Balanced words and related concepts: applications and complexity issues

Yves Crama HEC Management School, University of Liège, Belgium

Liège, September 2016



 An operations researcher's perspective on "regular" words and sequences. – Based on joint work with Nadia Brauner and Vincent Jost (Grenoble).

- An operations researcher's perspective on "regular" words and sequences. – Based on joint work with Nadia Brauner and Vincent Jost (Grenoble).
- OR specificity: decision-making problems
 - application driven (management, economics, social sciences, engineering, etc.)

- An operations researcher's perspective on "regular" words and sequences. – Based on joint work with Nadia Brauner and Vincent Jost (Grenoble).
- OR specificity: decision-making problems
 - application driven (management, economics, social sciences, engineering, etc.)
 - mathematical modeling

- An operations researcher's perspective on "regular" words and sequences. – Based on joint work with Nadia Brauner and Vincent Jost (Grenoble).
- OR specificity: decision-making problems
 - application driven (management, economics, social sciences, engineering, etc.)
 - mathematical modeling
 - algorithmically oriented

- An operations researcher's perspective on "regular" words and sequences. – Based on joint work with Nadia Brauner and Vincent Jost (Grenoble).
- OR specificity: decision-making problems
 - application driven (management, economics, social sciences, engineering, etc.)
 - mathematical modeling
 - algorithmically oriented

Outline



- Maximum deviation JIT scheduling
- 3 JIT and Balanced words
- 4 Balanced words
- 5 Extensions and related concepts
- 6 Short bibliography

Apportionment

- A House of representatives has 50 seats
- 37000 citizens elect representatives
- Party A gets 19000 votes, party B gets 13000, party C gets 5000
- How should each party be represented in the House?

Apportionment

- A House of representatives has 50 seats
- 37000 citizens elect representatives
- Party A gets 19000 votes, party B gets 13000, party C gets 5000
- How should each party be represented in the House?
- Ideally, the apportionment of seats should be as close as possible to the ratios ¹/₃₇(19, 13, 5), i.e., to (25.6757, 17.5676, 6.7568)

Apportionment

- A House of representatives has 50 seats
- 37000 citizens elect representatives
- Party A gets 19000 votes, party B gets 13000, party C gets 5000
- How should each party be represented in the House?
- Ideally, the apportionment of seats should be as close as possible to the ratios ¹/₃₇(19, 13, 5), i.e., to (25.6757, 17.5676, 6.7568)
- Well-studied problem (for a couple of centuries)
- Apportionment algorithms are mostly sequential allocation methods: first seat, then second one,...
- E.g., Webster's method for above example: A B A B A A B C A B A C...

Applications

Just-In-Time production scheduling

- p product types (A, B, C,...; red, blue, green,...)
- *n_i* items of type *i* = 1,...,*p*
- $n = \sum_{i} n_i$ = total number of items
- unit production times
- $r_i = \frac{n_i}{n}$ = proportion of items of type *i*
- *kr_i* = expected number of items of type *i* in the interval [1, *k*].

Applications

Just-In-Time production scheduling

- p product types (A, B, C,...; red, blue, green,...)
- *n_i* items of type *i* = 1,...,*p*
- $n = \sum_{i} n_i$ = total number of items
- unit production times
- $r_i = \frac{n_i}{n}$ = proportion of items of type *i*
- *kr_i* = expected number of items of type *i* in the interval [1, *k*].
- Determine a production schedule of all items such that, at every instant *k*, the number of items of type *i* that have been produced is as close as possible to *kr_i*.

JIT scheduling: Example



JIT scheduling

- Related to apportionment, but "sequencing" aspect is central.
- E.g., 8 parties get (8,1,1,1,1,1,1) votes. House size is 15.

JIT scheduling

- Related to apportionment, but "sequencing" aspect is central.
- E.g., 8 parties get (8,1,1,1,1,1,1) votes. House size is 15.
- Webster's apportionment method allocates A A A A B C D E F G H A A A A

JIT scheduling

- Related to apportionment, but "sequencing" aspect is central.
- E.g., 8 parties get (8,1,1,1,1,1,1) votes. House size is 15.
- Webster's apportionment method allocates A A A A B C D E F G H A A A A
- JIT scheduling would expect
 A B A C A D A E A F A G A H A

Outline

Applications

- 3 JIT and Balanced words
- 4 Balanced words
- 5 Extensions and related concepts
- 6 Short bibliography

- Steiner and Yeomans (1993): Optimization model
- $r_i = \frac{n_i}{n}$ = proportion of items of type *i*
- Determine a production schedule of all items such that, at every instant *k*, the number of items of type *i* that have been produced is as close as possible to *kr_i*

- Steiner and Yeomans (1993): Optimization model
- $r_i = \frac{n_i}{n}$ = proportion of items of type *i*
- Determine a production schedule of all items such that, at every instant *k*, the number of items of type *i* that have been produced is as close as possible to *kr_i*
- x_{ik} = number of items of type i = 1, ..., p produced up to time k = 1, ..., n
- (MDJIT) minimize $\max_{i,k} |x_{ik} kr_i|$

JIT scheduling: Example



- $r_i = \frac{n_i}{n}$ = proportion of items of type *i*
- x_{ik} = number of items of type i = 1, ..., p produced up to time k = 1, ..., n
- (MDJIT) minimize $\max_{i,k} |x_{ik} kr_i|$

- $r_i = \frac{n_i}{n}$ = proportion of items of type *i*
- x_{ik} = number of items of type i = 1, ..., p produced up to time k = 1, ..., n
- (MDJIT) minimize $\max_{i,k} |x_{ik} kr_i|$
- Thresholding approach: fix maximum allowed deviation, say, *B*.
- Decide whether one can produce the *j*-th item of type *i* at time *k* so that |*j* − *kr_i*| ≤ *B*, for all *i*, *j*, *k*.

