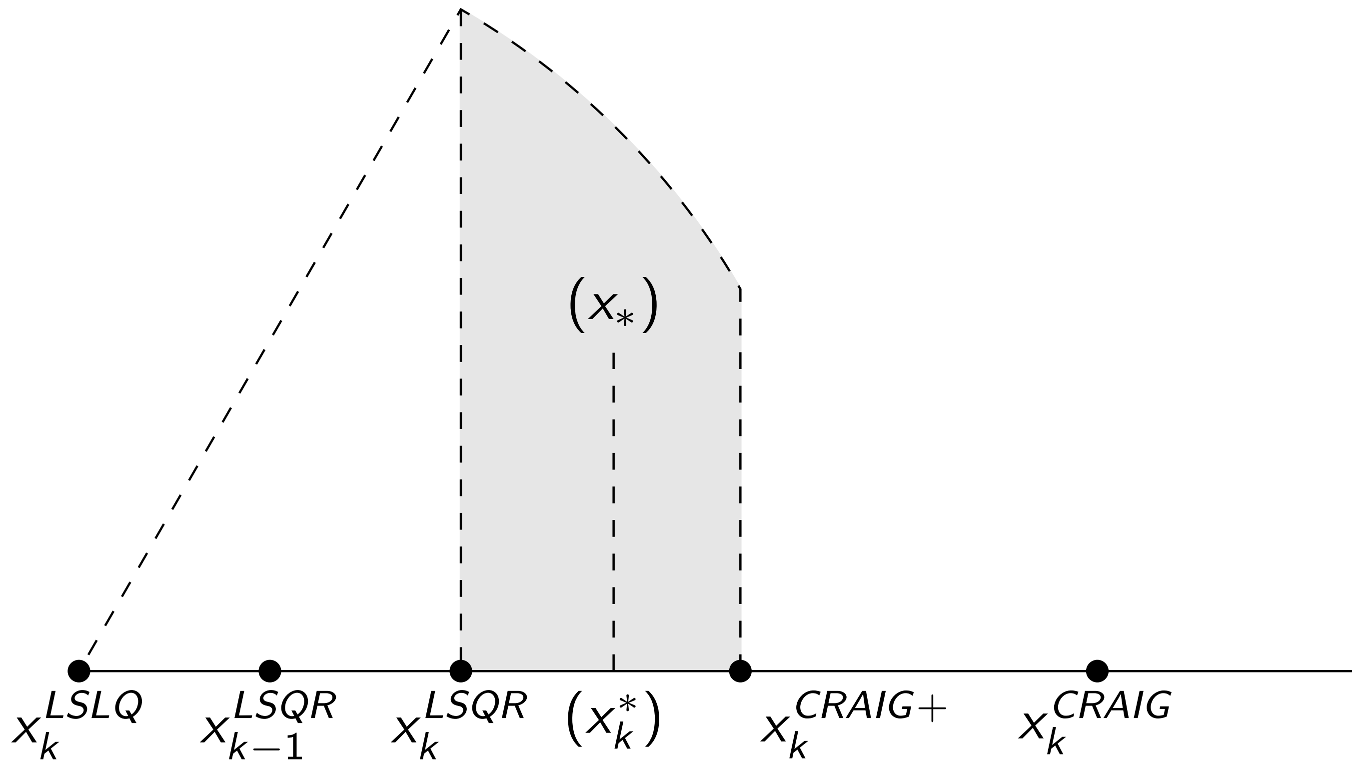
Outline

Main Result

We can minimize $\|x_k - x_*\|$ without knowing ξ_{k+1}!

