A note on reductions between compressed sensing guarantees

Tom Morgan* Jelani Nelson†

June 3, 2016

Abstract

In *compressed sensing*, one wishes to acquire an approximately sparse high-dimensional signal $x \in \mathbb{R}^n$ via $m \ll n$ noisy linear measurements, then later approximately recover x given only those measurement outcomes. Various guarantees have been studied in terms of the notion of approximation in recovery, and some isolated folklore results are known stating that some forms of recovery are stronger than others, via black-box reductions. In this note we provide a general theorem concerning the hierarchy of strengths of various recovery guarantees. As a corollary of this theorem, by reducing from well-known results in the compressed sensing literature, we obtain an efficient ℓ_p/ℓ_p scheme for any $0 < p < 1$ with the fewest number of measurements currently known amongst efficient schemes, improving recent bounds of [SY16].

1 Introduction

The field of *compressed sensing* [CT05, Don06] is concerned with recovering approximately sparse signals from few (possibly noisy) linear measurements. That is, given access to $y = \Phi x + e$, where $\Phi \in \mathbb{R}^{m\times n}$ is some matrix and $x, e \in \mathbb{R}^n$ has e of small norm and x being approximately sparse. That is, x can be decomposed as $x = f + e'$ where f is sparse, i.e. $k \stackrel{\text{def}}{=} \|f\|_0$ is small, and e' has small norm. Here $\|\cdot\|_0$ denotes support size. Ideally we would like $m \ll n$ (i.e. few measurements), and that there is an efficient algorithm $\mathcal{R}$ which, knowing Φ and given only access to y but not x, recovers some $\hat{x} = \mathcal{R}(y)$ such that $x - \hat{x}$ has small norm (in terms of the norms of e, e'). That is, for some norms $\|\cdot\|_X, \|\cdot\|_Y$ and some constants $C > 0$ and $c \geq 0$ depending only X, Y, we would like

$$\|x - \mathcal{R}(y)\|_X \leq Ck^{-c} \cdot \sigma_k(x)_Y, \tag{1}$$

where $\sigma_k(x)_Y = \inf_{\|z\|_0 \leq k} \|x - z\|_Y$ is the $\|\cdot\|_Y$-norm error of the best k-sparse approximation to x. Popular guarantees investigated previously include so-called ℓ_p/ℓ_q guarantees, of the form

$$\|x - \mathcal{R}(y)\|_{\ell_p^n} \leq Ck^{-c} \cdot \sigma_k(x)_{\ell_q^n} \tag{2}$$

where the constant c equals $1/q - 1/p$ [CDD08]. For the case $p = q$, one can often even take $C = 1 + \varepsilon$ and develop schemes with small m and efficient $\mathcal{R}$, in which a dependence on ε enters m (see e.g. [IR08] for $p = q = 1$). In all the cases of $1 \leq q \leq p < \infty$, it is known that any scheme achieving (2) must have $m \gtrsim k \log(n/k)$ (and sometimes much larger, depending on p, q), regardless of how inefficient $\mathcal{R}$ is allowed to be. This is due to a connection with known bounds on Gelfand widths [CDD08].

The case $p = q < 1$ was first investigated in [CS08, SL10] (although ℓ_p is not a norm in this case), where it was shown that there exists *some* recovery scheme $\mathcal{R}$ (though not efficient) that allows for $m \simeq C_1(p)k + C_2(p)k\log(n/k)$ measurements for some $C_2(p) \to 0$ as $p \to 0$. In other words, for sufficiently small $p \ll 1$, there exists an upper bound violating the lower bound for ℓ_q/ℓ_q norm guarantees for $q \geq 1$. For the

*Harvard University. tdmorgan@seas.harvard.edu. Supported by NSF grants CCF-1320231 and CCF-0915922.