- $r_i = \frac{n_i}{n}$ = proportion of items of type *i*
- *x_{ik}* = number of items of type *i* = 1,..., *p* produced up to time *k* = 1,..., *n*
- (MDJIT) minimize $\max_{i,k} |x_{ik} kr_i|$
- Thresholding approach: fix maximum allowed deviation, say, *B*.
- Decide whether one can produce the *j*-th item of type *i* at time *k* so that |*j* − *kr_i*| ≤ *B*, for all *i*, *j*, *k*.
- Bipartite matching model:
 - precompute the time-slots *k* to which *j*-th item of type *i* can be assigned so that $|j kr_i| \le B$
 - put an edge between (i, j) and k
 - determine whether the graph has a perfect matching

JIT scheduling: Bipartite graph

- 3 part types
- $n_1 = 3, n_2 = 3, n_3 = 1$
- n = 7 time slots
- *B* = 5/7



Binary search on B leads to O(n log n) algorithm for the MDJIT optimization problem

- Binary search on *B* leads to *O*(*n* log *n*) algorithm for the MDJIT optimization problem
- Pseudo-polynomial (input length is: $\log n_1 + \log n_2 + \ldots \log n_p$)

- Binary search on *B* leads to *O*(*n* log *n*) algorithm for the MDJIT optimization problem
- Pseudo-polynomial (input length is: $\log n_1 + \log n_2 + \ldots \log n_p$)
- Can we do better?

- Binary search on B leads to O(n log n) algorithm for the MDJIT optimization problem
- Pseudo-polynomial (input length is: $\log n_1 + \log n_2 + \ldots \log n_p$)
- Can we do better?
- Is the MDJIT problem in NP?

Algebraic characterization

Theorem (Brauner and Crama DAM 2004)

MDJIT has a solution with maximum deviation at most *B* if and only if the following hold for all $x_1, x_2 \in \{1, 2, ..., n\}$ with $x_1 \le x_2$: $\sum_i \max(0, \lfloor x_2 r_i + B \rfloor - \lceil (x_1 - 1)r_i - B \rceil) \ge x_2 - x_1 + 1$ $\sum_i \max(0, \lceil x_2 r_i - B \rceil - \lfloor (x_1 - 1)r_i + B \rfloor) \le x_2 - x_1 + 1$

Algebraic characterization

Theorem (Brauner and Crama DAM 2004)

MDJIT has a solution with maximum deviation at most *B* if and only if the following hold for all $x_1, x_2 \in \{1, 2, ..., n\}$ with $x_1 \le x_2$: $\sum_i \max(0, \lfloor x_2 r_i + B \rfloor - \lceil (x_1 - 1)r_i - B \rceil) \ge x_2 - x_1 + 1$ $\sum_i \max(0, \lceil x_2 r_i - B \rceil - \lfloor (x_1 - 1)r_i + B \rfloor) \le x_2 - x_1 + 1$

Corollary 1

MDJIT is in co-NP.

Algebraic characterization

Theorem (Brauner and Crama DAM 2004)

MDJIT has a solution with maximum deviation at most *B* if and only if the following hold for all $x_1, x_2 \in \{1, 2, ..., n\}$ with $x_1 \le x_2$: $\sum_i \max(0, \lfloor x_2 r_i + B \rfloor - \lceil (x_1 - 1)r_i - B \rceil) \ge x_2 - x_1 + 1$ $\sum_i \max(0, \lceil x_2 r_i - B \rceil - \lfloor (x_1 - 1)r_i + B \rfloor) \le x_2 - x_1 + 1$

Corollary 1

MDJIT is in co-NP.

Corollary 2

For fixed *p*, MDJIT can be solved in polynomial time.

- Proof. Express the NSC as linear inequalities in integer variables; use Lenstra's algorithm.
- Easy when p = 2. We know nothing smarter when $p \ge 3$

Bounds on the smallest deviation

Let $B^* = \min \max_{i,k} |x_{ik} - kr_i|$.

Bounds on the smallest deviation

Let $B^* = \min \max_{i,k} |x_{ik} - kr_i|$.

Corollary 3

For all instances of MDJIT, $B^* < 1$.

Bounds on the smallest deviation

Let $B^* = \min \max_{i,k} |x_{ik} - kr_i|$.

Corollary 3

For all instances of MDJIT, $B^* < 1$.

Corollary 4 (Jost 2006)

There always exists a 3-*balanced schedule* of items, i.e., a schedule such that the difference between the number of occurrences of parts of a same type in any two (time) intervals of the same length is at most 3.

- Recall: for $(n_1, n_2, \dots, n_8) = (8, 1, 1, 1, 1, 1, 1, 1)$,
- Webster's method yields A A A A B C D E F G H A A A A
- JIT scheduling yields A B A C A D A E A F A G A H A

Bounds on the smallest deviation

Let $B^* = \min \max_{i,k} |x_{ik} - kr_i|$.

Corollary 3

For all instances of MDJIT, $B^* < 1$.

Corollary 4 (Jost 2006)

There always exists a 3-balanced schedule of items, i.e., a schedule such that the difference between the number of occurrences of parts of a same type in any two (time) intervals of the same length is at most 3.

- Recall: for $(n_1, n_2, \dots, n_8) = (8, 1, 1, 1, 1, 1, 1, 1)$,
- Webster's method yields A A A A B C D E F G H A A A A
- JIT scheduling yields A B A C A D A E A F A G A H A
- The latter JIT schedule is *balanced* (difference is at most 1).

3-balance
- interval [s, t]: number of A's is in ($(t - s + 1) r_A - 2, (t - s + 1) r_A + 2$)

• interval [*s*, *t*]: number of A's is in

$$((t - s + 1) r_A - 2, (t - s + 1) r_A + 2)$$

- same holds for any other interval of length (t s)
- so, the number of A's differs by 3 units, at most.

Let $B^* = \min \max_{i,k} |x_{ik} - kr_i|$.

Let $B^* = \min \max_{i,k} |x_{ik} - kr_i|$.

• when does $B^* < \frac{1}{2}$ hold (meaning: $x_{ik} = [kr_i]$)?

Let $B^* = \min \max_{i,k} |x_{ik} - kr_i|$.

• when does $B^* < \frac{1}{2}$ hold (meaning: $x_{ik} = [kr_i]$)?