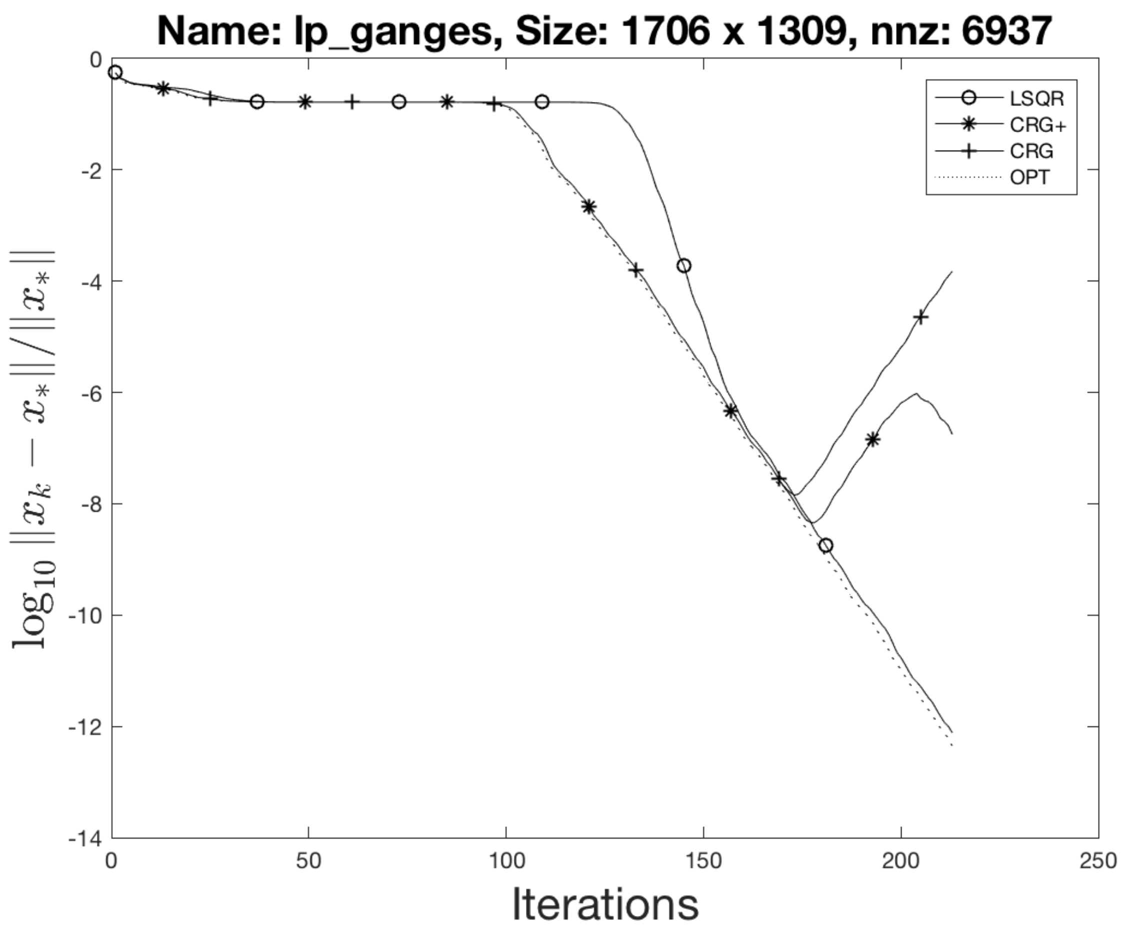
- $\|x_k - x_*\| = \left\| \begin{bmatrix} R_k^{-1} f_k - y_k - \theta_{k+1}(c_{k+1}^*)^2 \phi'_{k+1}/\rho'_{k+1} R_k^{-1} e_k \\ (c_{k+1}^*)^2 \phi'_{k+1}/(\rho'_{k+1}\xi_{k+1}) \end{bmatrix} \right\|$
- Minimized when $R_k y_k = f_k - \theta_{k+1}(c_{k+1}^*)^2 \frac{\phi'_{k+1}}{\rho'_{k+1}} e_k$
- If $c_{k+1}^* = 0$, we recover LSQR
- If $c_{k+1}^* = 1$, we recover Craig's method