†Harvard University. minilek@seas.harvard.edu. Supported by NSF grant IIS-1447471 and CAREER award CCF-1350670, ONR Young Investigator award N00014-15-1-2388, and a Google Faculty Research Award.

case $p = 1/2^j$ where j is any positive integer, and for any fixed constant $\epsilon > 0$, recent work of [SY16] provides a scheme with an *efficient* (polynomial-time in n) $\mathcal{R}$ for achieving a slight weakening of the ℓ_p/ℓ_p recovery guarantee, but with a larger number of measurements $m \simeq k^{2/p} \log n$. Specifically, their error guarantee is

$$\|x - \mathcal{R}(y)\|_{\ell_p^n} \leq C\sigma_k(x)_{\ell_q^n} + \epsilon \cdot \|x\|_\infty$$

It is folklore that some norm guarantees are stronger than others. By stronger, we mean a reduction in the sense that if the scheme $(\Phi, \mathcal{R}_1)$ achieves the $\|\cdot\|_{X_1}/\|\cdot\|_{Y_1}$ recovery guarantee, then there is an efficient algorithm $\mathcal{A}$ such that if $\mathcal{R}_2(y)$ is simply set to be $\mathcal{A}(\mathcal{R}_1(y), y)$, then $(\Phi, \mathcal{R}_2)$ achieves the $\|\cdot\|_{X_2}/\|\cdot\|_{Y_2}$ guarantee (sometimes in these reductions, the values of k and C may change by constant factors). In this sense, achieving the $\|\cdot\|_{X_1}/\|\cdot\|_{Y_1}$ guarantee is *stronger* than achieving the $\|\cdot\|_{X_2}/\|\cdot\|_{Y_2}$ guarantee (since devising an efficient scheme for the former implies an efficient scheme for the latter). It is folklore that, for example, an ℓ_2/ℓ_2 scheme is stronger than ℓ_2/ℓ_1, which in turn is stronger than ℓ_1/ℓ_1 (note though it is impossible to achieve the ℓ_2/ℓ_2 guarantee without weakening to a certain probabilistic guarantee [CDD08], i.e. *nonuniformity* — we discuss nonuniformity in Section 1.1). The main observation of this note is a common generalization of both of these folklore reductions. In particular, we show the following main theorem.

Theorem 1. *Suppose that either $p, 1 \geq r \geq s > 0$ and $q \geq s$, or $p > 0$ and $q \geq s > 0$. Then if there exists a recovery scheme $(\Phi, \mathcal{R}')$ s.t. for some constant C' independent of k, n*

$$\|x - \mathcal{R}'(y)\|_p \leq C' \cdot k^{\frac{1}{p} - \frac{1}{q}} \sigma_{2k}(x)_q,$$

then there exists a recovery scheme $(\Phi, \mathcal{R})$ such that for some constant C

$$\|x - \mathcal{R}(y)\|_r \leq C \cdot k^{\frac{1}{r} - \frac{1}{s}} \sigma_k(x)_s.$$

Furthermore, if $\mathcal{R}'$ runs in time T and outputs a vector $\hat{x}$ of support size S, then $\mathcal{R}$ runs in time $O(T + S)$.

The above theorem is implied by Corollary 7 and Theorem 8. The reduction of Theorem 1 is very simple: in the second case for p, q, s, $\mathcal{R}$ is simply $\mathcal{R}'$. In the first case for p, q, r, s, the recovery algorithm $\mathcal{R}$, given Φx, first computes $\mathcal{R}'(\Phi x)$. We then let $\mathcal{R}(\Phi x)$ be defined by projecting $\mathcal{R}'(\Phi x)$ to its largest $2k$ coordinates (in magnitude). This scheme is analyzed in Section 2 (see Theorem 8). It is worth noting that these are the same simple schemes that were previously known in folklore to reduce ℓ_2/ℓ_1 to ℓ_2/ℓ_2, and ℓ_1/ℓ_1 to ℓ_2/ℓ_1.

It is known how to achieve the ℓ_2/ℓ_1 guarantee with $m \simeq k \log(n/k)$ measurements and a $poly(n)$-time recovery algorithm $\mathcal{R}$ [Can08] (more accurately, there exists a deterministic algorithm $\mathcal{R}$ such that if Φ is drawn at random from a particular distribution with that number m of measurements, then with probability $1 - poly(1/\binom{n}{k})$, $(\Phi, \mathcal{R})$ is a scheme achieving the ℓ_2/ℓ_1 recovery guarantee). Thus by setting $r = s = p \leq 1$ in Theorem 1, we obtain the following corollary.