Note

When $B^* < \frac{1}{2}$, there exists a *balanced schedule* of items, i.e., a schedule such that the difference between the number of occurrences of parts of a same type in any two intervals of the same length is at most 1.

Let $B^* = \min \max_{i,k} |x_{ik} - kr_i|$.

• when does $B^* < \frac{1}{2}$ hold (meaning: $x_{ik} = [kr_i]$)?

Note

When $B^* < \frac{1}{2}$, there exists a *balanced schedule* of items, i.e., a schedule such that the difference between the number of occurrences of parts of a same type in any two intervals of the same length is at most 1.

More precisely:

- W: word (sequence of letters) associated with a schedule
- W^* : infinite word (W, W, ...) obtained by repeating W indefinitely

More precisely:

- W: word (sequence of letters) associated with a schedule
- W^* : infinite word (W, W, \ldots) obtained by repeating W indefinitely

Theorem (Crama and Brauner (2004))

If *W* is a schedule with $B(W) < \frac{1}{2}$ for the instance $(n_1, n_2, ..., n_p)$, then the word *W*^{*} is *balanced* and all numbers n_i are pairwise distinct.

More precisely:

- W: word (sequence of letters) associated with a schedule
- W^* : infinite word (W, W, \ldots) obtained by repeating W indefinitely

Theorem (Crama and Brauner (2004))

If *W* is a schedule with $B(W) < \frac{1}{2}$ for the instance $(n_1, n_2, ..., n_p)$, then the word *W*^{*} is *balanced* and all numbers n_i are pairwise distinct.

Conjecture (Crama and Brauner (2004))

If *W* is a schedule with $B(W) < \frac{1}{2}$ for the instance $(n_1, n_2, ..., n_p)$, then $(n_1, n_2, ..., n_p) = (2^{p-1}, 2^{p-2}, ..., 1)$.

More precisely:

- W: word (sequence of letters) associated with a schedule
- W^* : infinite word (W, W, \ldots) obtained by repeating W indefinitely

Theorem (Crama and Brauner (2004))

If *W* is a schedule with $B(W) < \frac{1}{2}$ for the instance $(n_1, n_2, ..., n_p)$, then the word *W*^{*} is *balanced* and all numbers n_i are pairwise distinct.

Conjecture (Crama and Brauner (2004))

If *W* is a schedule with $B(W) < \frac{1}{2}$ for the instance $(n_1, n_2, ..., n_p)$, then $(n_1, n_2, ..., n_p) = (2^{p-1}, 2^{p-2}, ..., 1)$.

• Proved by Kubiak (2003), Brauner and Jost (2008)

More precisely:

- W: word (sequence of letters) associated with a schedule
- W^* : infinite word (W, W, \ldots) obtained by repeating W indefinitely

Theorem (Crama and Brauner (2004))

If *W* is a schedule with $B(W) < \frac{1}{2}$ for the instance $(n_1, n_2, ..., n_p)$, then the word *W*^{*} is *balanced* and all numbers n_i are pairwise distinct.

Conjecture (Crama and Brauner (2004))

If *W* is a schedule with $B(W) < \frac{1}{2}$ for the instance $(n_1, n_2, ..., n_p)$, then $(n_1, n_2, ..., n_p) = (2^{p-1}, 2^{p-2}, ..., 1)$.

- Proved by Kubiak (2003), Brauner and Jost (2008)
- Nice connections with Fraenkel's conjecture on balanced words.

Outline



- Maximum deviation JIT scheduling
- IT and Balanced words
 - 4 Balanced words
 - 5 Extensions and related concepts
 - 6 Short bibliography

Definitions

Sequences and words

A sequence is a subset of \mathbb{Z} . A word on *p* letters (or colors) is a partition of \mathbb{Z} into *p* sequences S_1, \ldots, S_p or, equivalently, a mapping $\mathbb{Z} \to \{1, 2, \ldots, p\}$.

Definitions

Sequences and words

A sequence is a subset of \mathbb{Z} . A word on *p* letters (or colors) is a partition of \mathbb{Z} into *p* sequences S_1, \ldots, S_p or, equivalently, a mapping $\mathbb{Z} \to \{1, 2, \ldots, p\}$.

Balanced sequences

A balanced sequence is a sequence *S* such that, for every two intervals I_1 and I_2 of the same length, the difference between the number of elements of the sequence in the two intervals is at most 1: that is, if $I_1 = \{i_1, \ldots, i_1 + t\}$ and $I_2 = \{i_2, \ldots, i_2 + t\}$, then $-1 \le |I_1 \cap S| - |I_2 \cap S| \le 1$.

Definitions

Sequences and words

A sequence is a subset of \mathbb{Z} . A word on *p* letters (or colors) is a partition of \mathbb{Z} into *p* sequences S_1, \ldots, S_p or, equivalently, a mapping $\mathbb{Z} \to \{1, 2, \ldots, p\}$.

Balanced sequences

A balanced sequence is a sequence *S* such that, for every two intervals I_1 and I_2 of the same length, the difference between the number of elements of the sequence in the two intervals is at most 1: that is, if $I_1 = \{i_1, \ldots, i_1 + t\}$ and $I_2 = \{i_2, \ldots, i_2 + t\}$, then $-1 \le |I_1 \cap S| - |I_2 \cap S| \le 1$.

Balanced wordss

A word *W* is *balanced* if all its associated sequences S_i , $i \in \{1, ..., p\}$ are balanced.

Balanced words

Examples:

• abacaba abacaba ...

Balanced words

Examples:

- abacaba abacaba ...
- abacaba abacaba ...: balanced

Balanced words

Examples:

- abacaba abacaba ...
- abacaba abacaba ...: balanced
- abacabababacaaba...
- abacabababacaaba...: not balanced

Densities

Densities

Every balanced word has a *density vector* δ , where δ_i , the *density* of letter *i*, is the limit, when $t \to \infty$, of the proportion of occurrences of letter *i* in the interval $\{1, \ldots, t\}$.

Densities

Densities

Every balanced word has a *density vector* δ , where δ_i , the *density* of letter *i*, is the limit, when $t \to \infty$, of the proportion of occurrences of letter *i* in the interval $\{1, \ldots, t\}$.

