Random projections for Quadratic Programming over a Euclidean ball

Leo Liberti, CNRS LIX Ecole Polytechnique

joint work with P.L. Poirion, Vu Khac Ky and C. D'Ambrosio

Aussois COW 2019

Random projections: the gist

- **Let X be a $n \times p$ data matrix**
- **P is $d \times n$ short & fat, normally sampled componentwise**

$$\underbrace{\begin{pmatrix} & \cdots & \end{pmatrix}}_{P} \underbrace{\begin{pmatrix} \vdots & \vdots & \vdots \end{pmatrix}}_{A} = \underbrace{\begin{pmatrix} & \cdots & \end{pmatrix}}_{PA}$$

- **Then $\forall i < j \; \|X_i - X_j\|_2 \approx \|PX_i - PX_j\|_2$ "wahp"**
  X_i, X_j: i, j-th columns of A
- **P is a *random projection* (RP) matrix**
  there are also sparse versions of RPs

wahp

"wahp" = "with arbitrarily high probability"

the probability of E_d
(depending on some parameter d)
approaches 1 *"exponentially fast"* as d increases

$$\mathbf{P}(E_d) \approx 1 - O(e^{-d})$$

Variant "with overwhelming probability" gives "wop", not adopted as used to insult Italians in UK

Johnson-Lindenstrauss Lemma (JLL)

Thm.

Given $X \subseteq \mathbb{R}^n$ with $|X| = p$ and $\varepsilon > 0$ there is $d \sim O(\frac{1}{\varepsilon^2} \ln p)$ and a $d \times n$ matrix P s.t.

$$\forall x, y \in X \quad (1-\varepsilon)\|x-y\|_2 \;\leq\; \|Px - Py\|_2 \;\leq\; (1+\varepsilon)\|x-y\|_2$$

If P is sampled componentwise from $\mathsf{N}(0, \frac{1}{\sqrt{d}})$, then

$$\mathbf{P}(X \text{ and } PX \text{ are approximately congruent}) \geq \frac{1}{p}$$

result follows by probabilistic method

Henceforth, all norms will be Euclidean unless stated otherwise

What we can achieve

- Assume $Ax \leq b$ is $m \times n$ full-dimensional
- Consider ball-constrained QP:

$$\max_{x \in \mathbb{R}^n} \{x^\top Q x + cx \mid Ax \leq b \wedge \|x\| \leq 1\} \quad [\mathsf{R}]$$

- RP-based approximating reformulation:
  $\bar{Q} = PQP^\top, \bar{c} = cP^\top, \bar{A} = AP^\top$

$$\max_{u \in \mathbb{R}^d} \{u^\top \bar{Q} u + \bar{c}u \mid \bar{A}u \leq b \wedge \|u\| \leq 1\} \quad [P\mathsf{R}]$$

  with $d \sim O(\ln m)$

- Approximate optimum of [R] can be retrieved from opt. of [PR]

Approximating reformulation

$$\left.\begin{array}{rrcl} \min\limits_{x\in\mathbb{R}^n} & (x^\top P^\top)PQP^\top(Px) + c^\top P^\top(Px) & & \\ & AP^\top(Px) & \le & b \\ & \|Px\| & \le & 1 \end{array}\right\}$$

Set $u = Px$

$$\left.\begin{array}{rrcl} \min\limits_{u\in\mathbb{R}^d} & u^\top PQP^\top u + c^\top P^\top u & & \\ & AP^\top u & \le & b \\ & \|u\| & \le & 1 \end{array}\right\}$$

recall $\bar{Q} = PQP^\top, \bar{c} = c^\top P^\top, \bar{A} = AP^\top$

$$\left.\begin{array}{rrcl} \min\limits_{u\in\mathbb{R}^d} & u^\top \bar{Q}u + \bar{c}^\top u & & \\ & \bar{A}u & \le & b \\ & \|u\| & \le & 1 \end{array}\right\} [P\mathsf{R}]$$

$P\mathsf{R}$ much smaller than R

Why should $\mathsf{R} \approx P\mathsf{R}$?

