Fast & Forward Stable Randomized Algorithm for LLSP

Justin Mucelli and Marguerite Demasi

Linear Least Square Problem

The overdetermined linear least squares problem is given

$$\mathbf{x} = \arg\min_{\mathbf{y} \in \mathbb{R}^n} ||\mathbf{b} - \mathbf{A}\mathbf{y}||$$

for $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$

Algorithm – Iterative Sketching

Input: Matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$,RHS $\mathbf{b} \in \mathbb{R}^{m}$, Embedding dimension d iteration count q.

Output: Approx solution x_q to least squares.

(1) $\mathbf{S} \leftarrow d \times m$	(1) Sketching matrix
(2) $\mathbf{B} \leftarrow \mathbf{SA}$	(2) Applying the sketch
(3) $(\mathbf{Q}, \mathbf{R}) \leftarrow \mathrm{QR}(\mathbf{B}, \text{'econ'})$	
(4) $\mathbf{x}_0 \leftarrow \mathbf{R}^{-1}(\mathbf{Q}^*(\mathbf{Sb}))$	(3,4) Householder QR
(5) for $i = 0 : q-1$	
(6) $\quad \mathbf{x}_{i+1} \leftarrow \mathbf{x}_i + \mathbf{R}^{-1}\mathbf{R}^{-\top}\mathbf{A}^{\top}(\mathbf{b} - \mathbf{A}\mathbf{x}_i)$	(6) Update rule
(7) end	

Algorithm - Justification

(1) $(\mathbf{A}^T\mathbf{A})\mathbf{x} = \mathbf{A}^T\mathbf{b}$

(2) $(\mathbf{SA})^T(\mathbf{SA}) \approx \mathbf{A}^T\mathbf{A}$

(3) $\mathbf{A}^T\mathbf{A}(\mathbf{x} - \mathbf{x}_i) = \mathbf{A}^T(\mathbf{b} - \mathbf{A}\mathbf{x}_i)$

(4) $(\mathbf{SA})^T(\mathbf{SA})d_i = \mathbf{A}^T(\mathbf{b} - \mathbf{A}\mathbf{x}_i)$

(5) where, $\mathbf{x} \approx \mathbf{x}_{i+1} = \mathbf{x}_i + \mathbf{d}_i$

(6) $\mathbf{SA} = \mathbf{QR}$

(7) $(\mathbf{SA})^T(\mathbf{SA}) = \mathbf{R}^T\mathbf{R}$

(8) $\mathbf{d}_i = \mathbf{R}^{-1}\mathbf{R}^{-T}\mathbf{A}^T(\mathbf{b} - \mathbf{A}\mathbf{x}_i)$

(9) $\mathbf{x}_{i+1} = \mathbf{x}_i + \mathbf{R}^{-1}\mathbf{R}^{-T}\mathbf{A}^T(\mathbf{b} - \mathbf{A}\mathbf{x}_i)$

Notational Notes

- **Relative Forward Error:** $FE(\hat{\mathbf{x}}) = \frac{\|\mathbf{x}-\hat{\mathbf{x}}\|}{\|\mathbf{x}\|}$
- **Relative Residual:** $RE(\hat{\mathbf{x}}) = \frac{\|\mathbf{r}(\mathbf{x})-\mathbf{r}(\hat{\mathbf{x}})\|}{\|\mathbf{r}(\mathbf{x})\|}$ where $\mathbf{r}(\mathbf{y}) = \mathbf{b} - \mathbf{A}\mathbf{y}$
- **Relative Backward Error:** $BE(\hat{\mathbf{x}}) = \min\left\{\frac{\|\Delta\mathbf{A}\|_F}{\|\mathbf{A}\|_F} : \hat{\mathbf{x}} = \arg\min_{\mathbf{v}\in\mathbb{R}^n}\|\mathbf{b} - (\mathbf{A}+\Delta\mathbf{A})\mathbf{v}\|\right\}$
- **Machine Epsilon:** $u = \epsilon_{mach} \approx 10^{-16}$
- **Condition Number:** $\kappa(\mathbf{A}) = \kappa = \|\mathbf{A}\|\|\mathbf{A}^{-1}\| = \sigma_{\max}(\mathbf{A})/\sigma_{min}(\mathbf{A})$

