First-order methods for the impatient: support identification in finite time with convergent Frank-Wolfe variants

Immanuel Bomze, University of Vienna

joint work with:

Francesco Rinaldi and **Samuel Rota Bulò**

GDO 2019 @ UBB

Cluj-Napoca, 10 April 2019

Overview

1. **Simplex-constrained smooth optimization**

Overview

1. Simplex-constrained smooth optimization

2. Frank-Wolfe method with away steps

Overview

1. Simplex-constrained smooth optimization
2. Frank-Wolfe method with away steps
3. Global convergence of iterates

Overview

1. Simplex-constrained smooth optimization
2. Frank-Wolfe method with away steps
3. Global convergence of iterates
4. Support identification in finite time

A simple optimization model ...

... is $\min_{\mathbf{x}\in\Delta} f(\mathbf{x})$ with $\Delta = \left\{ \mathbf{x} \in \mathbb{R}^n_+ : \sum_i x_i = 1 \right\}$

and $f \in \mathcal{C}^2$.

A simple optimization model ...

... is $\min_{\mathbf{x}\in\Delta} f(\mathbf{x})$ with $\Delta = \left\{\mathbf{x} \in \mathbb{R}^n_+ : \sum_i x_i = 1\right\}$

and $f \in \mathcal{C}^2$. Line search starting in $\mathbf{x}$ along $\mathbf{d}$:

$$\varphi(\alpha) = f(\mathbf{x} + \alpha\mathbf{d}), \quad \alpha \in [0, 1].$$

A simple optimization model ...

$$\ldots \text{ is } \quad \min_{\mathbf{x}\in\Delta} f(\mathbf{x}) \quad \text{with} \quad \Delta = \left\{ \mathbf{x} \in \mathbb{R}^n_+ : \sum_i x_i = 1 \right\}$$

and $f \in \mathcal{C}^2$. Line search starting in $\mathbf{x}$ along $\mathbf{d}$:

$$\varphi(\alpha) = f(\mathbf{x} + \alpha\mathbf{d}), \quad \alpha \in [0,1].$$

Assumption: curvature of φ has constant sign:

A simple optimization model ...

$$\ldots \text{ is } \quad \min_{\mathbf{x}\in\Delta} f(\mathbf{x}) \quad \text{with} \quad \Delta = \left\{ \mathbf{x} \in \mathbb{R}^n_+ : \sum_i x_i = 1 \right\}$$

and $f \in \mathcal{C}^2$. Line search starting in $\mathbf{x}$ along $\mathbf{d}$:

$$\varphi(\alpha) = f(\mathbf{x} + \alpha \mathbf{d}) \,, \quad \alpha \in [0,1] \,.$$

Assumption: curvature of φ has constant sign:

$$\text{sign } \ddot{\varphi}(\alpha) = const. \quad \text{in } \alpha \in [0,1] \,,$$

A simple optimization model ...

$$\ldots \text{ is} \quad \min_{\mathbf{x}\in\Delta} f(\mathbf{x}) \quad \text{with} \quad \Delta = \left\{ \mathbf{x} \in \mathbb{R}^n_+ : \sum_i x_i = 1 \right\}$$

and $f \in \mathcal{C}^2$. Line search starting in $\mathbf{x}$ along $\mathbf{d}$:

$$\varphi(\alpha) = f(\mathbf{x} + \alpha\mathbf{d}), \quad \alpha \in [0, 1].$$

Assumption: curvature of φ has constant sign:

$$\text{sign } \ddot{\varphi}(\alpha) = const. \quad \text{in } \alpha \in [0, 1],$$

and moreover in the strictly convex descent case $\dot{\varphi}(0) < 0 < \ddot{\varphi}(0)$

A simple optimization model ...

$$\ldots \text{ is } \quad \min_{\mathbf{x}\in\Delta} f(\mathbf{x}) \quad \text{with} \quad \Delta = \left\{ \mathbf{x} \in \mathbb{R}^n_+ : \sum_i x_i = 1 \right\}$$

and $f \in \mathcal{C}^2$. Line search starting in $\mathbf{x}$ along $\mathbf{d}$:

$$\varphi(\alpha) = f(\mathbf{x} + \alpha\mathbf{d}) \, , \quad \alpha \in [0,1] \, .$$

Assumption: curvature of φ has constant sign:

$$\text{sign } \ddot{\varphi}(\alpha) = const. \quad \text{in } \alpha \in [0,1] \, ,$$

and moreover in the strictly convex descent case $\dot{\varphi}(0) < 0 < \ddot{\varphi}(0)$

$$\text{for some } \eta > 0 \, , \quad \ddot{\varphi}(\alpha) \geq \eta \quad \text{for all } \alpha \in [0,1] \, .$$

A simple optimization model ...

$$\ldots \text{ is } \quad \min_{\mathbf{x}\in\Delta} f(\mathbf{x}) \quad \text{with} \quad \Delta = \left\{ \mathbf{x} \in \mathbb{R}^n_+ : \sum_i x_i = 1 \right\}$$

and $f \in \mathcal{C}^2$. Line search starting in $\mathbf{x}$ along $\mathbf{d}$:

$$\varphi(\alpha) = f(\mathbf{x} + \alpha\mathbf{d})\,, \quad \alpha \in [0,1]\,.$$

Assumption: curvature of φ has constant sign:

$$\text{sign } \ddot{\varphi}(\alpha) = const. \quad \text{in } \alpha \in [0,1]\,,$$

and moreover in the strictly convex descent case $\dot{\varphi}(0) < 0 < \ddot{\varphi}(0)$

$$\text{for some } \eta > 0\,, \quad \ddot{\varphi}(\alpha) \geq \eta \quad \text{for all } \alpha \in [0,1]\,.$$

Holds for all (also non-convex) quadratic f and many more.