Intuitive (and informal) idea:

- P randomly sampled componentwise from $\mathsf{N}(0, \frac{1}{\sqrt{d}})$
- $\Rightarrow$ columns of P almost orthogonal
  see e.g. [Vershynin 2018]
- $\Rightarrow P^\top P$ approximately an identity I_n
- We have:

$$\begin{aligned} & x^\top (P^\top P) Q (P^\top P) x + c^\top (P^\top P) x \\ \approx \quad & x^\top I_n Q I_n x + c^\top I_n x \\ = \quad & x^\top Q x + c^\top x \end{aligned}$$

  and

$$A(P^\top P)x \approx A I_n x = Ax$$

- *This is <u>not</u> how the proof works*

Solution retrieval

Intuitive (and informal) idea:

- Solve $P\mathsf{R}$, get optimal solution u^*
- Recall $u = Px$ by approximating reformulation
- $P^\top P \approx I_n$ shows $P^\top$ "plays the role of P^{-1}"
- Let $x^* = P^\top u^*$
- *This can be made precise*

The approximation theorem

Consider:

$$(\mathsf{QP}^-_\varepsilon) \qquad \max\{u^\top \bar{Q} u + \bar{c}^\top u \mid \bar{A}u \le b, \qquad \|u\| \le 1-\varepsilon, u \in \mathbb{R}^d\}$$

$$(\mathsf{QP}^+_\varepsilon) \qquad \max\{u^\top \bar{Q} u + \bar{c}^\top u \mid \bar{A}u \le b+\varepsilon, \ \|u\| \le 1+\varepsilon, u \in \mathbb{R}^d\}$$

Thm.

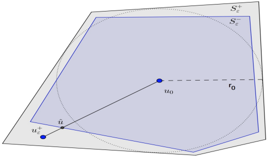
Let x^* be an opt. of original QP and $x^-_\varepsilon, x^+_\varepsilon$ of $\mathsf{QP}^-_\varepsilon, \mathsf{QP}^+_\varepsilon$. Then for d in $O((\ln m)/\varepsilon^2)$

$$f(x^*) \quad \in \quad [f(x^-_\varepsilon)\,,\ f(x^+_\varepsilon)]$$

wahp. Moreover,

$$f(x^+_\varepsilon) - f(x^-_\varepsilon) \quad \text{is} \quad O\left(\frac{\varepsilon\|c\|}{\text{“fullness” of polyhedron}}\right)$$

where $f(y) = y^\top \bar{Q} y + c^\top y$



The proof toolkit: additive distortions

- **Given** $\varepsilon \in (0,1)$
- $\exists$ **constant** $\mathcal{C}_0$ **s.t.**

1. $\forall x, y \in \mathbb{R}^n$

$$(Px)^\top Py \quad \in \quad x^\top y \pm \varepsilon \|x\| \, \|y\|$$

with prob. $\geq 1 - 4e^{-\mathcal{C}_0 \varepsilon^2 d}$

2. $\forall x \in \mathbb{R}^n$

$$AP^\top Px \quad \in \quad Ax \pm \varepsilon \|x\| \mathbf{1}$$

with prob. $\geq 1 - 4me^{-\mathcal{C}_0 \varepsilon^2 d}$

3. $\forall x, y \in \mathbb{R}^n$

$$x^\top P^\top PQP^\top Py \quad \in \quad x^\top Qy \pm 3\varepsilon \|x\| \, \|y\| \, \|Q\|_*$$

with prob. $\geq 1 - 8ke^{-\mathcal{C}_0 \varepsilon^2 d}$

$\|Q\|_*$: **nuclear norm,** $k = \mathsf{rank}(Q)$

The proof toolkit: additive distortions

- **Given** $\varepsilon \in (0,1)$
- $\exists$ **constant** $\mathcal{C}_0$ **s.t.**

1. $\forall x, y \in \mathbb{R}^n$
$$x^\top y - \varepsilon \|x\| \, \|y\| \le (Px)^\top Py \le x^\top y + \varepsilon \|x\| \, \|y\|$$
with prob. $\ge 1 - 4e^{-\mathcal{C}_0 \varepsilon^2 d}$
2. $\forall x \in \mathbb{R}^n$
$$Ax - \varepsilon \|x\| \mathbf{1} \le AP^\top Px \le Ax + \varepsilon \|x\| \mathbf{1}$$
with prob. $\ge 1 - 4me^{-\mathcal{C}_0 \varepsilon^2 d}$
3. $\forall x, y \in \mathbb{R}^n$
$$x^\top Qy - 3\varepsilon \|x\| \, \|y\| \, \|Q\|_* \le x^\top P^\top PQP^\top Py \le x^\top Qy + 3\varepsilon \|x\| \, \|y\| \, \|Q\|_*$$
with prob. $\ge 1 - 8ke^{-\mathcal{C}_0 \varepsilon^2 d}$
$\|Q\|_*$: **nuclear norm,** $k = \mathsf{rank}(Q)$