If BE is small, then computed solution is the *true solution to the nearly the right least-square problem*
Informally, if BE ≅ *u* then the algorithm is *backwards stable*

Comparing with Backward Stable Algo

- Householder QR is *backwards stable* i.e. $\hat{\mathbf{x}} = \arg\min_{y \in \mathbb{R}^n} \|(\mathbf{b} + \Delta\mathbf{b}) - (\mathbf{A} + \Delta\mathbf{A})\mathbf{y}\|$ with perturbations of size $\|\Delta\mathbf{A}\| \leq cu\|\mathbf{A}\|$ and $\|\Delta\mathbf{b}\| \leq cu\|\mathbf{b}\|$ provided $cu < 1$

- **Main Result:** iterative sketching has forward errors and residual errors comparable to a backward stable algorithm (e.g. Householder QR)

But what are the forward and residual errors for a backward stable algorithm?

Wedin's Perturbation Theorem

Let $\mathbf{A}, \Delta\mathbf{A} \in \mathbb{R}^{m\times n}$ and$\mathbf{b}, \Delta\mathbf{b} \in \mathbb{R}^{m}$and suppose that$\|\Delta\mathbf{A}\| \leq \epsilon\|\mathbf{A}\|$ and $\|\Delta\mathbf{b}\| \leq \epsilon\|\mathbf{b}\|$ for some $(0,1)$. Then set

$$\mathbf{x} = \arg\min_{\mathbf{y}\in\mathbb{R}^n} \|\mathbf{b} - \mathbf{A}\mathbf{y}\|$$

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{y}\in\mathbb{R}^n} \|(\mathbf{b} + \Delta\mathbf{b}) - (\mathbf{A} + \Delta\mathbf{A})\mathbf{y}\|$$

Then if $\epsilon\kappa(\mathbf{A}) \leq 0.1$

$$\|\mathbf{x} - \hat{\mathbf{x}}\| \leq 2.23\kappa(\mathbf{A}) \left(\|\mathbf{x}\| + \frac{\kappa(\mathbf{A})}{\|\mathbf{A}\|}\|\mathbf{r}(\mathbf{x})\|\right)\epsilon$$

$$\|\mathbf{r}(\mathbf{x}) - \mathbf{r}(\hat{\mathbf{x}})\| \leq 2.23\left(\|\mathbf{A}\|\,\|\mathbf{x}\| + \kappa(\mathbf{A})\|\mathbf{r}(\mathbf{x})\|\right)\epsilon$$

Results from Wedin's Theorem

Assuming κu sufficiently small, then the solution $\hat{\mathbf{x}}$ to a *backwards stable* least square solver satisfies

$$\|\mathbf{x} - \hat{\mathbf{x}}\| \leq c\prime \kappa(\mathbf{A}) \left(\|\mathbf{x}\| + \frac{\kappa}{\|\mathbf{A}\|} \|\mathbf{r}(\mathbf{x})\| \right) u$$

$$\|\mathbf{r}(\mathbf{x}) - \mathbf{r}\hat{\mathbf{x}}\| \leq c\prime \left(\|\mathbf{A}\| \|\mathbf{x}\| + \kappa \|\mathbf{r}(\mathbf{x})\| \right) u$$

An algorithm is said to be *forward stable* if the computed solution satisfies bounds of this form

Theorem VI: Iterative sketching is forward stable

Let $\mathbf{S} \in \mathbb{R}^{d \times m}$ be a subspace embedding for range($[\mathbf{A}\ \mathbf{b}]$) with distortion $\epsilon \in (0, 0.29]$. There exists a constant $c_1 > 0$ which depends polynomially on m, n, and d such that $c_1 \kappa u < 1$ and multiplication by $\mathbf{S}$ is forward stable, then the numerically computed iterates $\hat{\mathbf{x}}_i$ by Iterative Method satisfy bounds

