Fast algorithms for 1-fair packing (and its dual)

Sebastian Pokutta

Technische Universität Berlin
and
Zuse Institute Berlin

pokutta@math.tu-berlin.de
@spokutta

Joint work with: Francisco Criado and David Martínez-Rubio

HIM Workshop: Continuous approaches to discrete optimization
October 13, 2021

Berlin Mathematics Research Center

Solving the 1-fair packing problem

$$\max_{x \geq 0} \left\{ f(x) \stackrel{\mathrm{def}}{=} \sum_{i=1}^{n} \log x_i : Ax \leq \mathbb{1}_m \right\}. \tag{1FP}$$

What is this talk about?

Introduction

Solving the 1-fair packing problem

$$\max_{x \geq 0} \left\{ f(x) \stackrel{\text{def}}{=} \sum_{i=1}^{n} \log x_i : Ax \leq \mathbb{1}_m \right\}. \tag{1FP}$$

Why? Problem occurs in many relevant applications.

What is this talk about?

Introduction

Solving the 1-fair packing problem

$$\max_{x \geq 0} \left\{ f(x) \stackrel{\text{def}}{=} \sum_{i=1}^{n} \log x_i : Ax \leq \mathbb{1}_m \right\}. \tag{1FP}$$

Why? Problem occurs in many relevant applications.

Today: A fast (and distributed) algorithm for the (1FP)

What is this talk about?

Introduction

Solving the 1-fair packing problem

$$\max_{x \geq 0} \left\{ f(x) \stackrel{\text{def}}{=} \sum_{i=1}^{n} \log x_i : Ax \leq \mathbb{1}_m \right\}. \tag{1FP}$$

Why? Problem occurs in many relevant applications.

Today: A fast (and distributed) algorithm for the (1FP)

Outline

- Quick overview of the 1-fair packing problem
- Accelerated distributed method for its solution

(Hyperlinked) References are not exhaustive; check references contained therein.

The 1-fair packing problem

Problem Definition

The 1-fair packing problem

Given $A \in \mathcal{M}_{m\times n}(\mathbb{R}_{\geq 0})$, the 1-fair packing problem is defined as

$$\max_{x\geq 0}\left\{f(x) \stackrel{\mathrm{def}}{=} \sum_{i=1}^{n} \log x_i : Ax \leq \mathbb{1}_m\right\}. \tag{1FP}$$

Problem Definition

The 1-fair packing problem

Given $A \in \mathcal{M}_{m\times n}(\mathbb{R}_{\geq 0})$, the 1-fair packing problem is defined as

$$\max_{x\geq 0}\left\{f(x) \stackrel{\text{def}}{=} \sum_{i=1}^{n} \log x_i : Ax \leq \mathbb{1}_m\right\}. \tag{1FP}$$

The width ρ of A is defined as the maximum ratio of the non-zero entries of A:

$$\rho \stackrel{\text{def}}{=} \max\{A_{ij}\}/\min_{A_{ij}\neq 0}\{A_{ij}\}.$$

Problem Definition

The 1-fair packing problem

Given $A \in \mathcal{M}_{m\times n}(\mathbb{R}_{\geq 0})$, the 1-fair packing problem is defined as

$$\max_{x \geq 0} \left\{ f(x) \stackrel{\text{def}}{=} \sum_{i=1}^{n} \log x_i : Ax \leq \mathbb{1}_m \right\}. \tag{1FP}$$

The width ρ of A is defined as the maximum ratio of the non-zero entries of A:

$$\rho \stackrel{\text{def}}{=} \max\{A_{ij}\} / \min_{A_{ij} \neq 0} \{A_{ij}\}.$$

We are interested in algorithms that are:

1. width-independent (usually: width-dependence $\Rightarrow$ non-polynomiality)
2. provide additive ε-minimizers

Why should we care?

The 1-fair packing problem

The 1-fair packing problem (aka *proportional fair allocation*):

1. Arises under a natural set of fairness axioms [Bertsimas et al., 2011, Lan et al., 2010]
2. Corresponds to Nash bargaining solutions [Nash, 1950]
3. Multi-resource allocation [Bonald and Roberts, 2015, Jin and Hayashi, 2018, Joe-Wong et al., 2012]
4. Rate control in networks [Kelly, 1997]
5. Game theory [Jain and Vazirani, 2010, 2007]

Why should we care?

The 1-fair packing problem

The 1-fair packing problem (aka *proportional fair allocation*):

1. Arises under a natural set of fairness axioms [Bertsimas et al., 2011, Lan et al., 2010]
2. Corresponds to Nash bargaining solutions [Nash, 1950]
3. Multi-resource allocation [Bonald and Roberts, 2015, Jin and Hayashi, 2018, Joe-Wong et al., 2012]
4. Rate control in networks [Kelly, 1997]
5. Game theory [Jain and Vazirani, 2010, 2007]

Other important allocations:

1. linear objectives (no fairness)
2. max-min allocations [Mo and Walrand, 2000]
3. α-fair allocations (generalization) [Atkinson, 1970, Mo and Walrand, 2000, McCormick et al., 2014]

Overview of Results

The 1-fair packing problem

Paper	Problem	Iterations	Width-dependence?
[Beck et al., 2014]	Primal	$O(\rho^2 mn/\epsilon)$	Yes
[Marašević et al., 2016]	Primal	$\widetilde{O}(n^5/\epsilon^5)$	nearly No (polylog)
[Diakonikolas et al., 2020]	Primal	$\widetilde{O}(n^2/\epsilon^2)$	nearly No (polylog)
[Criado et al., 2021]	Primal	$\widetilde{O}(n/\epsilon)$	No
[Beck et al., 2014]	Dual	$O(\rho\sqrt{mn/\epsilon})$	Yes
[Criado et al., 2021]	Dual	$\widetilde{O}(n^2/\epsilon)$	No

Table: Comparison of algorithms for 1-fair packing and its dual. The work of one iteration is linear in N, the number of non-zero entries in A.