Corollary 2. *For any $0 < p \leq 1$, there exists a scheme achieving the ℓ_p/ℓ_p recovery guarantee with $m \lesssim k \log(n/k)$ measurements and recovery time $poly(n)$.*

We point out that although Corollary 2 does not achieve the $C_1(p)k + C_2(p)k \log(n/k)$ measurements for $C_2(p) \to 0$ that was shown achievable (with an inefficient recovery algorithm) in [CS08, SL10], it achieves a number of measurements that is much less than the $m \simeq k^{2/p} \log n$ of [SY16], even for $p = 1$.

1.1 Nonuniform guarantees

Up until this point we have only discussed *uniform* recovery guarantees (this terminology appears in e.g. [FR13]). A uniform scheme is one where Eq. (1) holds for all $x \in \mathbb{R}^n$ simultaneously, for single pair $(\Phi, \mathcal{R})$. Indeed many such schemes pick Φ randomly, but then the desired guarantee for uniformity is

$$\Pr_{\Phi}(\forall x \in \mathbb{R}^n,\ (1) \text{ holds}) \geq 1 - \delta.$$

A *nonuniform* recovery guarantee is one where, in a randomized scheme (in which Φ, and possibly also $\mathcal{R}$, are chosen at random from some distribution)

$$\forall x \in \mathbb{R}^n \Pr_{\Phi,\mathcal{R}}((1) \text{ holds}) \geq 1-\delta.$$

Uniform and nonuniform schemes are also sometimes called "for all" and "for each" schemes in the literature. It is known for example that the ℓ_2/ℓ_2 guarantee cannot be achieved uniformly by *any* (even possibly inefficient) recovery algorithm unless $m \gtrsim n$ [CDD08], however it is achievable by a nonuniform scheme with failure probability $1/poly(n)$, $m \simeq k\log n$, recovery time $T \lesssim n\log n$, and output sparsity $S = \|\hat{x}\|_0 \lesssim k$ by combining the CountSketch [CCFC04] with a reduction from ℓ_2/ℓ_2 recovery to the ℓ_2 heavy hitters problem [CM06] (see also [GI10, Section II.B]). One could also replace the CountSketch with the more efficient ExpanderSketch to keep all parameters the same while reducing the recovery time to $T \lesssim k\log^c n$. We note that to achieve $C = 1+\varepsilon$, the work of [GLPS10] achieves a better bound on m than the ExpanderSketch in terms of ε by a factor of $1/\varepsilon$, albeit with a worse value of S and only constant failure probability. However, since Theorem 1 only preserves C up to a constant factor, it cannot be used to convert a scheme with $C = 1+O(\varepsilon)$ for one recovery guarantee into a scheme with $C = 1+\varepsilon$ for another guarantee.

We now mention that Theorem 1 holds regardless of whether $(\Phi, \mathcal{R}')$ is uniform or nonuniform scheme, and the reduction is uniformity-preserving. Thus by combining the observations of the last paragraph with Theorem 1, we obtain the following corollary.

Corollary 3. *For any $0 < p \leq 1$, there exists a nonuniform scheme achieving the ℓ_p/ℓ_p recovery guarantee with $m \lesssim k\log n$ measurements, recovery time $k \cdot poly(\log n)$, and failure probability $1/poly(n)$.*

2 Main result

Given $v \in \mathbb{R}^n$ and $S \subseteq [n]$ we will use v_S to denote the vector where $(v_S)_i = v_i$ if $i \in S$ and 0 otherwise.

Fact 4. *For any $v \in \mathbb{R}^n$ and $0 < a \leq b$, $\|v\|_b \leq \|v\|_a$.*

Proof. Observe that for all $i \in [n]$, $\frac{|v_i|}{\|v_i\|_b} \leq 1$. This, together with the fact that $a \leq b$, gives us

$$\left(\frac{|v_i|}{\|v\|_b}\right)^b \leq \left(\frac{|v_i|}{\|v\|_b}\right)^a. \tag{3}$$

We now have

$$\begin{aligned}
\|v\|_a &= \left(\sum_{i=1}^n |v_i|^a\right)^{\frac{1}{a}} \\
&\geq \left(\sum_{i=1}^n \left(\frac{|v_i|}{\|v\|_b}\right)^b \|v\|_b^a\right)^{\frac{1}{a}} \quad \text{(Equation 3)} \\
&= \|v\|_b^{1-\frac{b}{a}} \left(\sum_{i=1}^n |v_i|^b\right)^{\frac{1}{a}} \\
&= \|v\|_b.
\end{aligned}$$