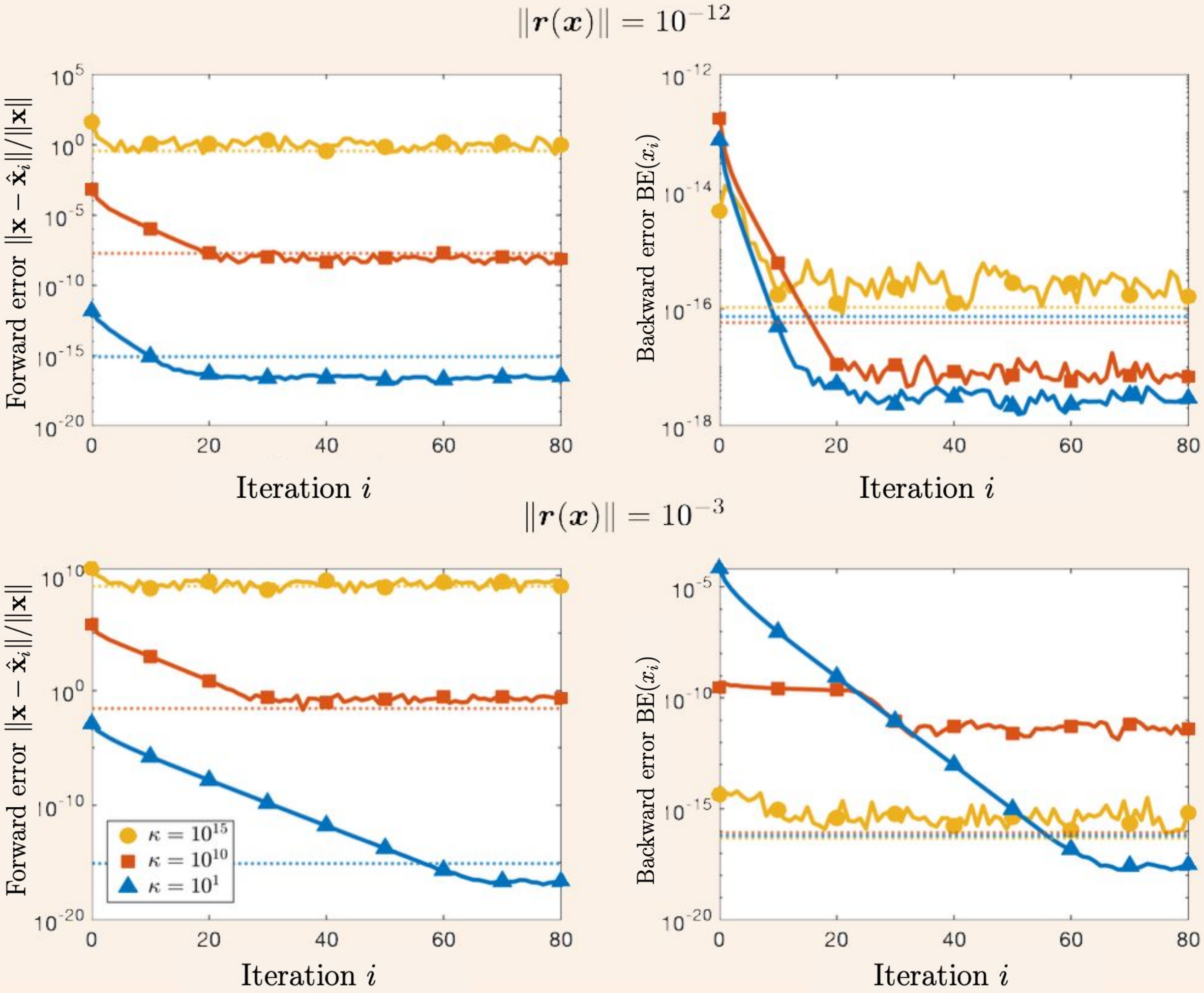
$$\|\mathbf{x} - \hat{\mathbf{x}}_i\| \leq 20\sqrt{\epsilon}\kappa(g_{IS} + c_1\kappa u)^i \frac{\|\mathbf{r}(\mathbf{x})\|}{\|\mathbf{A}\|} + c_1\kappa u \left[\|\mathbf{x}\| + \frac{\kappa}{\|\mathbf{A}\|}\|\mathbf{r}(\mathbf{x})\|\right]$$