Overview of Results

The 1-fair packing problem

Paper	Problem	Iterations	Width-dependence?
[Beck et al., 2014]	Primal	$O(\rho^2 mn/\epsilon)$	Yes
[Marašević et al., 2016]	Primal	$\widetilde{O}(n^5/\epsilon^5)$	nearly No (polylog)
[Diakonikolas et al., 2020]	Primal	$\widetilde{O}(n^2/\epsilon^2)$	nearly No (polylog)
[Criado et al., 2021]	Primal	$\widetilde{O}(n/\epsilon)$	No
[Beck et al., 2014]	Dual	$O(\rho\sqrt{mn/\epsilon})$	Yes
[Criado et al., 2021]	Dual	$\widetilde{O}(n^2/\epsilon)$	No

Table: Comparison of algorithms for 1-fair packing and its dual. The work of one iteration is linear in N, the number of non-zero entries in A.

Today: Primal algorithm via accelerated first-order method.

A distributed accelerated algorithm for the 1-fair packing problem

Key ingredients and algorithmic properties

Key ingredients.

1. Problem transformation to ensure width independence
2. Smoothed objective + special barrier

   Smoothening + accelerated coordinate descent used for 0-fair packing with multiplicative guarantees [Allen-Zhu and Orecchia, 2019]
3. Truncated gradients for the Mirror Descent step

   Smoothening + truncated gradients used for 1-fair packing without acceleration but additive guarantees [Diakonikolas et al., 2020]
4. Acceleration via Linear Coupling of Mirror Descent and Gradient Descent

   Very flexible approach to achieve acceleration [Allen-Zhu and Orecchia, 2017]

Key ingredients and algorithmic properties

Key ingredients.

1. Problem transformation to ensure width independence
2. Smoothed objective + special barrier

   Smoothening + accelerated coordinate descent used for 0-fair packing with multiplicative guarantees [Allen-Zhu and Orecchia, 2019]
3. Truncated gradients for the Mirror Descent step

   Smoothening + truncated gradients used for 1-fair packing without acceleration but additive guarantees [Diakonikolas et al., 2020]
4. Acceleration via Linear Coupling of Mirror Descent and Gradient Descent

   Very flexible approach to achieve acceleration [Allen-Zhu and Orecchia, 2017]

Algorithmic Properties.

1. Accelerated first-order method
2. Distributed

   Each worker j only needs x_j, column A_j, and global parameters [Kelly and Yudovina, 2014, Awerbuch and Khandekar, 2008]
3. Width-independent
4. Deterministic

Problem transformation

Preliminaries

Step 1. Rescale the original problem data via rescaling columns of A, so that

$$\max_{j\in[m]}\{A_{ij}\} = 1, \text{ for all } i \in [n]. \qquad \text{(rescaling)}$$

Note. Via rescaling of primal coordinates. Changes the objectives by an additive constant. Obtained additive guarantees are preserved.

Problem transformation

Preliminaries

Step 1. Rescale the original problem data via rescaling columns of A, so that

$$\max_{j\in[m]}\{A_{ij}\} = 1, \text{ for all } i \in [n]. \tag{rescaling}$$

Note. Via rescaling of primal coordinates. Changes the objectives by an additive constant. Obtained additive guarantees are preserved.

Step 2. Reparametrize the problem, via $\hat{f} : \mathbb{R}^n \to \mathbb{R}, x \mapsto f(\exp(x)) = \langle \mathbb{1}_n, x\rangle$:

$$\max_{x\in\mathbb{R}^n}\left\{\hat{f}(x) \stackrel{\text{def}}{=} \langle \mathbb{1}_n, x\rangle : A\exp(x) \leq \mathbb{1}_m\right\}. \tag{expSpace}$$

Note. Now we have a simple objective but more complex constraints.

Problem transformation

Preliminaries

Step 1. Rescale the original problem data via rescaling columns of A, so that

$$\max_{j\in[m]}\{A_{ij}\}=1, \text{ for all } i\in[n]. \qquad \text{(rescaling)}$$

Note. Via rescaling of primal coordinates. Changes the objectives by an additive constant. Obtained additive guarantees are preserved.

Step 2. Reparametrize the problem, via $\hat{f}:\mathbb{R}^n\to\mathbb{R}, x\mapsto f(\exp(x))=\langle \mathbb{1}_n, x\rangle$:

$$\max_{x\in\mathbb{R}^n}\left\{\hat{f}(x)\stackrel{\text{def}}{=}\langle \mathbb{1}_n, x\rangle : A\exp(x)\le \mathbb{1}_m\right\}. \qquad \text{(expSpace)}$$

Note. Now we have a simple objective but more complex constraints.

Step 3. Regularize objective by putting constraints into barrier

$$f_r(x)\stackrel{\text{def}}{=}-\langle \mathbb{1}_n, x\rangle+\frac{\beta}{1+\beta}\sum_{i=1}^{m}(A\exp(x))_i^{\frac{1+\beta}{\beta}} \quad \text{with} \quad \beta\stackrel{\text{def}}{=}\frac{\epsilon}{6n\log(2mn^2/\epsilon)}$$

Note. Unconstrained minimization problem now (technically we add a box later but...)