□

Fact 5. *For any $v \in \mathbb{R}^n$ and $0 < a \leq b$, $\|v\|_a \leq n^{\frac{1}{a}-\frac{1}{b}}\|v\|_b$. In particular, if v is κ-sparse then $\|v\|_a \leq \kappa^{\frac{1}{a}-\frac{1}{b}}\|v\|_b$.*

Proof. Hölder's inequality states that for $p \geq 1$,

$$\|fg\|_1 \leq \|f\|_p \|g\|_{\frac{p}{p-1}}.$$

We will choose $f, g \in \mathbb{R}^n$ such that for all $i \in [n]$, $|f_i| = |v_i|^a$ and $|g_i| = 1$. Letting $p = b/a$, we then have

$$\|v\|_a^a = \|fg\|_1 \leq \|f\|_{\frac{b}{a}} \|g\|_{\frac{b}{b-a}} = \|v\|_b^a \cdot n^{1-\frac{a}{b}},$$

and thus

$$\|v\|_a \leq \|v\|_b \cdot n^{\frac{1}{a}-\frac{1}{b}}$$

as desired. □

The following lemma is via a common technique in compressed sensing that has come to be known as *shelling*, in which one sorts coordinates of a vector by magnitude of entries, blocks consecutive groups of coordinates together, then compares some norm of one group with some norm of the previous group.

Lemma 6. *For any $v \in \mathbb{R}^n$ and $0 < a \leq b$,*

$$\|v_{\mathrm{tail}(2\kappa)}\|_b \leq \kappa^{\frac{1}{b}-\frac{1}{a}} \|v_{\mathrm{tail}(\kappa)}\|_a.$$

Proof. Let $i_1, \ldots, i_n$ be a permutation of $[n]$ such that $|v_{i_1}| \leq |v_{i_2}| \leq \ldots \leq |v_{i_n}|$. For $j \geq 0$ an integer define $B_j = \{i_{jk+1}, i_{jk+2}, \ldots, i_{(j+1)k}\}$. Then $v_{tail(2\kappa)}$ is simply $v_{B\setminus(B_0\cup B_1)}$, so that

$$\|v_{tail(2\kappa)}\|_b = \left(\sum_{j\geq 2} \|v_{B_j}\|_b^b\right)^{\frac{1}{b}}. \tag{4}$$

Next observe that for any $i \in B_j$, $|v_i| \leq \|v_{B_{j-1}}\|_a / \kappa^{1/a}$. This follows since $\|v_{B_{j-1}}\|_a^a/\kappa$ is the average ath power of all $|v_{i'}|$ in B_{j-1}, and all terms in this average are at least as big as $|v_i|^a$. Combining with (4)

$$\begin{aligned}
\|v_{\mathrm{tail}(2\kappa)}\|_b &\leq \left(\sum_{j\geq 2} \frac{\kappa}{\kappa^{b/a}} \|v_{B_{j-1}}\|_a^b\right)^{\frac{1}{b}} \\
&= \kappa^{\frac{1}{b}-\frac{1}{a}} \cdot \left(\sum_{j\geq 1} \|v_{B_j}\|_a^b\right)^{\frac{1}{b}} \\
&\leq \kappa^{\frac{1}{b}-\frac{1}{a}} \cdot \left(\sum_{j\geq 1} \|v_{B_j}\|_a^a\right)^{\frac{1}{a}} \quad \text{(Fact 4)} \\
&= \kappa^{\frac{1}{b}-\frac{1}{a}} \cdot \|v_{\mathrm{tail}(\kappa)}\|_a.
\end{aligned}$$

□

This lemma immediately gives us the following reduction.

Corollary 7. *For $p > 0$, $q \geq s > 0$, if there exists a recovery scheme $(\Phi, \mathcal{R})$ such that for some constant $C_{p,q}$*

$$\|x - \mathcal{R}(\Phi x)\|_p \leq C_{p,q} \cdot k^{\frac{1}{p}-\frac{1}{q}} \|x_{\mathrm{tail}(2k)}\|_q,$$

then

$$\|x - \mathcal{R}(\Phi x)\|_p \leq C_{p,q} \cdot k^{\frac{1}{p}-\frac{1}{s}} \|x_{\mathrm{tail}(k)}\|_s.$$

Now we have all the tools to prove our main theorem.