$$\|\mathbf{r}(\mathbf{x}) - \mathbf{r}(\hat{\mathbf{x}}_i)\| \leq 20\sqrt{\epsilon}(g_{IS} + c_1\kappa u)^i \|\mathbf{r}(\mathbf{x})\| + c_1 u \left[\|\mathbf{A}\|\|\mathbf{x}\| + \kappa\|\mathbf{r}(\mathbf{x})\|\right]$$

In particular, if $g_{IS} + c_i\kappa u \leq 0.9$ then the iterative sketching produces a solution $\hat{\mathbf{x}}$ such it satisfies the forward stability error bounds

Informally...

(**Informal Theorem I**) In FPA, iterative sketching produces iteratives $\hat{\mathbf{x}}_0, \hat{\mathbf{x}}_i, \ldots$ with errors $\|\mathbf{x} - \hat{\mathbf{x}}_i\|$ and residual errors $\|\mathbf{r}(\mathbf{x}) - \mathbf{r}(\hat{\mathbf{x}}_i)\|$ that converge geometrically until they reach roughly the same accuracy as Householder QR.



$\|\boldsymbol{r}(\boldsymbol{x})\| = 10^{-12}$
Forward error $\|\mathbf{x} - \hat{\mathbf{x}}_i\|/\|\mathbf{x}\|$
Backward error $\mathrm{BE}(x_i)$
Iteration i
$\|\boldsymbol{r}(\boldsymbol{x})\| = 10^{-3}$
$\kappa = 10^{15}$
$\kappa = 10^{10}$
$\kappa = 10^{1}$

Convergence – Theorem V

Let $\mathbf{S}$ be a subspace embedding with distortion $0 < \epsilon < 1 - \frac{1}{\sqrt{2}}$. Then the $\mathbf{x}_i$ satisfy the following bounds,

(1) $||\mathbf{x} - \mathbf{x}_i|| < (8 - 2\sqrt{2})\sqrt{\epsilon}\kappa g_{IS}^i \frac{||\mathbf{r}(\mathbf{x})||}{||\mathbf{A}||}$

(2) $||\mathbf{r}(\mathbf{x}) - \mathbf{r}(\mathbf{x}_i)|| < (8 - 2\sqrt{2})\sqrt{\epsilon} g_{IS}^i ||\mathbf{r}(\mathbf{x})||$

The convergence rate g_{IS} is,

$$g_{IS} = \frac{\epsilon(2-\epsilon)}{(1-\epsilon)^2} \leq (2+\sqrt{2})\epsilon$$

Lemma: Singular Value Bounds

Let $\mathbf{S}$ be a subspace embedding for range($\mathbf{A}$) with distortion $\epsilon \in (0,1)$ and $\mathbf{QR} = \mathbf{SA}$ be a reduced `QR` decomposition of $\mathbf{SA}$. Then $\mathbf{R}$ satisfies the bounds

$$\sigma_{\max}(\mathbf{R}) \le (1+\epsilon)\sigma_{\max}(\mathbf{A})$$
$$\sigma_{\min}(\mathbf{R}) \ge (1-\epsilon)\sigma_{\min}(\mathbf{A})$$

In addition, $\mathbf{AR}^{-1}$ satisfies the bounds

$$\sigma_{\max}(\mathbf{AR}^{-1}) \le \frac{1}{1-\epsilon}$$
$$\sigma_{\min}(\mathbf{AR}^{-1}) \ge \frac{1}{1+\epsilon}$$

Proof of Lemma

$$\sigma_{\max}(\mathbf{R}) \leq (1+\epsilon)\sigma_{\max}(\mathbf{A})$$

Recall, $\mathbf{SA} = \mathbf{QR}$ and $\mathbf{S}$ is a subspace embedding for subspace V such that $(1-\epsilon)\|\mathbf{v}\| \leq \|\mathbf{Sv}\| \leq (1+\epsilon)\|\mathbf{v}\|$ for $\epsilon \in (0,1)$ and $\mathbf{v} \in V$.