Main Result



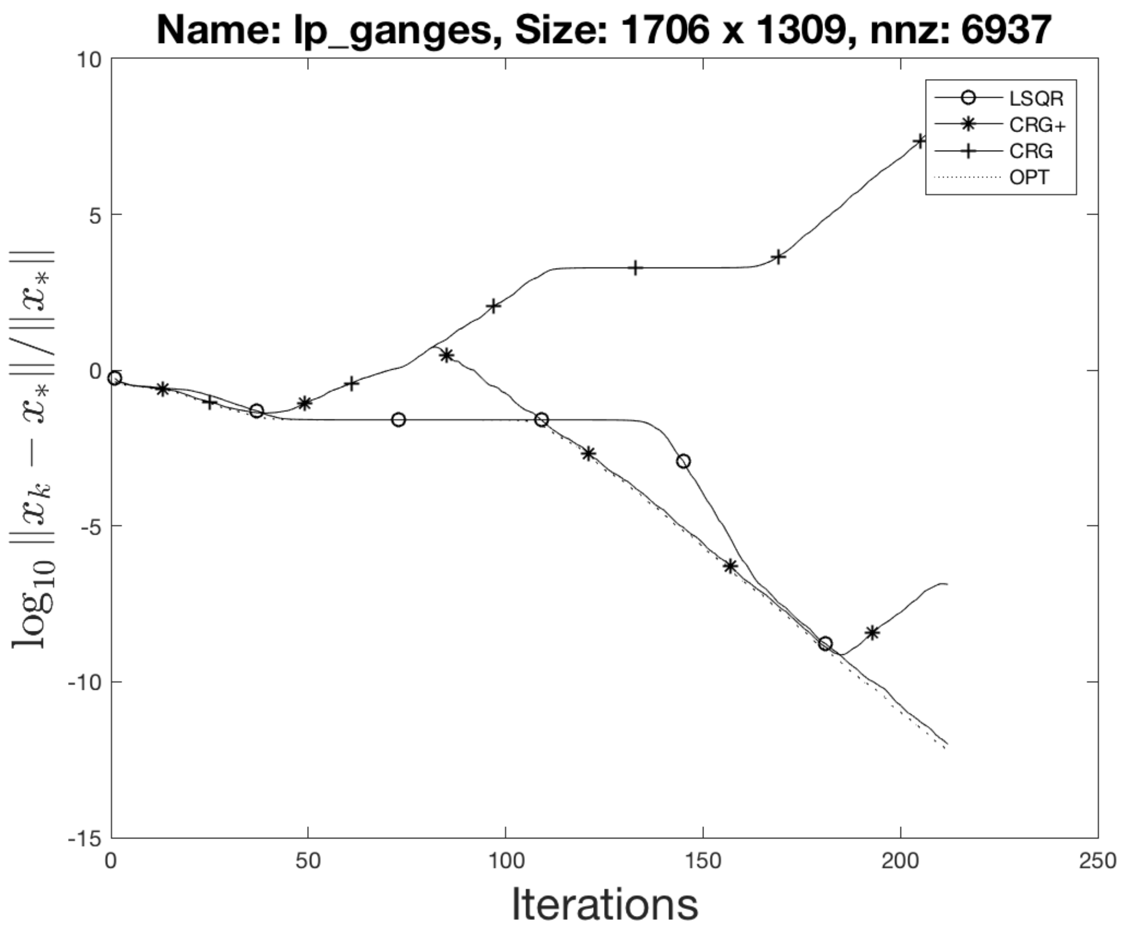
- Six collinear points
- Given $0 < \widetilde{\sigma} \leq \sigma_{\min}(A)$, x_k^{CRAIG+} will converge to x_*
- Minimizing $\|x_k - x_*\|$ is equivalent to measuring $\|P_A r_k\|$
- x_k^{CRAIG+} beats x_k^{LSQR} if and only if $\tilde{c}_{k+1} \leq 2c_{k+1}^*$