The gradient and the barrier function visualized

Gradients of f_r. At coordinate $j \in [m]$ given by

$$\nabla_j f_r(x) = -1 + \sum_{i=1}^{n} (A\exp(x))_i^{\frac{1}{\beta}} a_{ij} \exp(x_j), \qquad \text{(nabla-fr)}$$

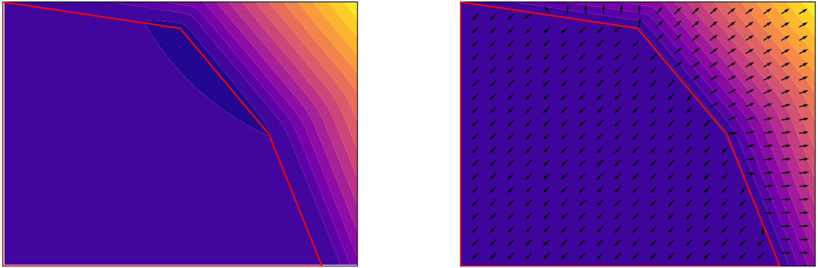

Figure: Regularized objective f_r (left) and its gradient (right), for a sample matrix $A \in \mathcal{M}_{3\times 2}$. For visualization purposes we show $\log f_r(x)$ and $\log \|\nabla f_r(x)\|$, represented by color, and we indicate the direction of the gradient with normalized arrows. Results are shown in original space but the gradient was computed as originally defined in (nabla-fr).

Lemma (Key Lemma (together with rescaling) to width-independence)

Let A satisfy the normalization in (rescaling)*, and let $\bar{x}^*$ be the optimizer of Problem* (1FP)*. Then $\bar{x}_i^* \geq 1/n$ for all $i \in [n]$.*

Lemma (Key Lemma (together with rescaling) to width-independence)

Let A satisfy the normalization in (rescaling)*, and let $\bar{x}^*$ be the optimizer of Problem* (1FP)*. Then $\bar{x}_i^* \geq 1/n$ for all $i \in [n]$.*

Let $\omega \stackrel{\text{def}}{=} \log(mn/(1-\epsilon/n))$ and define box $B \stackrel{\text{def}}{=} [-\omega, 0]^n$. It suffices to solve

$$\min_{x \in B} f_r(x). \qquad \text{(1FP-primalReg)}$$

Reformulation and Width-independence

Preliminaries

Lemma (Key Lemma (together with rescaling) to width-independence)

Let A satisfy the normalization in (rescaling)*, and let $\bar{x}^*$ be the optimizer of Problem* (1FP)*. Then $\bar{x}^*_i \geq 1/n$ for all $i \in [n]$.*

Let $\omega \stackrel{\text{def}}{=} \log(mn/(1-\epsilon/n))$ and define box $B \stackrel{\text{def}}{=} [-\omega, 0]^n$. It suffices to solve

$$\min_{x \in B} f_r(x). \tag{1FP-primalReg}$$

Proposition (Solving (1FP-primalReg) is good enough)

Let $\epsilon \in (0, n/2]$. Let x_r^* be the minimizer of (1FP-primalReg)*, let x^* be the maximizer of $\hat{f}$ and let $x_r^\epsilon \in B$ be a point such that $f_r(x_r^\epsilon) - f_r(x_r^*) \leq \epsilon$, i.e., an ϵ-minimizer of* (1FP-primalReg)*. Then we have:*

1. *The point x_r^* satisfies $\hat{f}(x^*) - \hat{f}(x_r^*) \leq \hat{f}(x^*) + f_r(x_r^*) \leq 3\epsilon$.*
2. *The point x_r^ε satisfies $A\exp(x_r^\epsilon) \leq (1+\epsilon/n)\mathbb{1}_m$*
3. *The point $u = x_r^\epsilon - \log(1+\epsilon/n)\mathbb{1}_n$ satisfies*

$$f(\exp(x^*)) - f(\exp(u)) = \hat{f}(x^*) - \hat{f}(u) \leq 5\epsilon \text{ and } A\exp(u) \leq \mathbb{1}_m.$$

Algorithm Accelerated descent method for 1-Fair Packing

Input: Matrix $A \in \mathcal{M}_{m\times n}(\mathbb{R}_{\geq 0})$ normalized as in (rescaling). Accuracy $\epsilon \in (0, n/2]$.

$\beta \leftarrow \frac{\epsilon}{6n\log(2mn^2/\epsilon)}$; $\omega \leftarrow \log(\frac{mn}{1-\epsilon/n})$; $L = \max\left\{\frac{4\omega(1+\beta)}{\beta}, \frac{16n\log(2mn)}{3\epsilon} + \frac{1}{3}\right\} = \widetilde{O}(n/\epsilon)$

$\eta_0 \leftarrow \frac{1}{3L}$; $C_k = 3\eta_k L$; $\tau \leftarrow \tau_k = \eta_k/C_k = 1/(3L)$.

$T \leftarrow \lceil \log(\frac{4n\log(2mn)}{\epsilon})/\log(\frac{1}{1-\tau})\rceil \leq \lceil 3L\log(\frac{4n\log(2mn)}{\epsilon})\rceil = \widetilde{O}(n/\epsilon)$

$x^{(0)} \leftarrow y^{(0)} \leftarrow z^{(0)} \leftarrow -\log(mn/(1-\epsilon/n))\mathbb{1}_n$

for $k = 1$ **to** T **do**

$\eta_k \leftarrow C_k - C_{k-1} = \frac{1}{1-\tau}\eta_{k-1}$

$x^{(k)} \leftarrow \tau z^{(k-1)} + (1-\tau)y^{(k-1)}$ ◇ Linear coupling

$z^{(k)} \leftarrow \arg\min_{z\in B}\left\{\left\langle \eta_k \overline{\nabla f_r}(x^{(k)}), z\right\rangle + \frac{1}{2\omega}\|z - z^{(k-1)}\|_2^2\right\}$ ◇ Mirror descent step

$y^{(k)} \leftarrow x^{(k)} + \frac{1}{\eta_k L}(z^{(k)} - z^{(k-1)})$ ◇ Gradient descent step

end for

The Algorithm

Algorithm and Convergence Guarantee

Algorithm Accelerated descent method for 1-Fair Packing

Input: Matrix $A \in \mathcal{M}_{m\times n}(\mathbb{R}_{\geq 0})$ normalized as in (rescaling). Accuracy $\epsilon \in (0, n/2]$.