Theorem 8. *For $p, 1 \geq r \geq s > 0$ and $q \geq s$, if there exists a recovery scheme $(\Phi, \mathcal{R}')$ such that for some constant $C'_{p,q}$*

$$\|x - \mathcal{R}'(\Phi x)\|_p \leq C'_{p,q} \cdot k^{\frac{1}{p}-\frac{1}{q}} \|x_{\text{tail}(2k)}\|_q,$$

then there exists a recovery scheme $(\Phi, \mathcal{R})$ such that for some constant $C_{p,q,r}$

$$\|x - \mathcal{R}(\Phi x)\|_r \leq C_{p,q,r} \cdot k^{\frac{1}{r}-\frac{1}{s}} \|x_{\text{tail}(k)}\|_s.$$

Furthermore, if $(\Phi, \mathcal{R}')$ achieves a uniform guarantee, then so does $(\Phi, \mathcal{R})$. Also, if $\mathcal{R}'$ runs in time T and outputs a vector of support size S, then $\mathcal{R}$ runs in time $O(T + S)$.

Proof. The recovery algorithm $\mathcal{R}$, given Φx, first computes $w = \mathcal{R}'(\Phi x)$. We then let $z = \mathcal{R}(\Phi x)$ be defined by projecting w to its largest $2k$ coordinates (in magnitude). It is clear that this reduction is uniformity-preserving, and it runs in time $O(T + S)$ since the $(2k)^{\text{th}}$ largest element of w (in magnitude) can be found in $O(S)$ time using the linear-time selection algorithm of [BFP+72].

We now analyze this scheme. Let $A \subseteq [n]$ be the largest $2k$ coordinates of x in magnitude, and $B \subseteq [n]$ be the largest $2k$ coordinates of w in magnitude (i.e. $z = w_B$). We will use the fact that ℓ_r for $r \leq 1$ is a quasinorm, with $d(f, g) := \|f - g\|_r^r$ satisfying the triangle inequality.

$$\begin{aligned}
\|x - z\|_r^r &\leq \|x_B - w_B\|_r^r + \|x - x_B\|_r^r \text{ (triangle inequality)} \\
&= \|x_B - w_B\|_r^r + \|x\|_r^r - \|x_B\|_r^r \\
&\leq \|x_B - w_B\|_r^r + \|x\|_r^r - (\|w_B\|_r^r - \|x_B - w_B\|_r^r) \text{ (triangle inequality)} \\
&= 2\|x_B - w_B\|_r^r + \|x\|_r^r - \|w_B\|_r^r \\
&\leq 2\|x_B - w_B\|_r^r + \|x\|_r^r - \|w_A\|_r^r \text{ (definition of } B\text{)} \\
&\leq 2\|x_B - w_B\|_r^r + \|x\|_r^r - (\|x_A\|_r^r - \|w_A - x_A\|_r^r) \text{ (triangle inequality)} \\
&= 2\|x_B - w_B\|_r^r + \|w_A - x_A\|_r^r + \|x_{\text{tail}(2k)}\|_r^r \\
&\leq 2(2k)^{1-\frac{r}{p}} \|x_B - w_B\|_p^r + (2k)^{1-\frac{r}{p}} \|w_A - x_A\|_p^r + \|x_{\text{tail}(2k)}\|_r^r \text{ (Fact 5)} \\
&\leq 3(2k)^{1-\frac{r}{p}} \|x - w\|_p^r + \|x_{\text{tail}(2k)}\|_r^r \\
&\leq 3 \cdot 2^{1-\frac{r}{p}} \left(C'_{p,q}\right)^r k^{1-\frac{r}{q}} \|x_{\text{tail}(2k)}\|_q^r + \|x_{\text{tail}(2k)}\|_r^r \text{ (by assumption)} \\
&\leq 3 \cdot 2^{1-\frac{r}{p}} \left(C'_{p,q}\right)^r k^{1-\frac{r}{s}} \|x_{\text{tail}(k)}\|_s^r + k^{1-\frac{r}{s}} \|x_{\text{tail}(k)}\|_s^r \text{ (Lemma 6)} \\
&= \left(1 + 3 \cdot 2^{1-\frac{r}{p}} \left(C'_{p,q}\right)^r\right) k^{1-\frac{r}{s}} \|x_{\text{tail}(k)}\|_s^r,
\end{aligned}$$