Main Result



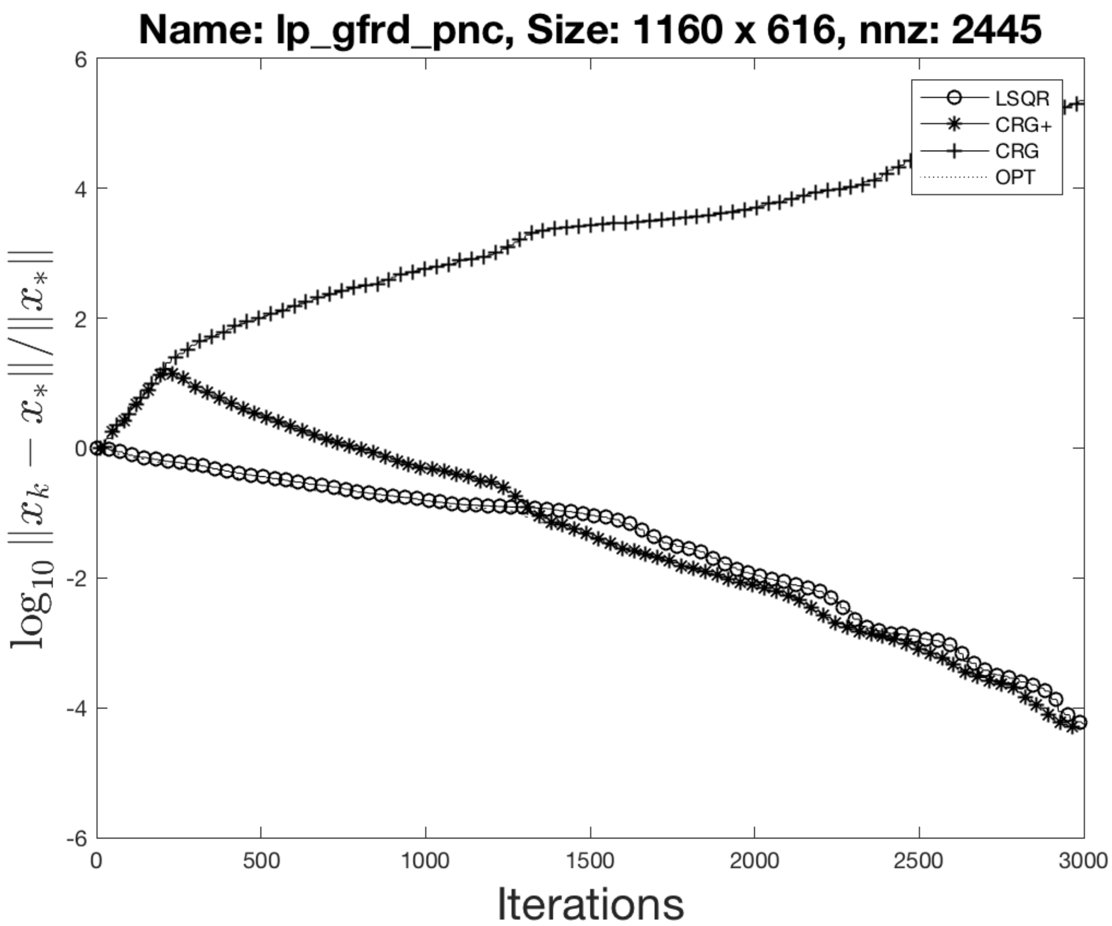
- Craig's method can sometimes outperform LSQR
- It is sometimes possible for our method to outperform both