$\beta \leftarrow \frac{\epsilon}{6n\log(2mn^2/\epsilon)}$; $\omega \leftarrow \log(\frac{mn}{1-\epsilon/n})$; $L = \max\left\{\frac{4\omega(1+\beta)}{\beta}, \frac{16n\log(2mn)}{3\epsilon} + \frac{1}{3}\right\} = \widetilde{O}(n/\epsilon)$

$\eta_0 \leftarrow \frac{1}{3L}$; $C_k = 3\eta_k L$; $\tau \leftarrow \tau_k = \eta_k / C_k = 1/(3L)$.

$T \leftarrow \lceil \log(\frac{4n\log(2mn)}{\epsilon}) / \log(\frac{1}{1-\tau}) \rceil \leq \lceil 3L\log(\frac{4n\log(2mn)}{\epsilon}) \rceil = \widetilde{O}(n/\epsilon)$

$x^{(0)} \leftarrow y^{(0)} \leftarrow z^{(0)} \leftarrow -\log(mn/(1-\epsilon/n))\mathbb{1}_n$

for $k = 1$ **to** T **do**

$\quad \eta_k \leftarrow C_k - C_{k-1} = \frac{1}{1-\tau}\eta_{k-1}$

$\quad x^{(k)} \leftarrow \tau z^{(k-1)} + (1-\tau)y^{(k-1)}$ ◇ Linear coupling

$\quad z^{(k)} \leftarrow \arg\min_{z\in B}\left\{\left\langle \eta_k \overline{\nabla f_r}(x^{(k)}), z\right\rangle + \frac{1}{2\omega}\|z - z^{(k-1)}\|_2^2\right\}$ ◇ Mirror descent step

$\quad y^{(k)} \leftarrow x^{(k)} + \frac{1}{\eta_k L}(z^{(k)} - z^{(k-1)})$ ◇ Gradient descent step

end for

Algorithm Accelerated descent method for 1-Fair Packing

Input: Matrix $A \in \mathcal{M}_{m\times n}(\mathbb{R}_{\geq 0})$ normalized as in (rescaling). Accuracy $\epsilon \in (0, n/2]$.

$\beta \leftarrow \frac{\epsilon}{6n\log(2mn^2/\epsilon)}$; $\omega \leftarrow \log(\frac{mn}{1-\epsilon/n})$; $L = \max\left\{\frac{4\omega(1+\beta)}{\beta}, \frac{16n\log(2mn)}{3\epsilon} + \frac{1}{3}\right\} = \widetilde{O}(n/\epsilon)$

$\eta_0 \leftarrow \frac{1}{3L}$; $C_k = 3\eta_k L$; $\tau \leftarrow \tau_k = \eta_k/C_k = 1/(3L)$.

$T \leftarrow \lceil \log(\frac{4n\log(2mn)}{\epsilon})/\log(\frac{1}{1-\tau})\rceil \leq \lceil 3L\log(\frac{4n\log(2mn)}{\epsilon})\rceil = \widetilde{O}(n/\epsilon)$

$x^{(0)} \leftarrow y^{(0)} \leftarrow z^{(0)} \leftarrow -\log(mn/(1-\epsilon/n))\mathbb{1}_n$

for $k = 1$ **to** T **do**

$\quad \eta_k \leftarrow C_k - C_{k-1} = \frac{1}{1-\tau}\eta_{k-1}$

$\quad x^{(k)} \leftarrow \tau z^{(k-1)} + (1-\tau)y^{(k-1)}$ ◇ Linear coupling

$\quad z^{(k)} \leftarrow \arg\min_{z\in B}\left\{\left\langle \eta_k \overline{\nabla f_r}(x^{(k)}), z\right\rangle + \frac{1}{2\omega}\|z - z^{(k-1)}\|_2^2\right\}$ ◇ Mirror descent step

$\quad y^{(k)} \leftarrow x^{(k)} + \frac{1}{\eta_k L}(z^{(k)} - z^{(k-1)})$ ◇ Gradient descent step

end for

The Algorithm

Algorithm and Convergence Guarantee

Algorithm Accelerated descent method for 1-Fair Packing

Input: Matrix $A \in \mathcal{M}_{m\times n}(\mathbb{R}_{\geq 0})$ normalized as in (rescaling). Accuracy $\epsilon \in (0, n/2]$.

$\beta \leftarrow \frac{\epsilon}{6n\log(2mn^2/\epsilon)}$; $\omega \leftarrow \log(\frac{mn}{1-\epsilon/n})$; $L = \max\left\{\frac{4\omega(1+\beta)}{\beta}, \frac{16n\log(2mn)}{3\epsilon} + \frac{1}{3}\right\} = \widetilde{O}(n/\epsilon)$

$\eta_0 \leftarrow \frac{1}{3L}$; $C_k = 3\eta_k L$; $\tau \leftarrow \tau_k = \eta_k / C_k = 1/(3L)$.