As $\mathbf{Q}$ unitary, $\mathbf{QR}$ and $\mathbf{R}$ have the same singular values. Establishing the upper bound,

$$\begin{aligned}
\sigma_{\max}(\mathbf{R}) &= \sigma_{\max}(\mathbf{QR}) = \sigma_{\max}(\mathbf{SA}) && (1)\\
\sigma_{\max}(\mathbf{SA}) &= \|\mathbf{SA}\| && (2)\\
&= \max_{\|\mathbf{v}\|=1} \|\mathbf{SAv}\| && (3)\\
&\leq (1+\epsilon) \max_{\|\mathbf{v}\|=1} \|\mathbf{Av}\| && (4)\\
&= (1+\epsilon)\sigma_{\max}(\mathbf{A}) && (5)\\
& && (6)
\end{aligned}$$

Proof cont.

$$\sigma_{\min}(\mathbf{R}) \geq (1-\epsilon)\sigma_{\min}(\mathbf{A})$$

Establishing the lower bound,

$$\begin{aligned}
\sigma_{\min}(\mathbf{R}) &= \sigma_{\min}(\mathbf{SA}) && (1)\\
&= \min_{\|\mathbf{v}\|=1} \|\mathbf{SAv}\| && (2)\\
&\geq (1-\epsilon) \min_{\|\mathbf{v}\|=1} \|\mathbf{Av}\| && (3)\\
&= (1-\epsilon)\sigma_{\min}(\mathbf{A}) && (4)
\end{aligned}$$

Proof cont.

$$\sigma_{\max}(\mathbf{AR}^{-1}) \leq \frac{1}{1-\epsilon}$$

For the bounds of $\mathbf{AR}^{-1}$, establishing the upper bounds

$$\begin{aligned}
\sigma_{\max}(\mathbf{AR}^{-1}) &= \max_{\|\mathbf{v}\|=1} \|\mathbf{AR}^{-1}\mathbf{v}\| \\
&\leq \frac{1}{1-\epsilon} \max_{\|\mathbf{v}\|=1} \|\mathbf{SAR}^{-1}\mathbf{v}\| \\
&= \frac{1}{1-\epsilon} \max_{\|\mathbf{v}\|=1} \|\mathbf{Qv}\| \\
&= \frac{1}{1-\epsilon} \sigma_{\max}(\mathbf{Q}) \\
&= \frac{1}{1-\epsilon}
\end{aligned}$$

Proof cont.

$$\sigma_{\min}(\mathbf{A}\mathbf{R}^{-1}) \geq \frac{1}{1+\epsilon}$$

Establishing the lower bound,

$$\begin{aligned}
\sigma_{\min}(\mathbf{A}\mathbf{R}^{-1}) &= \min_{\|\mathbf{v}\|=1} \|\mathbf{A}\mathbf{R}^{-1}\mathbf{v}\| \\
&\geq \frac{1}{1+\epsilon} \min_{\|\mathbf{v}\|=1} \|\mathbf{S}\mathbf{A}\mathbf{R}^{-1}\mathbf{v}\| \\
&= \frac{1}{1+\epsilon} \min_{\|\mathbf{v}\|=1} \|\mathbf{Q}\mathbf{v}\| \\
&= \frac{1}{1+\epsilon}\sigma_{\min}(\mathbf{Q}) \\
&= \frac{1}{1+\epsilon}
\end{aligned}$$

Lemma: Sketch-and-Solve

In the sketch-and-solve, compute

$$\mathbf{x}_0 = \arg\min_{\mathbf{y}\in\mathbb{R}^n} \|(\mathbf{SA})\mathbf{y} - \mathbf{Sb}\|$$