The proof toolkit: Lemma 1

Lemma

There is a $d \times n$ RP P, an $\varepsilon \in (0,1)$ and a constant $\mathcal{C}_0$ s.t.
$\forall x, y \in \mathbb{R}^n \; (Px)^\top Py \in x^\top y \pm \varepsilon \|x\| \, \|y\|$ with prob. $\geq 1 - 4e^{-\mathcal{C}_0 \varepsilon^2 d}$

Proof

By JLL,

$$\mathsf{Prob}\left[(1-\varepsilon)\|x\|^2 \leq \|Px\|^2 \leq (1+\varepsilon)\|x\|^2\right] \geq 1 - 2e^{-\mathcal{C}\varepsilon^2 d}$$

Apply to any two vectors $u+v, u-v$ and use union bound, get

$$\begin{aligned} |(Pu)^\top (Pv) - u^\top v| &= \frac{1}{4}\left| \|P(u+v)\|^2 - \|P(u-v)\|^2 - \|u+v\|^2 + \|u-v\|^2 \right| \\ &\leq \frac{1}{4}\left| \|P(u+v)\|^2 - \|u+v\|^2 \right| + \frac{1}{4}\left| \|P(u-v)\|^2 - \|u-v\|^2 \right| \\ &\leq \frac{\varepsilon}{4}(\|u+v\|^2 + \|u-v\|^2) = \frac{\varepsilon}{2}(\|u\|^2 + \|v\|^2), \end{aligned}$$

with prob. $\geq 1 - 4e^{-\mathcal{C}_0 \varepsilon^2 d}$; apply to $u = \frac{x}{\|x\|}$ and $v = \frac{y}{\|y\|}$

The proof toolkit: Lemma 2

Lemma

There is a $d \times n$ RP P, an $\varepsilon \in (0,1)$ and a constant $\mathcal{C}_0$ s.t. $\forall x \in \mathbb{R}^n$ and $m \times n$ matrix A with unit row vectors, $AP^\top Px \in Ax \pm \varepsilon \|x\| \mathbf{1}$ with prob. $\geq 1 - 4me^{-\mathcal{C}_0 \varepsilon^2 d}$

Proof

Let $A_1, \ldots, A_m$ be the unit row vectors of A

$$AP^\top Px - Ax = \begin{pmatrix} A_1^\top P^\top Px - A_1^\top x \\ \ldots \\ A_m^\top P^\top Px - A_m^\top x \end{pmatrix} = \begin{pmatrix} (PA_1)^\top Px - {A_1}^\top x \\ \ldots \\ (PA_m)^\top Px - {A_m}^\top x \end{pmatrix}$$

Apply Lemma 1 and union bound

The proof toolkit: Lemma 3

Lemma

There is a $d \times n$ RP P, an $\varepsilon \in (0,1)$ and a constant $\mathcal{C}_0$ s.t. $\forall x, y \in \mathbb{R}^n$ we have $x^\top P^\top PQP^\top Py \in x^\top Qy \pm 3\varepsilon \|x\|\|y\|\|Q\|_*$
with prob. $\geq 1 - 8ke^{-\mathcal{C}_0 \varepsilon^2 d}$

Proof

(Sketch) Let $U\Sigma V^\top$ be SVD decomp. of Q, then

$$\begin{aligned} x^\top P^\top PQP^\top Py &= (U^\top P^\top Px)^\top \Sigma (V^\top P^\top Py) \\ &= \left[U^\top x + U^\top (P^\top P - I_n)x\right]^\top \Sigma \left[V^\top y + V^\top (P^\top P - I_n)y\right] \end{aligned}$$

Apply Lemma 2 and union bound, then

$$\begin{aligned} (U^\top x - \varepsilon\|x\|\mathbf{1}_k)^\top \, \Sigma \, (V^\top y - \varepsilon\|y\|\mathbf{1}_k) &\leq x^\top P^\top PQP^\top Py \\ (U^\top x + \varepsilon\|x\|\mathbf{1}_k)^\top \, \Sigma \, (V^\top y + \varepsilon\|y\|\mathbf{1}_k) &\geq x^\top P^\top PQP^\top Py \end{aligned}$$

with prob. $\geq 1 - 8ke^{-\mathcal{C}\varepsilon^2 d}$

Lots of elementary algebra and rearrangements later, you get the result

Approximation theorem: proof by epigraph

> Despite its triviality, the union bound is probably the most useful fact in all of theoretical computer science. I use it maybe 200 times in every paper I write.
>
> Scott Aaronson