$T \leftarrow \lceil \log(\frac{4n\log(2mn)}{\epsilon}) / \log(\frac{1}{1-\tau}) \rceil \leq \lceil 3L\log(\frac{4n\log(2mn)}{\epsilon}) \rceil = \widetilde{O}(n/\epsilon)$

$x^{(0)} \leftarrow y^{(0)} \leftarrow z^{(0)} \leftarrow -\log(mn/(1-\epsilon/n))\mathbb{1}_n$

for $k = 1$ **to** T **do**

$\quad \eta_k \leftarrow C_k - C_{k-1} = \frac{1}{1-\tau}\eta_{k-1}$

$\quad x^{(k)} \leftarrow \tau z^{(k-1)} + (1-\tau) y^{(k-1)}$ ⋄ Linear coupling

$\quad z^{(k)} \leftarrow \arg\min_{z\in B}\left\{\left\langle \eta_k \overline{\nabla f_r}(x^{(k)}), z\right\rangle + \frac{1}{2\omega}\|z - z^{(k-1)}\|_2^2\right\}$ ⋄ Mirror descent step

$\quad y^{(k)} \leftarrow x^{(k)} + \frac{1}{\eta_k L}(z^{(k)} - z^{(k-1)})$ ⋄ Gradient descent step

end for

Truncated Gradient. Run Mirror Descent with $\overline{\nabla f_r}(x^{(k)}) \in [-1,1]^n$ defined as

$$\overline{\nabla_i f_r}(x^{(k)}) \stackrel{\text{def}}{=} \min\{1, \nabla_i f_r(x^{(k)})\} \text{ for all } i \in [n]. \qquad \text{(truncGradient)}$$

The Mirror Descent Step

Algorithm and Convergence Guarantee

Truncated Gradient. Run Mirror Descent with $\overline{\nabla f_r}(x^{(k)}) \in [-1, 1]^n$ defined as

$$\overline{\nabla_i f_r}(x^{(k)}) \stackrel{\text{def}}{=} \min\{1, \nabla_i f_r(x^{(k)})\} \text{ for all } i \in [n]. \qquad \text{(truncGradient)}$$

Lemma (Mirror Descent Guarantee)

Let $u \in B$ and choose L as in the Algorithm. It holds that:

$$\left\langle \eta_k \overline{\nabla f_r}(x^{(k)}), z^{(k-1)} - u \right\rangle \leq \eta_k^2 L \left\langle \overline{\nabla f_r}(x^{(k)}), x^{(k)} - y^{(k)} \right\rangle + \frac{1}{2\omega} \|z^{(k-1)} - u\|_2^2 - \frac{1}{2\omega} \|z^{(k)} - u\|_2^2.$$

(Basically classical MD guarantee without CS and using $z^{(k-1)} - z^{(k)} = \eta_k L(x^{(k)} - y^{(k)})$)

Truncated Gradient. Run Mirror Descent with $\overline{\nabla f_r}(x^{(k)}) \in [-1, 1]^n$ defined as

$$\overline{\nabla_i f_r}(x^{(k)}) \stackrel{\text{def}}{=} \min\{1, \nabla_i f_r(x^{(k)})\} \text{ for all } i \in [n]. \qquad \text{(truncGradient)}$$

Lemma (Mirror Descent Guarantee)

Let $u \in B$ and choose L as in the Algorithm. It holds that:

$$\left\langle \eta_k \overline{\nabla f_r}(x^{(k)}), z^{(k-1)} - u \right\rangle \le \eta_k^2 L \left\langle \overline{\nabla f_r}(x^{(k)}), x^{(k)} - y^{(k)} \right\rangle + \frac{1}{2\omega}\|z^{(k-1)} - u\|_2^2 - \frac{1}{2\omega}\|z^{(k)} - u\|_2^2.$$

(Basically classical MD guarantee without CS and using $z^{(k-1)} - z^{(k)} = \eta_k L(x^{(k)} - y^{(k)})$)

Note. We use 'wrong' gradient; need to pay for this.

The Gradient Descent Step

Algorithm and Convergence Guarantee

Descent Lemma can be derived from a lemma in [Diakonikolas et al., 2020]

Lemma (Descent Lemma)

Given $x^{(k)}$ and $y^{(k)}$ as defined in the Algorithm, the following holds:

$$f_r(x^{(k)}) - f_r(y^{(k)}) \geq \frac{1}{2}\left\langle \nabla f_r(x^{(k)}), x^{(k)} - y^{(k)} \right\rangle \geq 0.$$

The Gradient Descent Step

Algorithm and Convergence Guarantee

Descent Lemma can be derived from a lemma in [Diakonikolas et al., 2020]

Lemma (Descent Lemma)

Given $x^{(k)}$ and $y^{(k)}$ as defined in the Algorithm, the following holds:

$$f_r(x^{(k)}) - f_r(y^{(k)}) \geq \frac{1}{2}\left\langle \nabla f_r(x^{(k)}), x^{(k)} - y^{(k)} \right\rangle \geq 0.$$

Analysis similar to Lemma 3.10 in [Allen-Zhu and Orecchia, 2019]

Lemma (GD compensates for MD regret *and* truncation)

Let $C_k \stackrel{\text{def}}{=} 3\eta_k L$, and let $\nu^{(k)} \stackrel{\text{def}}{=} \nabla f_r(x^{(k)}) - \overline{\nabla f_r}(x^{(k)}) \in [0, \infty)^n$. For all $u \in B$, we have

$$\left\langle \eta_k \nu^{(k)}, z^{(k-1)} - u \right\rangle + \eta_k^2 L \left\langle \overline{\nabla f_r}(x^{(k)}), x^{(k)} - y^{(k)} \right\rangle \leq C_k(f_r(x^{(k)}) - f_r(y^{(k)})).$$