which gives us that

$$\|x - z\|_r \leq C_{p,q,r} \cdot k^{\frac{1}{r}-\frac{1}{s}} \|x_{\text{tail}(k)}\|_s$$

for $C_{p,q,r} = \left(1 + 3 \cdot 2^{1-\frac{r}{p}} \left(C'_{p,q}\right)^r\right)^{\frac{1}{r}}$.

□

Note that this reduction loses a factor of two in k due to its application of Lemma 6. If $q = r = s$ then we don't have to apply Lemma 6 and we can therefore avoid losing this factor of two. This gives us the following corollary.

Corollary 9. *For $p, 1 \geq q > 0$, if there exists a recovery scheme $(\Phi, \mathcal{R}')$ such that for some constant $C'_{p,q}$*

$$\|x - \mathcal{R}'(\Phi x)\|_p \leq C'_{p,q} \cdot k^{\frac{1}{p}-\frac{1}{q}} \|x_{\text{tail}(k)}\|_q,$$

then there exists a recovery scheme $(\Phi, \mathcal{R})$ such that for some constant $C_{p,q}$

$$\|x - \mathcal{R}(\Phi x)\|_q \leq C_{p,q} \cdot \|x_{\text{tail}(k)}\|_q.$$

Note that in this case our recovery algorithm $\mathcal{R}$ projects $\mathcal{R}'(\Phi x)$ to its largest k coordinates, rather than the largest $2k$.

References

[BFP+72] Manuel Blum, Robert W. Floyd, Vaughan R. Pratt, Ronald L. Rivest, and Robert Endre Tarjan. Linear time bounds for median computations. In *Proceedings of the 4th Annual ACM Symposium on Theory of Computing (STOC)*, pages 119–124, 1972.

[Can08] Emmanuel Candès. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346(9-10):589–592, 2008.

[CCFC04] Moses Charikar, Kevin Chen, and Martin Farach-Colton. Finding frequent items in data streams. *Theor. Comput. Sci.*, 312(1):3–15, 2004.

[CDD08] Albert Cohen, Wolfgang Dahmen, and Ronald DeVore. Compressed sensing and best k-term approximation. *Journal of the American Mathematical Society*, 22(1):211–231, 2008.

[CM06] Graham Cormode and S. Muthukrishnan. Combinatorial algorithms for compressed sensing. In *Structural Information and Communication Complexity, 13th International Colloquium (SIROCCO)*, pages 280–294, 2006.

[CS08] Rick Chartrand and Valentina Staneva. Restricted isometry properties and nonconvex compressive sensing. *Inverse Problems*, 24(3), 2008.

[CT05] Emmanuel Candès and Terence Tao. Decoding by linear programming. *IEEE Trans. Inf. Theory*, 51(12):4203–4215, 2005.

[Don06] D. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[FR13] Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing*. Applied and Numerical Harmonic Analysis. Birkhäuser, 2013.

[GI10] Anna C. Gilbert and Piotr Indyk. Sparse recovery using sparse matrices. *Proceedings of the IEEE*, 98(6):937–947, 2010.

[GLPS10] Anna C. Gilbert, Yi Li, Ely Porat, and Martin J. Strauss. Approximate sparse recovery: optimizing time and measurements. In *Proceedings of the 42nd ACM Symposium on Theory of Computing (STOC)*, pages 475–484, 2010.

[IR08] Piotr Indyk and Milan Ruzic. Near-optimal sparse recovery in the L1 norm. In *49th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 199–207, 2008.

[SL10] Yi Shen and Song Li. Restricted p-isometry property and its applications for nonconvex compressive sensing. *Adv. Comput. Math.*, 37(3):441–452, 2010.

[SY16] Tasuku Soma and Yuichi Yoshida. Non-convex compressed sensing with the sum-of-squares method. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 570–579, 2016.