Example:

• abacaba abacaba ... $\delta = (\frac{4}{7}, \frac{2}{7}, \frac{1}{7})$

Fraenkel's conjecture (1973)

For all $p \ge 3$, W^p is a balanced word on p letters such that all components of its density vector δ^p are pairwise distinct if and only if

$$\delta_i^p = \frac{2^{p-i}}{2^p - 1}, \quad i = 1, \dots, p.$$

Fraenkel's conjecture (1973)

For all $p \ge 3$, W^p is a balanced word on p letters such that all components of its density vector δ^p are pairwise distinct if and only if

$$\delta_i^p = \frac{2^{p-i}}{2^p - 1}, \quad i = 1, \dots, p.$$

• W^{p} is of the form $(F^{p})^{*}$ where $F^{p} = (F^{p-1}, p, F^{p-1})$.

Fraenkel's conjecture (1973)

For all $p \ge 3$, W^p is a balanced word on p letters such that all components of its density vector δ^p are pairwise distinct if and only if

$$\delta_i^p = \frac{2^{p-i}}{2^p - 1}, \quad i = 1, \dots, p.$$

W^ρ is of the form (*F^ρ*)* where *F^ρ* = (*F^{ρ-1}*, *p*, *F^{ρ-1}*). *p* = 3: (aba c aba)*; δ³ = (⁴/₇, ²/₇, ¹/₇).

Fraenkel's conjecture (1973)

For all $p \ge 3$, W^p is a balanced word on p letters such that all components of its density vector δ^p are pairwise distinct if and only if

$$\delta_i^p = \frac{2^{p-i}}{2^p - 1}, \quad i = 1, \dots, p.$$

- W^{p} is of the form $(F^{p})^{*}$ where $F^{p} = (F^{p-1}, p, F^{p-1})$.
- p = 3: (aba c aba)*; $\delta^3 = (\frac{4}{7}, \frac{2}{7}, \frac{1}{7})$.
- p = 4: (aba c aba d aba c aba)*; $\delta^4 = (\frac{8}{15}, \frac{4}{15}, \frac{2}{15}, \frac{1}{15})$.

Fraenkel's conjecture (1973)

For all $p \ge 3$, W^p is a balanced word on p letters such that all components of its density vector δ^p are pairwise distinct if and only if

$$\delta_i^p = \frac{2^{p-i}}{2^p - 1}, \quad i = 1, \dots, p.$$

- W^{p} is of the form $(F^{p})^{*}$ where $F^{p} = (F^{p-1}, p, F^{p-1})$.
- p = 3: (aba c aba)*; $\delta^3 = (\frac{4}{7}, \frac{2}{7}, \frac{1}{7})$.
- p = 4: (aba c aba d aba c aba)*; $\delta^4 = (\frac{8}{15}, \frac{4}{15}, \frac{2}{15}, \frac{1}{15})$.
- F^{p} has length $2^{p} 1$, and the letter frequencies are $(2^{p-1}, 2^{p-2}, \dots, 1)$.

JIT and Balanced words

Links with MDJIT

• Theorem (C&B (2004)): If *W* is a schedule with $B(W) < \frac{1}{2}$ for the instance $(n_1, n_2, ..., n_p)$, then W^* is balanced and all numbers n_i are pairwise distinct.

- Theorem (C&B (2004)): If W is a schedule with B(W) < ¹/₂ for the instance (n₁, n₂,..., n_p), then W* is balanced and all numbers n_i are pairwise distinct.
- Conjecture (C&B (2004)): If *W* is a schedule with $B(W) < \frac{1}{2}$ for the instance $(n_1, n_2, ..., n_p)$, then $n_i = 2^{p-i}$ for all i = 1, ..., p.

- Theorem (C&B (2004)): If W is a schedule with B(W) < ¹/₂ for the instance (n₁, n₂,..., n_p), then W* is balanced and all numbers n_i are pairwise distinct.
- Conjecture (C&B (2004)): If *W* is a schedule with $B(W) < \frac{1}{2}$ for the instance $(n_1, n_2, ..., n_p)$, then $n_i = 2^{p-i}$ for all i = 1, ..., p.
- Conjecture would follow from Fraenkel's conjecture.

- Theorem (C&B (2004)): If W is a schedule with B(W) < ¹/₂ for the instance (n₁, n₂,..., n_p), then W* is balanced and all numbers n_i are pairwise distinct.
- Conjecture (C&B (2004)): If *W* is a schedule with $B(W) < \frac{1}{2}$ for the instance (n_1, n_2, \dots, n_p) , then $n_i = 2^{p-i}$ for all $i = 1, \dots, p$.
- Conjecture would follow from Fraenkel's conjecture.

Theorem (Brauner and Jost (2008))

If *W* is a schedule with $B(W) < \frac{1}{2}$ for the instance $(n_1, n_2, ..., n_p)$, then *W* is *symmetric*.

- Theorem (C&B (2004)): If W is a schedule with B(W) < ¹/₂ for the instance (n₁, n₂,..., n_p), then W* is balanced and all numbers n_i are pairwise distinct.
- Conjecture (C&B (2004)): If *W* is a schedule with $B(W) < \frac{1}{2}$ for the instance (n_1, n_2, \ldots, n_p) , then $n_i = 2^{p-i}$ for all $i = 1, \ldots, p$.
- Conjecture would follow from Fraenkel's conjecture.

Theorem (Brauner and Jost (2008))

If *W* is a schedule with $B(W) < \frac{1}{2}$ for the instance $(n_1, n_2, ..., n_p)$, then *W* is *symmetric*.

Theorem (Symmetric case of Fraenkel's conjecture; B&J (2008))

For all $p \ge 3$, W^p is a symmetric and balanced word on p letters such that all components of its density vector δ^p are pairwise distinct if and only if $\delta_i^p = \frac{2^{p-i}}{2^p-1}$ for all i = 1, ..., p.