Linear Coupling: Bringing it all together (Proof Sketch)

Algorithm and Convergence Guarantee

Step one. Consider gap with respect to $x^{(k)}$:

$$\eta_k (f_r(x^{(k)}) - f_r(u))$$

Linear Coupling: Bringing it all together (Proof Sketch)

Algorithm and Convergence Guarantee

Step one. Consider gap with respect to $x^{(k)}$:

$$
\begin{aligned}
&\eta_k(f_r(x^{(k)}) - f_r(u)) \\
&\leq \left\langle \eta_k \nabla f_r(x^{(k)}), x^{(k)} - u \right\rangle && \text{(Convexity)} \\
&= \left\langle \eta_k \nabla f_r(x^{(k)}), x^{(k)} - z^{(k-1)} \right\rangle + \left\langle \eta_k \nu^{(k)}, z^{(k-1)} - u \right\rangle + \left\langle \eta_k \overline{\nabla f_r}(x^{(k)}), z^{(k-1)} - u \right\rangle && \text{(Expansion)}
\end{aligned}
$$

Linear Coupling: Bringing it all together (Proof Sketch)

Algorithm and Convergence Guarantee

Step one. Consider gap with respect to $x^{(k)}$:

$$
\begin{aligned}
&\eta_k (f_r(x^{(k)}) - f_r(u)) \\
&\leq \left\langle \eta_k \nabla f_r(x^{(k)}), x^{(k)} - u \right\rangle && \text{(Convexity)} \\
&= \left\langle \eta_k \nabla f_r(x^{(k)}), x^{(k)} - z^{(k-1)} \right\rangle + \left\langle \eta_k \nu^{(k)}, z^{(k-1)} - u \right\rangle + \left\langle \eta_k \overline{\nabla f_r}(x^{(k)}), z^{(k-1)} - u \right\rangle && \text{(Expansion)} \\
&= \frac{(1-\tau)\eta_k}{\tau} \left\langle \nabla f_r(x^{(k)}), y^{(k-1)} - x^{(k)} \right\rangle + \left\langle \eta_k \nu^{(k)}, z^{(k-1)} - u \right\rangle + \left\langle \eta_k \overline{\nabla f_r}(x^{(k)}), z^{(k-1)} - u \right\rangle && \text{(Definition of } x^{(k)}\text{)}
\end{aligned}
$$

Step one. Consider gap with respect to $x^{(k)}$:

$$
\begin{aligned}
&\eta_k (f_r(x^{(k)}) - f_r(u)) \\
&\leq \left\langle \eta_k \nabla f_r(x^{(k)}), x^{(k)} - u \right\rangle && \text{(Convexity)} \\
&= \left\langle \eta_k \nabla f_r(x^{(k)}), x^{(k)} - z^{(k-1)} \right\rangle + \left\langle \eta_k \nu^{(k)}, z^{(k-1)} - u \right\rangle + \left\langle \eta_k \overline{\nabla f_r}(x^{(k)}), z^{(k-1)} - u \right\rangle && \text{(Expansion)} \\
&= \frac{(1-\tau)\eta_k}{\tau} \left\langle \nabla f_r(x^{(k)}), y^{(k-1)} - x^{(k)} \right\rangle + \left\langle \eta_k \nu^{(k)}, z^{(k-1)} - u \right\rangle + \left\langle \eta_k \overline{\nabla f_r}(x^{(k)}), z^{(k-1)} - u \right\rangle && \text{(Definition of } x^{(k)}\text{)} \\
&\leq \frac{(1-\tau)\eta_k}{\tau} (f_r(y^{(k-1)}) - f_r(x^{(k)})) + \left\langle \eta_k \nu^{(k)}, z^{(k-1)} - u \right\rangle + \left\langle \eta_k^2 L \overline{\nabla f_r}(x^{(k)}), x^{(k)} - y^{(k)} \right\rangle \\
&\quad + \frac{1}{2\omega} \|z^{(k-1)} - u\|_2^2 - \frac{1}{2\omega} \|z^{(k)} - u\|_2^2] && \text{(Convexity for first term + MD Lemma)}
\end{aligned}
$$

Linear Coupling: Bringing it all together (Proof Sketch)

Algorithm and Convergence Guarantee

Step one. Consider gap with respect to $x^{(k)}$:

$$
\begin{aligned}
&\eta_k(f_r(x^{(k)}) - f_r(u)) \\
&\leq \left\langle \eta_k \nabla f_r(x^{(k)}), x^{(k)} - u \right\rangle && \text{(Convexity)} \\
&= \left\langle \eta_k \nabla f_r(x^{(k)}), x^{(k)} - z^{(k-1)} \right\rangle + \left\langle \eta_k \nu^{(k)}, z^{(k-1)} - u \right\rangle + \left\langle \eta_k \overline{\nabla f_r}(x^{(k)}), z^{(k-1)} - u \right\rangle && \text{(Expansion)} \\
&= \frac{(1-\tau)\eta_k}{\tau} \left\langle \nabla f_r(x^{(k)}), y^{(k-1)} - x^{(k)} \right\rangle + \left\langle \eta_k \nu^{(k)}, z^{(k-1)} - u \right\rangle + \left\langle \eta_k \overline{\nabla f_r}(x^{(k)}), z^{(k-1)} - u \right\rangle && \text{(Definition of } x^{(k)}) \\
&\leq \frac{(1-\tau)\eta_k}{\tau} (f_r(y^{(k-1)}) - f_r(x^{(k)})) + \left\langle \eta_k \nu^{(k)}, z^{(k-1)} - u \right\rangle + \left\langle \eta_k^2 L \overline{\nabla f_r}(x^{(k)}), x^{(k)} - y^{(k)} \right\rangle \\
&\quad + \frac{1}{2\omega} \|z^{(k-1)} - u\|_2^2 - \frac{1}{2\omega} \|z^{(k)} - u\|_2^2] && \text{(Convexity for first term + MD Lemma)} \\
&\leq \frac{(1-\tau)\eta_k}{\tau} (f_r(y^{(k-1)}) - f_r(x^{(k)})) + C_k (f_r(x^{(k)}) - f_r(y^{(k)}) + \frac{1}{2\omega} \|z^{(k-1)} - u\|_2^2 \\
&\quad - \frac{1}{2\omega} \|z^{(k)} - u\|_2^2 && \text{(MD + truncation compensation via GD)}
\end{aligned}
$$