Main Result



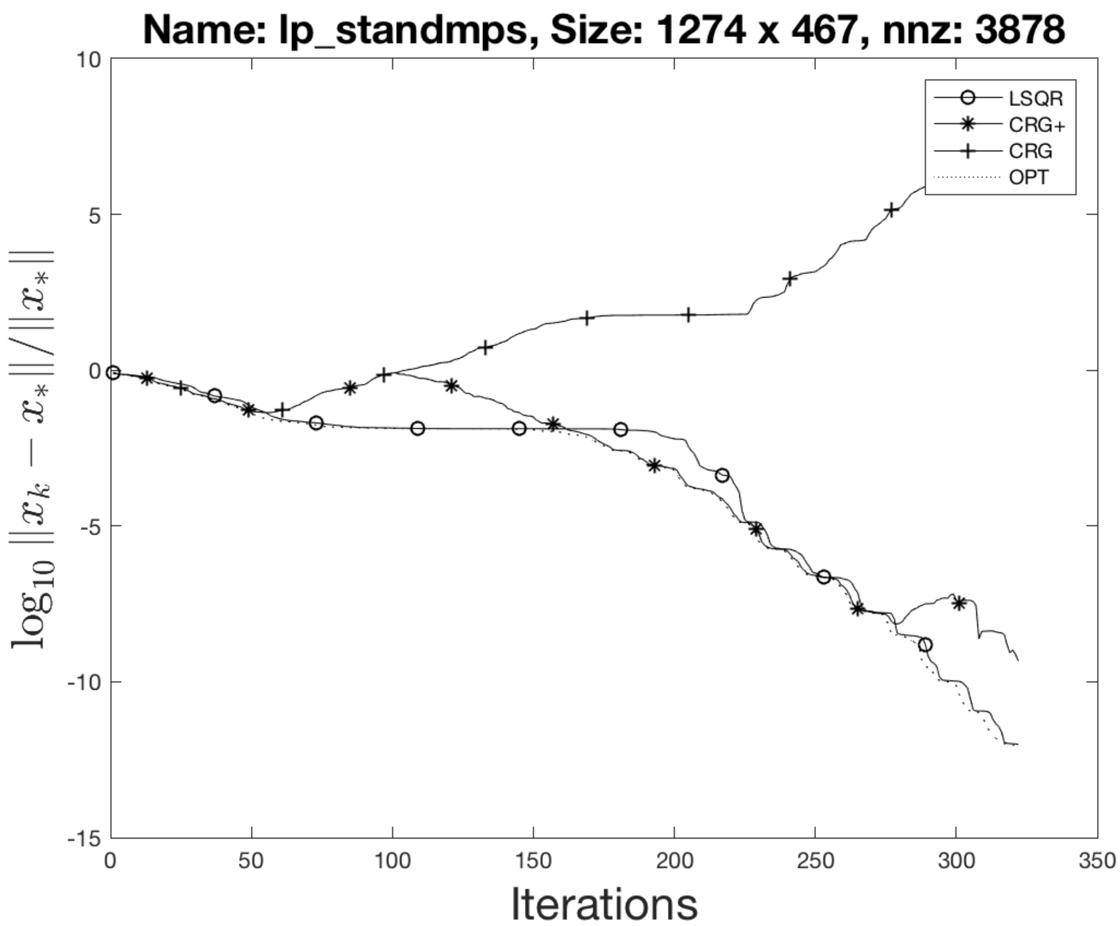
- Craig's method can sometimes outperform LSQR
- It is sometimes possible for our method to outperform both

Main Result



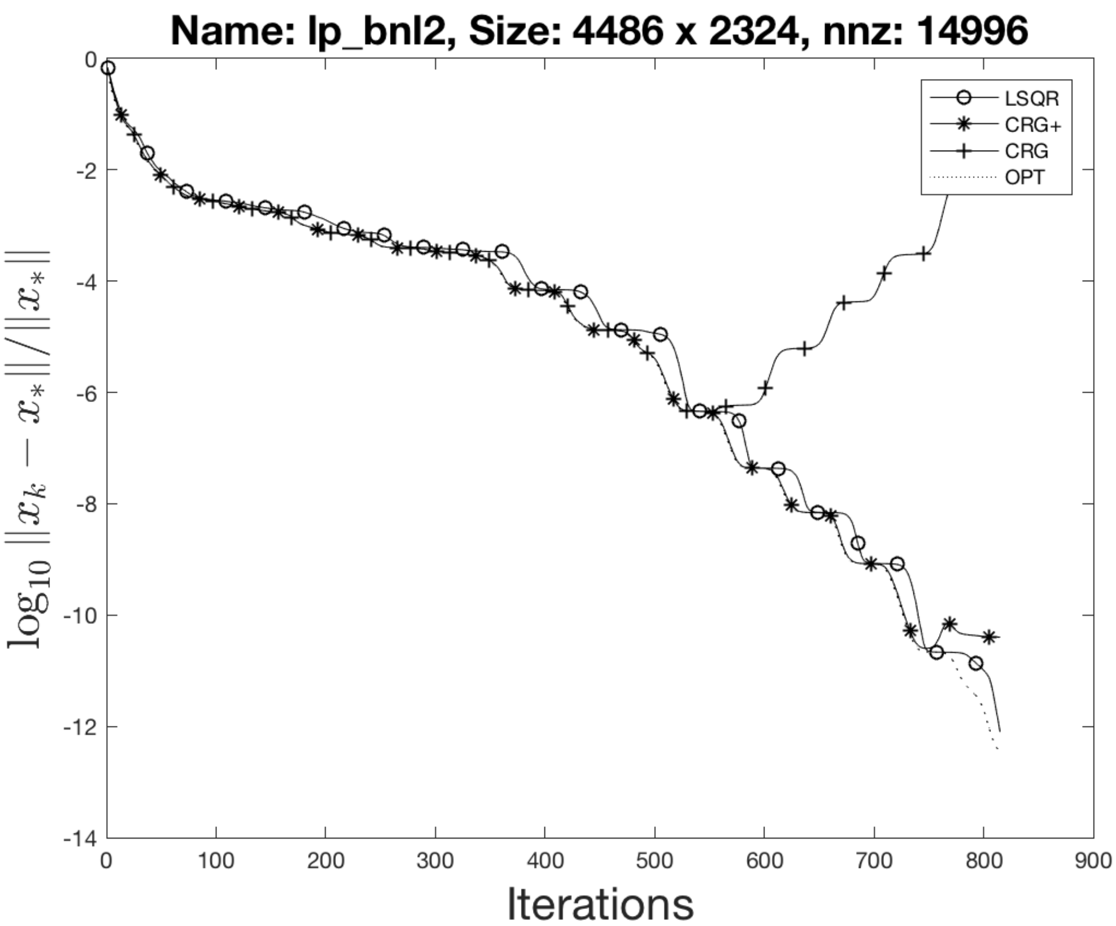
- Craig's method can sometimes outperform LSQR
- It is sometimes possible for our method to outperform both