Computed numerically,

$$\mathbf{x}_0 = \mathbf{R}^{-1}(\mathbf{Q}^*(\mathbf{Sb}))$$

Then, we have the guarantee that

$$\|\mathbf{r}(\mathbf{x}_0)\| \leq \frac{1+\epsilon}{1-\epsilon}\|\mathbf{r}(\mathbf{x})\| \tag{1}$$

$$\|\mathbf{r}(\mathbf{x}) - \mathbf{r}(\mathbf{x}_0)\| \leq \frac{2\sqrt{\epsilon}}{1-\epsilon}\|\mathbf{r}(\mathbf{x})\| \tag{2}$$

$$\|\mathbf{x} - \mathbf{x}_0\| \leq \frac{2\sqrt{\epsilon}}{1-\epsilon}\frac{\kappa}{\|\mathbf{A}\|}\|\mathbf{r}(\mathbf{x})\| \tag{3}$$

Convergence Proof

(1) $\mathbf{x}_{i+1} = (\mathbf{I} - \mathbf{R}^{-1}\mathbf{R}^{-T}\mathbf{A}^T\mathbf{A})\mathbf{x}_i + \mathbf{R}^{-1}\mathbf{R}^{-T}\mathbf{A}^T\mathbf{b}$

(2) $\mathbf{x} = (\mathbf{I} - \mathbf{R}^{-1}\mathbf{R}^{-T}\mathbf{A}^T\mathbf{A})\mathbf{x} + \mathbf{R}^{-1}\mathbf{R}^{-T}\mathbf{A}^T\mathbf{b}$

(3) $\mathbf{x} - \mathbf{x}_{i+1} = \mathbf{R}^{-1}\mathbf{G}\mathbf{R}(\mathbf{x} - \mathbf{x}_i)$ where $\mathbf{G} := \mathbf{I} - \mathbf{R}^{-T}\mathbf{A}^T\mathbf{A}\mathbf{R}^{-1}$

(4) $\mathbf{x} - \mathbf{x}_i = \mathbf{R}^{-1}\mathbf{G}^i\mathbf{R}(\mathbf{x} - \mathbf{x}_0)$

(5) $||\mathbf{SA}(\mathbf{x} - \mathbf{x}_i)|| \leq ||\mathbf{Q}||||\mathbf{G}||^i||\mathbf{R}(\mathbf{x} - \mathbf{x}_0)|| = ||\mathbf{G}||^i||\mathbf{SA}(\mathbf{x} - \mathbf{x}_0)||$

(6) $||\mathbf{G}|| \leq \max\left\{\left(\frac{1}{1-\epsilon}\right)^2 - 1, 1 - \left(\frac{1}{1+\epsilon}\right)^2\right\} = \frac{\epsilon(2-\epsilon)}{(1-\epsilon)^2} = g_{IS}$

(7) $||\mathbf{SA}(\mathbf{x} - \mathbf{x}_0)|| \leq (1+\epsilon)||\mathbf{A}(\mathbf{x} - \mathbf{x}_0)|| = (1+\epsilon)||\mathbf{r}(\mathbf{x}) - \mathbf{r}(\mathbf{x}_0)|| \leq 2\frac{1+\epsilon}{1-\epsilon}\sqrt{\epsilon}||\mathbf{r}(\mathbf{x})||$

(8) $||\mathbf{r}(x) - \mathbf{r}(\mathbf{x}_i)|| \leq \frac{1}{1-\epsilon}||\mathbf{SA}(\mathbf{x} - \mathbf{x}_i)|| \leq \frac{1}{1-\epsilon}||\mathbf{G}||^i||\mathbf{SA}(\mathbf{x} - \mathbf{x}_0)|| \leq 2\frac{1+\epsilon}{(1-\epsilon)^2}\sqrt{\epsilon}g_{IS}^i||\mathbf{r}(\mathbf{x})||$

(9) $||\mathbf{r}(\mathbf{x}) - \mathbf{r}(\mathbf{x}_i)|| = ||\mathbf{A}(\mathbf{x} - \mathbf{x}_i)|| \geq \sigma_{min}(\mathbf{A})||\mathbf{x} - \mathbf{x}_i||$

Thanks!