Summary

- JIT scheduling asks for "regular" scheduling of item types with given densities.
- MDJIT asks for a schedule minimizing the maximum deviation from "ideal frequencies" *kr_i*.
- MDJIT is in co-NP.
- MDJIT can be solved in pseudo-polynomial time, and even in polynomial time when p is fixed.
- Complexity is unknown in general.
- Optimal schedules are almost balanced (3-balanced).
- When B^{*} < ¹/₂, the optimal schedule is balanced. But this is a rare instance.
- What about 2-balanced schedules?

Extensions

- JIT scheduling asks for "regular" scheduling of item types with given densities.
- MDJIT asks for a schedule minimizing the maximum deviation from "ideal frequencies" *kr_i*.
- JIT scheduling asks for "regular" scheduling of item types with given densities.
- MDJIT asks for a schedule minimizing the maximum deviation from "ideal frequencies" *kr_i*.
- These are just models! Very meaningful to ask for balanced sequences when they exist.

- JIT scheduling asks for "regular" scheduling of item types with given densities.
- MDJIT asks for a schedule minimizing the maximum deviation from "ideal frequencies" *kr_i*.
- These are just models! Very meaningful to ask for balanced sequences when they exist.

More generally:

Given a vector δ in ℝ^p, can we decide (efficiently) whether there exists a "nicely regular" word with density δ?

- JIT scheduling asks for "regular" scheduling of item types with given densities.
- MDJIT asks for a schedule minimizing the maximum deviation from "ideal frequencies" *kr_i*.
- These are just models! Very meaningful to ask for balanced sequences when they exist.

More generally:

- Given a vector δ in ℝ^p, can we decide (efficiently) whether there exists a "nicely regular" word with density δ?
- Can we better understand the structure of such "nicely regular" words?

- JIT scheduling asks for "regular" scheduling of item types with given densities.
- MDJIT asks for a schedule minimizing the maximum deviation from "ideal frequencies" *kr_i*.
- These are just models! Very meaningful to ask for balanced sequences when they exist.

More generally:

- Given a vector δ in ℝ^p, can we decide (efficiently) whether there exists a "nicely regular" word with density δ?
- Can we better understand the structure of such "nicely regular" words?
- How do we define a "nicely regular" word??

Questions, questions,...

Outline



- 2 Maximum deviation JIT scheduling
- 3 JIT and Balanced words
- 4 Balanced words
- 5 Extensions and related concepts
- 6 Short bibliography

What words are balanced?

Balanceable vectors

Vector $\delta \in \mathbb{R}^{p}$ is *balanceable* if there exists a balanced word on p letters with density vector δ .

• $(\frac{4}{7}, \frac{2}{7}, \frac{1}{7})$ is a balanceable vector.

What words are balanced?

Balanceable vectors

Vector $\delta \in \mathbb{R}^{p}$ is *balanceable* if there exists a balanced word on p letters with density vector δ .

•
$$\left(\frac{4}{7}, \frac{2}{7}, \frac{1}{7}\right)$$
 is a balanceable vector.

Question:

Can we characterize all balanceable vectors?

- Probably very ambitious, so let's start slowly...
- What is known already?

Balanceable vectors

On 2 letters:

 $(\alpha, 1 - \alpha)$ is balanceable for all $0 < \alpha < 1$.

Balanceable vectors

On 2 letters:

 $(\alpha, 1 - \alpha)$ is balanceable for all $0 < \alpha < 1$.

On more than 2 letters:

Much more complex

Congruence sequence

 $S_i = \{a_i n + b_i : n \in \mathbb{Z}\}$ with a_i, b_i integers.

Congruence sequence

 $S_i = \{a_i n + b_i : n \in \mathbb{Z}\}$ with a_i, b_i integers.

Congruence word

A word consisting of congruence sequences S_1, S_2, \ldots, S_p . The density of letter *i* is $1/a_i$.

Example: $W = (abacabad)^*$

- Positions of *a* : 1, 3, 5, 7, ... = 2*n* + 1
- Positions of b : 2, 6, 10... = 4n + 2
- Positions of *c* : 4, 12, 20, ... = 8*n* + 4

Congruence sequence

 $S_i = \{a_i n + b_i : n \in \mathbb{Z}\}$ with a_i, b_i integers.

Congruence word

A word consisting of congruence sequences S_1, S_2, \ldots, S_p . The density of letter *i* is $1/a_i$.

Example: $W = (abacabad)^*$

- Positions of *a* : 1, 3, 5, 7, ... = 2*n* + 1
- Positions of *b* : 2, 6, 10... = 4*n* + 2
- Positions of *c* : 4, 12, 20, ... = 8*n* + 4

Property

Every congruence word is balanced.

Congruence substitution $W_{A,j}$

given a word W, a *congruence word* A, and a letter j of W, replace the k-th occurrence of j in W by k-th letter of A, cyclically.

```
W = (abacaba)^* and A = (de)^*
W_{A,b} =
```

Congruence substitution $W_{A,i}$

given a word W, a *congruence word* A, and a letter j of W, replace the k-th occurrence of j in W by k-th letter of A, cyclically.

 $W = (abacaba)^*$ and $A = (de)^*$ $W_{A,b} = (adacaea)^*$

Congruence substitution $W_{A,j}$

given a word W, a *congruence word* A, and a letter j of W, replace the k-th occurrence of j in W by k-th letter of A, cyclically.

 $W = (abacaba)^*$ and $A = (de)^*$ $W_{A,b} = (adacaea)^*$ $W_{A,c} = (abadabaabaeaba)^*$

Congruence substitution $W_{A,j}$

given a word W, a *congruence word* A, and a letter j of W, replace the k-th occurrence of j in W by k-th letter of A, cyclically.

 $W = (abacaba)^*$ and $A = (de)^*$ $W_{A,b} = (adacaea)^*$ $W_{A,c} = (abadabaabaeaba)^*$

Congruence expansion

A congruence expansion of a word W is the result of iterative applications of congruence substitutions on W.

Congruence substitution $W_{A,j}$

given a word W, a *congruence word* A, and a letter j of W, replace the k-th occurrence of j in W by k-th letter of A, cyclically.