Main Result



- Craig's method can sometimes outperform LSQR
- It is sometimes possible for our method to outperform both

Main Result



- Craig's method can sometimes outperform LSQR
- It is sometimes possible for our method to outperform both

Outline

1. Previous Work
   - Golub-Kahan bidiagonalization and LSQR
   - LSLQ and estimating $\|x_k - x_*\|$
2. Our Work
   - Craig's method and minimizing $\|x_k - x_*\|$
   - Main results
   - Open problems and future directions

Remaining Issues

- Stably estimate $\tilde{c}_{k+1}$ (equivalently, ω) from $\widetilde{R}_{k+1}$
- Improve the estimate of $\|x_k - x_*\|$

Both are necessary for our method to be practical!

Estimating ω

We would like to solve

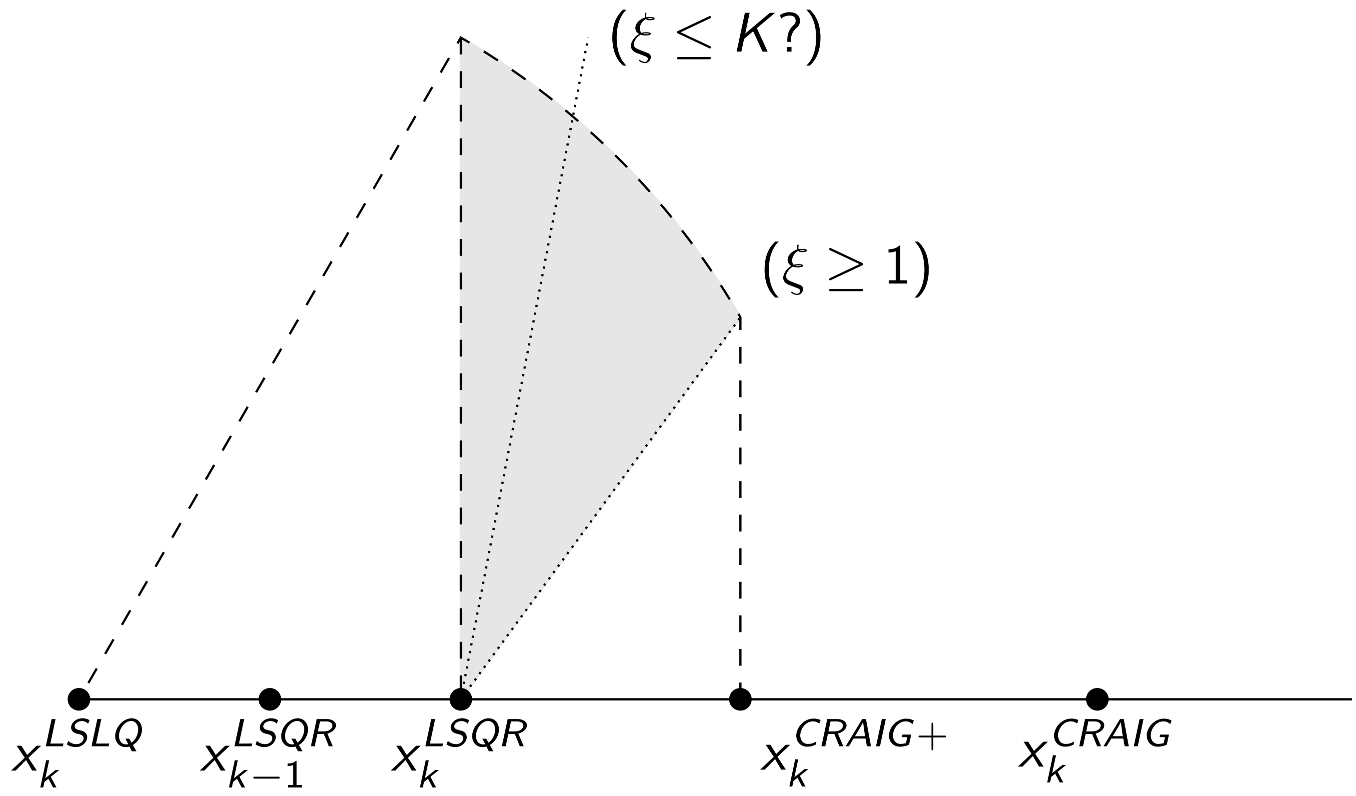
$$\max\ \omega\ :\ \sigma_{\min}\left(\begin{bmatrix} R_k & \theta_{k+1} e_k \\ 0 & \omega \end{bmatrix}\right) \leq \widetilde{\sigma} \leq \sigma_{\min}(A)$$