Linear Coupling: Bringing it all together (Proof Sketch)

Algorithm and Convergence Guarantee

Step one. Consider gap with respect to $x^{(k)}$:

$$
\begin{aligned}
&\eta_k(f_r(x^{(k)}) - f_r(u)) \\
&\leq \left\langle \eta_k \nabla f_r(x^{(k)}), x^{(k)} - u \right\rangle && \text{(Convexity)} \\
&= \left\langle \eta_k \nabla f_r(x^{(k)}), x^{(k)} - z^{(k-1)} \right\rangle + \left\langle \eta_k \nu^{(k)}, z^{(k-1)} - u \right\rangle + \left\langle \eta_k \overline{\nabla f_r}(x^{(k)}), z^{(k-1)} - u \right\rangle && \text{(Expansion)} \\
&= \frac{(1-\tau)\eta_k}{\tau} \left\langle \nabla f_r(x^{(k)}), y^{(k-1)} - x^{(k)}) \right\rangle + \left\langle \eta_k \nu^{(k)}, z^{(k-1)} - u \right\rangle + \left\langle \eta_k \overline{\nabla f_r}(x^{(k)}), z^{(k-1)} - u \right\rangle && \text{(Definition of } x^{(k)}\text{)} \\
&\leq \frac{(1-\tau)\eta_k}{\tau} (f_r(y^{(k-1)}) - f_r(x^{(k)})) + \left\langle \eta_k \nu^{(k)}, z^{(k-1)} - u \right\rangle + \left\langle \eta_k^2 L \overline{\nabla f_r}(x^{(k)}), x^{(k)} - y^{(k)} \right\rangle \\
&\qquad + \frac{1}{2\omega} \|z^{(k-1)} - u\|_2^2 - \frac{1}{2\omega} \|z^{(k)} - u\|_2^2] && \text{(Convexity for first term + MD Lemma)} \\
&\leq \frac{(1-\tau)\eta_k}{\tau} (f_r(y^{(k-1)}) - f_r(x^{(k)})) + C_k (f_r(x^{(k)}) - f_r(y^{(k)}) + \frac{1}{2\omega} \|z^{(k-1)} - u\|_2^2 \\
&\qquad - \frac{1}{2\omega} \|z^{(k)} - u\|_2^2 && \text{(MD + truncation compensation via GD)} \\
&\leq \eta_k f_r(x^{(k)}) + (C_k - \eta_k) f_r(y^{(k-1)}) - C_k f_r(y^{(k)}) + \frac{1}{2\omega} \|z^{(k-1)} - u\|_2^2 - \frac{1}{2\omega} \|z^{(k)} - u\|_2^2 && \text{(Substitute } \tau \text{ to cancel } \eta_k f_r(x^{(k)})\text{)}
\end{aligned}
$$

Linear Coupling: Bringing it all together (Proof Sketch)

Algorithm and Convergence Guarantee

Step two. Telescope using $C_k - \eta_k = C_{k-1}$ and setting $u = x_r^*$.

$$\left(-C_0 - \sum_{k=1}^{T} \eta_k\right) f_r(x_r^*) \leq C_0(f_r(y^{(0)}) - f_r(x_r^*)) - C_T f_r(y^{(T)}) + \frac{1}{2\omega} \|z^{(0)} - x_r^*\|_2^2.$$

Linear Coupling: Bringing it all together (Proof Sketch)

Algorithm and Convergence Guarantee

Step two. Telescope using $C_k - \eta_k = C_{k-1}$ and setting $u = x_r^*$.

$$\left(-C_0 - \sum_{k=1}^{T} \eta_k\right) f_r(x_r^*) \leq C_0(f_r(y^{(0)}) - f_r(x_r^*)) - C_T f_r(y^{(T)}) + \frac{1}{2\omega}\|z^{(0)} - x_r^*\|_2^2.$$

Step three. Clean up and using C_0 (via η_0) and T

$$\begin{aligned} f_r(y^{(T)}) &\leq f_r(x_r^*) + \frac{1}{C_T}\left(C_0(f_r(y^{(0)}) - f_r(x_r^*)) + \frac{1}{2\omega}\|z^{(0)} - x_r^*\|_2^2\right) \\ &\leq f_r(x_r^*) + \frac{1}{C_T}\left(C_0(n(\log(2mn) + 1) + \frac{n\log(mn)}{2}\right) \\ &\leq f_r(x_r^*) + \epsilon \end{aligned}$$

□

Thank you!

References I

Z. Allen-Zhu and L. Orecchia. Linear coupling: An ultimate unification of gradient and mirror descent. In C. H. Papadimitriou, editor, *8th Innovations in Theoretical Computer Science Conference, ITCS 2017, January 9-11, 2017, Berkeley, CA, USA*, volume 67 of *LIPIcs*, pages 3:1–3:22. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017. doi: 10.4230/LIPIcs.ITCS.2017.3. URL https://doi.org/10.4230/LIPIcs.ITCS.2017.3.