 $W = (abacaba)^*$ and $A = (de)^*$ $W_{A,b} = (adacaea)^*$ $W_{A,c} = (abadabaabaeaba)^*$

Congruence expansion

A congruence expansion of a word W is the result of iterative applications of congruence substitutions on W.

Property.

Every congruence expansion of a balanced word is balanced.

• How general is this construction?

Proposition (Hubert 2000)

A word with *irrational densities* is balanced if and only if it is a congruence expansion of a balanced word on two letters. These words are non-periodic.

Proposition (Hubert 2000)

A word with *irrational densities* is balanced if and only if it is a congruence expansion of a balanced word on two letters. These words are non-periodic.

• Algorithmically, irrational densities are not really relevant.

Proposition (Hubert 2000)

A word with *irrational densities* is balanced if and only if it is a congruence expansion of a balanced word on two letters. These words are non-periodic.

- Algorithmically, irrational densities are not really relevant.
- What about rational density vectors?

Proposition (Hubert 2000)

A word with *irrational densities* is balanced if and only if it is a congruence expansion of a balanced word on two letters. These words are non-periodic.

- Algorithmically, irrational densities are not really relevant.
- What about rational density vectors?

Experiment:

for small *p* and *D*, generate all balanceable vectors with rational densities of the form $\left(\frac{d_1}{D}, \frac{d_2}{D}, \dots, \frac{d_p}{D}\right)$.

In all cases, the balanceable vectors on *p* letters fall into one of the following classes:

- density vectors of congruence expansions of balanced words on fewer letters
- Fraenkel density $\left(\frac{2^{p-1}}{2^p-1}, \frac{2^{p-2}}{2^p-1}, \dots, \frac{1}{2^p-1}\right)$
- and not much more...

N = 3 : (α/2, α/2, 1 - α), for all 0 < α < 1, and (⁴/₇, ²/₇, ¹/₇) (this is the complete list)

- N = 3 : (α/2, α/2, 1 α), for all 0 < α < 1, and (⁴/₇, ²/₇, ¹/₇) (this is the complete list)
- N = 4, results for $D \le 200$; sporadic cases:

```
 \begin{pmatrix} \frac{6}{11}, \frac{3}{11}, \frac{1}{11}, \frac{1}{11} \end{pmatrix} \begin{pmatrix} \frac{6}{11}, \frac{2}{11}, \frac{2}{11}, \frac{1}{11} \end{pmatrix} \begin{pmatrix} \frac{4}{11}, \frac{4}{11}, \frac{2}{11}, \frac{1}{11} \end{pmatrix} \begin{pmatrix} \frac{8}{14}, \frac{4}{14}, \frac{1}{14}, \frac{1}{14} \end{pmatrix} \\ \begin{pmatrix} \frac{8}{15}, \frac{4}{15}, \frac{2}{15}, \frac{1}{15} \end{pmatrix}
```

(proved to be complete)

- N = 3 : (α/2, α/2, 1 α), for all 0 < α < 1, and (⁴/₇, ²/₇, ¹/₇) (this is the complete list)
- N = 4, results for $D \le 200$; sporadic cases:

 $\begin{pmatrix} \frac{6}{11}, \frac{3}{11}, \frac{1}{11}, \frac{1}{11} \end{pmatrix} \begin{pmatrix} \frac{6}{11}, \frac{2}{11}, \frac{2}{11}, \frac{1}{11} \end{pmatrix} \begin{pmatrix} \frac{4}{11}, \frac{4}{11}, \frac{2}{11}, \frac{1}{11} \end{pmatrix} \begin{pmatrix} \frac{8}{14}, \frac{4}{14}, \frac{1}{14}, \frac{1}{14} \end{pmatrix} \\ \begin{pmatrix} \frac{8}{15}, \frac{4}{15}, \frac{2}{15}, \frac{1}{15} \end{pmatrix}$

(proved to be complete)

• N = 5, results for $D \le 130$; sporadic cases:

$$\begin{pmatrix} \frac{8}{13}, \frac{2}{13}, \frac{1}{13}, \frac{1}{13}, \frac{1}{13} \end{pmatrix} \begin{pmatrix} \frac{6}{17}, \frac{6}{17}, \frac{2}{17}, \frac{2}{17}, \frac{1}{17} \end{pmatrix} \begin{pmatrix} \frac{12}{23}, \frac{6}{23}, \frac{3}{23}, \frac{1}{23}, \frac{1}{23} \end{pmatrix} \\ \begin{pmatrix} \frac{6}{13}, \frac{3}{13}, \frac{2}{13}, \frac{1}{13}, \frac{1}{13} \end{pmatrix} \begin{pmatrix} \frac{9}{17}, \frac{3}{17}, \frac{3}{17}, \frac{1}{17}, \frac{1}{17} \end{pmatrix} \begin{pmatrix} \frac{12}{23}, \frac{6}{23}, \frac{2}{23}, \frac{2}{23}, \frac{1}{23} \end{pmatrix} \\ \begin{pmatrix} \frac{4}{13}, \frac{3}{13}, \frac{2}{13}, \frac{2}{13}, \frac{2}{13}, \frac{2}{13} \end{pmatrix} \begin{pmatrix} \frac{6}{17}, \frac{6}{17}, \frac{3}{17}, \frac{1}{17}, \frac{1}{17} \end{pmatrix} \begin{pmatrix} \frac{16}{31}, \frac{8}{31}, \frac{4}{31}, \frac{2}{31}, \frac{1}{31} \end{pmatrix}$$

• N = 6, test exhaustif pour $D \le 80$; sporadic cases:

 $\left(\frac{5}{13}, \frac{3}{13}, \frac{2}{13}, \frac{1}{13}, \frac{1}{13}, \frac{1}{13}, \frac{1}{13}\right) \left(\frac{10}{19}, \frac{5}{19}, \frac{1}{19}, \frac{1}{19}, \frac{1}{19}, \frac{1}{19}\right) \left(\frac{8}{21}, \frac{8}{21}, \frac{2}{21}, \frac{1}{21}, \frac{1}{21}, \frac{1}{21}\right)$ $\left(\frac{12}{35}, \frac{12}{35}, \frac{6}{35}, \frac{2}{35}, \frac{2}{35}, \frac{1}{35}\right) \left(\frac{9}{16}, \frac{3}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}\right) \left(\frac{10}{19}, \frac{3}{19}, \frac{2}{19}, \frac{2}{19}, \frac{1}{19}, \frac{1}{19}\right)$ $\left(\frac{12}{25}, \frac{6}{25}, \frac{3}{25}, \frac{2}{25}, \frac{1}{25}, \frac{1}{25}\right) \left(\frac{12}{35}, \frac{12}{35}, \frac{4}{35}, \frac{4}{35}, \frac{2}{35}, \frac{1}{35}\right) \left(\frac{10}{17}, \frac{2}{17}, \frac{2}{17}, \frac{1}{17}, \frac{1}{17}, \frac{1}{17}\right)$ $\left(\frac{10}{19}, \frac{2}{19}, \frac{2}{19}, \frac{2}{19}, \frac{2}{19}, \frac{1}{19}\right) \left(\frac{9}{26}, \frac{9}{26}, \frac{3}{26}, \frac{3}{26}, \frac{3}{26}, \frac{1}{26}, \frac{1}{26}\right) \left(\frac{24}{47}, \frac{12}{47}, \frac{6}{47}, \frac{3}{47}, \frac{1}{47}, \frac{1}{47}\right)$ $\left(\frac{9}{17}, \frac{3}{17}, \frac{2}{17}, \frac{1}{17}, \frac{1}{17}, \frac{1}{17}\right) \left(\frac{9}{19}, \frac{3}{19}, \frac{3}{19}, \frac{2}{19}, \frac{1}{19}, \frac{1}{19}\right) \left(\frac{18}{35}, \frac{9}{35}, \frac{3}{35}, \frac{3}{35}, \frac{1}{35}, \frac{1}{35}\right)$ $\left(\frac{24}{47},\frac{12}{47},\frac{6}{47},\frac{2}{47},\frac{2}{47},\frac{1}{47}\right)\left(\frac{8}{17},\frac{3}{17},\frac{2}{17},\frac{2}{17},\frac{1}{17},\frac{1}{17}\right)\left(\frac{6}{19},\frac{6}{19},\frac{3}{19},\frac{2}{19},\frac{1}{19},\frac{1}{19}\right)$ $\left(\frac{18}{35}, \frac{6}{35}, \frac{3}{35}, \frac{3}{35}, \frac{1}{35}, \frac{1}{35}\right) \left(\frac{24}{47}, \frac{12}{47}, \frac{4}{47}, \frac{4}{47}, \frac{2}{47}, \frac{1}{47}\right) \left(\frac{6}{17}, \frac{4}{17}, \frac{3}{17}, \frac{2}{17}, \frac{1}{17}, \frac{1}{17}\right)$ $\left(\frac{6}{19}, \frac{4}{19}, \frac{4}{19}, \frac{2}{19}, \frac{2}{19}, \frac{1}{19}\right) \left(\frac{18}{35}, \frac{6}{35}, \frac{6}{35}, \frac{2}{35}, \frac{2}{35}, \frac{1}{35}\right) \left(\frac{24}{47}, \frac{8}{47}, \frac{8}{47}, \frac{4}{47}, \frac{2}{47}, \frac{1}{47}\right)$ $\left(\frac{6}{17}, \frac{4}{17}, \frac{2}{17}, \frac{2}{17}, \frac{2}{17}, \frac{1}{17}\right) \left(\frac{4}{19}, \frac{4}{19}, \frac{4}{19}, \frac{4}{19}, \frac{2}{19}, \frac{1}{19}\right) \left(\frac{12}{35}, \frac{12}{35}, \frac{6}{35}, \frac{3}{35}, \frac{1}{35}, \frac{1}{35}\right)$ $\left(\frac{16}{47}, \frac{16}{47}, \frac{8}{47}, \frac{4}{47}, \frac{2}{47}, \frac{1}{47}\right)\left(\frac{4}{17}, \frac{4}{17}, \frac{3}{17}, \frac{2}{17}, \frac{2}{17}, \frac{2}{17}\right)\left(\frac{32}{63}, \frac{16}{63}, \frac{8}{63}, \frac{4}{63}, \frac{2}{63}, \frac{1}{63}\right)$

Which leads us to...

Conjecture (Brauner, Crama, Jost (2013))

Besides well-identified infinite families, there only exists a finite number of balanceable vectors for each *p*.

Which leads us to ...

Conjecture (Brauner, Crama, Jost (2013))

Besides well-identified infinite families, there only exists a finite number of balanceable vectors for each *p*.

More precisely:

Conjecture

If a word W on p letters is balanced, then

- (1) *W* is a *congruence expansion* of a balanced word on two letters, or
- (2) W is D-periodical for some $D \le 2^p 1$.
 - For irrational densities, Condition (1) holds.
 - Condition (2) implies that the number of "sporadic cases" is finite for each p.
 - Proof for *p* ≥ 5??

More algorithmic questions

 How difficult is it to recognize whether a vector δ is the density vector of a balanced word?

More algorithmic questions

- How difficult is it to recognize whether a vector δ is the density vector of a balanced word?
- How difficult is it to recognize whether a vector δ is the density vector of a congruence word?

More algorithmic questions

- How difficult is it to recognize whether a vector δ is the density vector of a balanced word?
- How difficult is it to recognize whether a vector δ is the density vector of a congruence word?
- Example: $\delta = \frac{1}{15}(5, 5, 3, 3, 3, 2, 2, 1, 1, 1, 1, 1, 1, 1)$ is the density of the congruence word

132465172xx31425xx162374152xxx.

But when applied to δ , classical methods for building "regular schedules" (e.g., MDJIT algorithms, or Webster's method of divisors) do not produce a congruence word.

 Given p congruence sequences S(a_i, b_i), how difficult is it to recognize whether they form a congruence word, i.e, whether they partition Z ?

Polynomial for fixed p. In NP otherwise. Hard?