Computational results

- **All validation on convex QPs to save on testing time**
  otherwise wait forever for sBB termination on nonconvex QPs...
- **Random projection matrices**
  - $\varepsilon \in \{0.1, 0.15, 0.2\}$
  - **density in** $\{0.2, 0.5, 1.0\}$
- **All random instances for**
  - $m \in \{10, 100, 1000\}$
  - $n \in \{2000, 3000\}$
  - **coefficients sampled from** $\mathsf{U}(0,1)$ **and** $\mathsf{U}(-1,1)$
  - A, Q **density in** $\{0.1, 0.6\}$
- **2 large portfolio instances from Kaggle datasets**

$$\max\{\mu^\top x - \lambda\, x^\top \Sigma x \mid \sum_{j \le n} x_j \le 1 \wedge -x \le 0 \wedge \|x\| \le 1\}$$

  λ **is a scalarization constant**
  sizes are $n \in \{1344, 7163\}$, $\mathsf{nnz}(\Sigma) \in \{1\mathrm{M}, 25\mathrm{M}\}$ **(fully dense** Q**)**
- **Julia+JuMP with IPOPT solver**
  on 4-CPU Xeon @2.1GHz 64GB RAM (each CPU has 8 cores)

Computational results

- Measures:
  - CPU time
  - Optimality: $\rho = |f^*_{\mathsf{org}} - f^*_{\mathsf{retr}}| / \max(f^*_{\mathsf{org}}, f^*_{\mathsf{retr}})$
  - Feasibility: avg feas. errors (ace,are) for $(Ax \leq b, \|x\| \leq 1)$
- Statistics over all random instance runs with all RPs:

	CPU_{org}	CPU_{proj}	ρ	ace	are
mean	37.691	**14.590**	0.103	0.0	0.0
stdev	49.984	**15.057**	0.070	0.0	0.0
min	8.750	**2.170**	0.000	0.0	0.0
max	198.350	**61.340**	0.485	0.0	0.0

- The two large portfolio instances

Instance	CPU_{org}	CPU_{proj}	ρ	ace	are
`etfs`	534.32	**11.38**	0.270	0.026	0.000
`stocks`	266713.40	**132.78**	0.007	0.023	0.000

Relationship with RP applied to LP

- **RP applied to LP in standard form [Vu et al. MOR 2018]:**

$$\min\{c^\top x \mid Ax = b \wedge x \geq 0\}$$

- **Replace $Ax = b$ by $PAx = Pb$**

$$\min\{c^\top x \mid PAx = Pb \wedge x \geq 0\}$$

- **Approximate LP with $O(\ln n)$ constraints**

Relationship with RP applied to LP

- **Consider LP in canonical form**

$$\max\{c^\top x \mid Ax \leq b\}$$

- **Assume bounded:** $\|x^*\| \leq K$
  scale problem data so $\|x^*\| \leq 1$
- **All QP results go through if** $Q = 0$:

$$\max\{(Pc)^\top u \mid AP^\top u \leq b\}$$

- **Approximate LP with** $O(\ln m)$ **variables**

Relationship with RP applied to LP

- **RP applied to LP in standard form:**

$$\min\{c^\top x \mid PAx = Pb \wedge x \geq 0\}$$

- **Take dual:**

$$\min\{(Pb)^\top u \mid (PA)^\top u \leq c\}$$

- **After renaming $c \leftrightarrow b$:**

$$\max\{(Pc)^\top u \mid AP^\top u \leq b\}$$

- **Same as RP applied to QP with $Q = 0$**

Projecting (some) LP rows *and* columns!

- Consider LP

$$\max\{c^\top x \mid Ax \le b \wedge Dx = d \wedge x \ge 0\}$$

A is $m \times n$, D is $p \times n$

- LP results hold if RP applied to subset of rows [Vu et al. MOR 2018]
- P a $d \times n$ RP, T a $k \times p$ RP
- Obtain

$$\max\{(Pc)^\top u \mid AP^\top u \le b \wedge TDx = Td \wedge x \ge 0 \wedge u = Px\}$$

- Snag: increases number of vars to $n + O(\ln m)$
- Nice: no need for retrieval (x in formulation)
- Computationally untested *I only thought of this a few days ago*
- Can't apply to QP: $Dx = d$ negates full-dim. assumption

Some references

- K. Vu, P.-L. Poirion, L. Liberti, *Random projections for linear programming*, Math. of OR, online 2018
- K. Vu, P.-L. Poirion, L. Liberti, *Gaussian random projections for Euclidean membership problems*, DAM, online 2018
- L. Liberti, P.-L. Poirion, K. Vu, *Solving LP using random projections*, ENDM **55**:53-56, 2016
- W. Johnson and J. Lindenstrauss, *Extensions of Lipschitz mappings into a Hilbert space*, in Contemporary Mathematics, Vol. 26, 189-206, 1984
- R. Vershynin, High-Dimensional Probability, CUP 2018.