as accurately as possible.

- Implicit Cholesky on $\widetilde{R}_{k+1}^T \widetilde{R}_{k+1} - \sigma^2 I$
- If $\widetilde{\sigma} \leq \sigma_{\min}(A)$ is too conservative, the bounds are weak
- If $\widetilde{\sigma}$ is too aggressive, the factorization will break down
- Suggestion: $\widetilde{\sigma} = \sigma_{\min}(A)(1 - 10^{-10})$ (Estrin, Orban, Saunders 2017)
- It seems that we cannot avoid subtraction (Tichý, Meurant, Strakoš 2014)

Estimating $\|x_k - x_*\|$

Without an upper bound on ξ (or some equivalent), we cannot beat the error bound for LSQR.



Summary

- LSLQ, LSQR, and Craig's method produce collinear iterates
- Also collinear with $x_k^* = \arg\min_{x_k \in \text{span}(V_k)} \|x_k - x_*\|$
- Finding x_k^* is equivalent to measuring $\|P_A r_k\|$
- Given $\widetilde{\sigma} > 0$ (or $\lambda > 0$), x_k^{CRAIG+} converges to x_*
- x_k^{CRAIG+} might outperform LSQR
- Performance depends on how close $\widetilde{\sigma}$ is to $\sigma_{\min}(A)$
- LSQR is often close to optimal

Future Directons

- Find practical upper bounds for ξ
- Extend to SPD problems–SYMMLQ and CG?

For Further Reading I

- Ron Estrin, Dominique Orban, and Michael Saunders
  LSLQ: An Iterative Method for Linear Least-Squares with an Error Minimization Property
  http://stanford.edu/ restrin/files/eos2017.pdf
- Ron Estrin, Dominique Orban, and Michael Saunders
  LNLQ: An Iterative Method for Least-Norm Problems with an Error Minimization Property
  http://stanford.edu/ restrin/files/eos2018.pdf
- Petr Tichý, Gérard Meurant, and Zdeněk Strakoš
  A New Algorithm for Computing Quadrature-Based Bounds in Conjugate Gradients
  http://www.cs.cas.cz/tichy/download/present/2014Spa.pdf
- Chris Paige and Michael Saunders
  LSQR: An Algorithm for Sparse Linear Equations and Sparse Least Squares
  ACM Transactions on Mathematical Software 8(1):43-71, 1982.