Outline

Applications

- 2 Maximum deviation JIT scheduling
- 3 JIT and Balanced words
- 4 Balanced words
- 5 Extensions and related concepts
- 6 Short bibliography

m-balanced words

Various related notions have been considered in the OR literature.

m-balanced words

Various related notions have been considered in the OR literature.

- **Definition.** A word *W* is *m*-balanced if, for all *i* and all *t*, every subword of *W* of length *t* contains the same number of occurrences of letter *i*, up to *m* units.
- 1-balanced \equiv balanced.

m-balanced words

Various related notions have been considered in the OR literature.

- **Definition.** A word *W* is *m*-balanced if, for all *i* and all *t*, every subword of *W* of length *t* contains the same number of occurrences of letter *i*, up to *m* units.
- 1-balanced \equiv balanced.
- How difficult is it to recognize whether a vector δ is the density vector of an *m*-balanced word? (or to minimize *m*?)
- Proposition: For every rational vector δ, there exists a 3-balanced word with density δ.
 (Follows from B* < 1 for the MDJIT problem.)
m-balanced words

Various related notions have been considered in the OR literature.

- **Definition.** A word *W* is *m*-balanced if, for all *i* and all *t*, every subword of *W* of length *t* contains the same number of occurrences of letter *i*, up to *m* units.
- 1-balanced \equiv balanced.
- How difficult is it to recognize whether a vector δ is the density vector of an *m*-balanced word? (or to minimize *m*?)
- Proposition: For every rational vector δ, there exists a 3-balanced word with density δ.
 (Follows from B* < 1 for the MDJIT problem.)
- Question: What about 2-balance??

- **Definition.** A *tree word* (or tree schedule) *W* is recursively built as follows:
 - start with the constant word $W = (a)^*$
 - in the current word W, pick a letter j and substitute it by a congruence word of the form (a₁...a_k)* for some integer k.

- **Definition.** A *tree word* (or tree schedule) *W* is recursively built as follows:
 - start with the constant word $W = (a)^*$
 - in the current word W, pick a letter j and substitute it by a congruence word of the form (a₁...a_k)* for some integer k.
- Tree words are a subclass of congruence words.

- **Definition.** A *tree word* (or tree schedule) *W* is recursively built as follows:
 - start with the constant word $W = (a)^*$
 - in the current word W, pick a letter j and substitute it by a congruence word of the form (a₁...a_k)* for some integer k.
- Tree words are a subclass of congruence words.
- How difficult is it to recognize whether a vector δ is the density vector of a tree word?

- **Definition.** A *tree word* (or tree schedule) *W* is recursively built as follows:
 - start with the constant word $W = (a)^*$
 - in the current word W, pick a letter j and substitute it by a congruence word of the form (a₁...a_k)* for some integer k.
- Tree words are a subclass of congruence words.
- How difficult is it to recognize whether a vector δ is the density vector of a tree word?

Conclusions

- Many interesting (and hard) questions relating to the structure of "almost regular" words and of their density vectors.
- More fundamentally: what is the "right" notion of regularity? *m*-balance (different versions), congruence words, weighted measure of deviation, etc.
- Other untouched connections: apportionment problems, queueing, Beatty sequences, billiard words, etc.
- Recognition problems: given a vector δ, decide whether δ is the density of a "regular" word.
- Optimization problems: given a vector δ, find a "regular" word whose density is as close as possible to δ.

Outline

Applications

- 2 Maximum deviation JIT scheduling
- 3 JIT and Balanced words
- 4 Balanced words
- 5 Extensions and related concepts
- 6 Short bibliography



E. Altman, B. Gaujal, and A. Hordijk. Balanced sequences and optimal routing. *Journal of the ACM* 47 (1999) 752–775.

M. Balinski and V. Ramirez,

Parametric methods of apportionment, rounding and production, *Mathematical Social Sciences* 37 (1999) 107-122.



A. Bar-Noy, V. Dreizin and B. Patt-Shamir, Efficient algorithms for periodic scheduling, *Computer Networks* 45 (2004) 155-173.



N. Brauner and Y. Crama,

The maximum deviation just-in-time scheduling problem, *Discrete Applied Mathematics* 134 (2004) 25-50.



N. Brauner and V. Jost,

Small deviations, JIT sequencing and symmetric case of Fraenkel's conjecture, *Discrete Mathematics* 308 (2008) 2319–2324.

A.S. Fraenkel,

Complementing and exactly covering sequences, Journal of Combinatorial Theory (Ser. A) 14 (1973) 8–20.



J.W. Herrmann,

Finding optimally balanced words for production planning and maintenance scheduling,

IIE Transactions 44 (2012), 215-229.



P. Hubert,

Suites équilibrées,

Theoretical Computer Science 242 (2000), 91–108.



V. Jost,

Ordonnancement chromatique; polyèdres, complexité et classification, Thèse de doctorat, Grenoble, 2006.



W. Kubiak,

Fair sequences,

in: J.Y-T. Leung, ed., *Handbook of Scheduling: Algorithms, Models and Performance Analysis*, Chapman & Hall/CRC, Boca Raton, FL, 2004, pp. 19-1–19–21.

W. Kubiak,

Proportional Optimization and Fairness, Springer, New York, NY, 2009.

S. Sano, N. Miyoshi and R. Kataoka,

m-Balanced words: a generalization of balanced words, *Theoretical Computer Science* 314(1-2) (2004) 97–120.

G. Steiner and J.S. Yeomans,

Level schedules for mixed-model, just-in-time processes, *Management Science* 39 (1993) 728–735.



R. Tijdeman,

Fraenkel's conjecture for six sequences,

Discrete Mathematics 222 (2000) 223–234.

R. Tijdeman,

Periodicity and almost-periodicity,

in: E. Győry, G.O.H. Katona and L. Lovász, eds., *More Sets, Graphs and Numbers,* Springer and Bolyai Mathematical Society, Berlin Heidelberg New York Budapest, 2006, pp. 381-405.

L. Vuillon,

Balanced words,

Bulletin of the Belgian Mathematical Society Simon Stevin 10 (2003) 787-805.