Z. Allen-Zhu and L. Orecchia. Nearly linear-time packing and covering LP solvers - achieving width-independence and 1/epsilon-convergence. *Math. Program.*, 175(1-2):307–353, 2019. doi: 10.1007/s10107-018-1244-x. URL https://doi.org/10.1007/s10107-018-1244-x.

A. B. Atkinson. On the measurement of inequality. *Journal of economic theory*, 2(3):244–263, 1970.

B. Awerbuch and R. Khandekar. Stateless distributed gradient descent for positive linear programs. In C. Dwork, editor, *Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008*, pages 691–700. ACM, 2008. doi: 10.1145/1374376.1374476. URL https://doi.org/10.1145/1374376.1374476.

A. Beck, A. Nedic, A. E. Ozdaglar, and M. Teboulle. An O(1/k) gradient method for network resource allocation problems. *IEEE Trans. Control. Netw. Syst.*, 1(1): 64–73, 2014. doi: 10.1109/TCNS.2014.2309751. URL https://doi.org/10.1109/TCNS.2014.2309751.

D. Bertsimas, V. F. Farias, and N. Trichakis. The price of fairness. *Oper. Res.*, 59(1):17–31, 2011. doi: 10.1287/opre.1100.0865. URL https://doi.org/10.1287/opre.1100.0865.

T. Bonald and J. W. Roberts. Multi-resource fairness: Objectives, algorithms and performance. In B. Lin, J. J. Xu, S. Sengupta, and D. Shah, editors, *Proceedings of the 2015 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, Portland, OR, USA, June 15-19, 2015*, pages 31–42. ACM, 2015. doi: 10.1145/2745844.2745869. URL https://doi.org/10.1145/2745844.2745869.

F. Criado, D. Martinez-Rubio, and S. Pokutta. Fast algorithms for packing proportional fairness and its dual. *preprint*, 9 2021.

J. Diakonikolas, M. Fazel, and L. Orecchia. Fair packing and covering on a relative scale. *SIAM J. Optim.*, 30(4):3284–3314, 2020. doi: 10.1137/19M1288516. URL https://doi.org/10.1137/19M1288516.

K. Jain and V. V. Vazirani. Eisenberg-gale markets: algorithms and structural properties. In D. S. Johnson and U. Feige, editors, *Proceedings of the 39th Annual ACM Symposium on Theory of Computing, San Diego, California, USA, June 11-13, 2007*, pages 364–373. ACM, 2007. doi: 10.1145/1250790.1250845. URL https://doi.org/10.1145/1250790.1250845.

K. Jain and V. V. Vazirani. Eisenberg-gale markets: Algorithms and game-theoretic properties. *Games Econ. Behav.*, 70(1):84–106, 2010. doi: 10.1016/j.geb.2008.11.011. URL https://doi.org/10.1016/j.geb.2008.11.011.

Y. Jin and M. Hayashi. Trade-off between fairness and efficiency in dominant alpha-fairness family. In *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops, INFOCOM Workshops 2018, Honolulu, HI, USA, April 15-19, 2018*, pages 391–396. IEEE, 2018. doi: 10.1109/INFCOMW.2018.8406869. URL https://doi.org/10.1109/INFCOMW.2018.8406869.

C. Joe-Wong, S. Sen, T. Lan, and M. Chiang. Multi-resource allocation: Fairness-efficiency tradeoffs in a unifying framework. In A. G. Greenberg and K. Sohraby, editors, *Proceedings of the IEEE INFOCOM 2012, Orlando, FL, USA, March 25-30, 2012*, pages 1206–1214. IEEE, 2012. doi: 10.1109/INFCOM.2012.6195481. URL https://doi.org/10.1109/INFCOM.2012.6195481.

References II

F. Kelly. Charging and rate control for elastic traffic. *Eur. Trans. Telecommun.*, 8(1):33–37, 1997. doi: 10.1002/ett.4460080106. URL https://doi.org/10.1002/ett.4460080106.

F. Kelly and E. Yudovina. *Stochastic networks*, volume 2. Cambridge University Press, 2014.

T. Lan, D. T. H. Kao, M. Chiang, and A. Sabharwal. An axiomatic theory of fairness in network resource allocation. In *INFOCOM 2010. 29th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies, 15-19 March 2010, San Diego, CA, USA*, pages 1343–1351. IEEE, 2010. doi: 10.1109/INFCOM.2010.5461911. URL https://doi.org/10.1109/INFCOM.2010.5461911.

J. Marašević, C. Stein, and G. Zussman. A fast distributed stateless algorithm for alpha-fair packing problems. In I. Chatzigiannakis, M. Mitzenmacher, Y. Rabani, and D. Sangiorgi, editors, *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, volume 55 of *LIPIcs*, pages 54:1–54:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2016. doi: 10.4230/LIPIcs.ICALP.2016.54. URL https://doi.org/10.4230/LIPIcs.ICALP.2016.54.

B. McCormick, F. Kelly, P. Plante, P. Gunning, and P. Ashwood-Smith. Real time alpha-fairness based traffic engineering. In A. Akella and A. G. Greenberg, editors, *Proceedings of the third workshop on Hot topics in software defined networking, HotSDN '14, Chicago, Illinois, USA, August 22, 2014*, pages 199–200. ACM, 2014. doi: 10.1145/2620728.2620762. URL https://doi.org/10.1145/2620728.2620762.

J. Mo and J. C. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Trans. Netw.*, 8(5):556–567, 2000. doi: 10.1109/90.879343. URL https://doi.org/10.1109/90.879343.

J. F. Nash. The bargaining problem. *Econometrica*, 18(2):155–162, 1950. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1907266.