Frank-Wolfe and away directions

Point $\mathbf{x} \in \Delta$ is (KKT) stationary if and only if

$$\nabla_r f(\mathbf{x}^*) \geq \nabla f(\mathbf{x}^*)^\top \mathbf{x}^* \quad \text{for all } r \text{ with equality if } x_r^* > 0\,.$$

Frank-Wolfe and away directions

Point $\mathbf{x} \in \Delta$ is (KKT) stationary if and only if

$$\nabla_r f(\mathbf{x}^*) \geq \nabla f(\mathbf{x}^*)^\top \mathbf{x}^* \quad \text{for all } r \text{ with equality if } x_r^* > 0\,.$$

Frank-Wolfe direction at $\mathbf{x}$:

$$\mathbf{d}_{FW}(\mathbf{x}) = \mathbf{e}_{\widehat{\imath}} - \mathbf{x}\,, \quad \widehat{\imath} \in Argmin\,\{\nabla_i f(\mathbf{x}) : 1 \leq i \leq n\}$$

where $\mathbf{e}_i$ is ith column of $n \times n$ identity matrix.

Frank-Wolfe and away directions

Point $\mathbf{x} \in \Delta$ is (KKT) stationary if and only if

$$\nabla_r f(\mathbf{x}^*) \geq \nabla f(\mathbf{x}^*)^\top \mathbf{x}^* \quad \text{for all } r \text{ with equality if } x_r^* > 0\,.$$

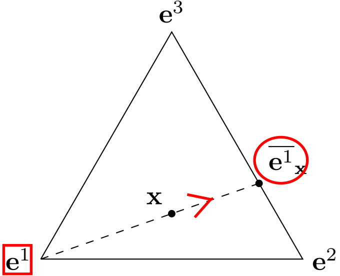
Frank-Wolfe direction at $\mathbf{x}$:

$$\mathbf{d}_{FW}(\mathbf{x}) = \mathbf{e}_{\hat{\imath}} - \mathbf{x}\,, \quad \hat{\imath} \in \mathit{Argmin}\,\{\nabla_i f(\mathbf{x}) : 1 \leq i \leq n\}$$

where $\mathbf{e}_i$ is ith column of $n \times n$ identity matrix.

Away direction at $\mathbf{x}$:

$$\mathbf{d}_A(\mathbf{x}) = \mathbf{x} - \mathbf{e}_{\hat{\jmath}}\,, \quad \hat{\jmath} \in \mathit{Argmax}\{\nabla_j f(\mathbf{x}) : x_j > 0\}\,.$$



e³
e¹ₓ
x
e¹
e²

Frank-Wolfe and away directions

Point $\mathbf{x} \in \Delta$ is (KKT) stationary if and only if

$$\nabla_r f(\mathbf{x}^*) \geq \nabla f(\mathbf{x}^*)^\top \mathbf{x}^* \quad \text{for all } r \text{ with equality if } x_r^* > 0\,.$$

Frank-Wolfe direction at $\mathbf{x}$:

$$\mathbf{d}_{FW}(\mathbf{x}) = \mathbf{e}_{\widehat{\imath}} - \mathbf{x}\,, \quad \widehat{\imath} \in Argmin\,\{\nabla_i f(\mathbf{x}) : 1 \leq i \leq n\}$$

where $\mathbf{e}_i$ is ith column of $n \times n$ identity matrix.

Away direction at $\mathbf{x}$:

$$\mathbf{d}_A(\mathbf{x}) = \mathbf{x} - \mathbf{e}_{\widehat{\jmath}}\,, \quad \widehat{\jmath} \in Argmax\{\nabla_j f(\mathbf{x}) : x_j > 0\}\,.$$

Both are feasible descent directions unless $\mathbf{x}$ is stationary.

Two variants of search directions

Based upon $(\widehat{\imath}, \widehat{\jmath})$, i.e., on $\mathbf{d}_{FW}(\mathbf{x})$ and $\mathbf{d}_A(\mathbf{x})$, we consider two variants.

Two variants of search directions

Based upon $(\widehat{\imath}, \widehat{\jmath})$, i.e., on $\mathbf{d}_{FW}(\mathbf{x})$ and $\mathbf{d}_A(\mathbf{x})$, we consider two variants.

Away-step Frank-Wolfe (AFW) direction:

$$\mathbf{d}_{AFW}(\mathbf{x}) = \begin{cases} \mathbf{d}_{FW}(\mathbf{x}), & \text{if } \mathbf{d}_{FW}(\mathbf{x})^\top \nabla f(\mathbf{x}) \leq \mathbf{d}_A(\mathbf{x})^\top \nabla f(\mathbf{x}), \\ & \end{cases}$$

Two variants of search directions

Based upon $(\hat{\imath}, \hat{\jmath})$, i.e., on $\mathbf{d}_{FW}(\mathbf{x})$ and $\mathbf{d}_A(\mathbf{x})$, we consider two variants.

Away-step Frank-Wolfe (AFW) direction:

$$\mathbf{d}_{AFW}(\mathbf{x}) = \begin{cases} \mathbf{d}_{FW}(\mathbf{x}), & \text{if } \mathbf{d}_{FW}(\mathbf{x})^\top \nabla f(\mathbf{x}) \le \mathbf{d}_A(\mathbf{x})^\top \nabla f(\mathbf{x}), \\ \dfrac{x_{\hat{\jmath}}}{1 - x_{\hat{\jmath}}} \mathbf{d}_A(\mathbf{x}), & \text{otherwise;} \end{cases}$$

Two variants of search directions

Based upon $(\widehat{\imath}, \widehat{\jmath})$, i.e., on $\mathbf{d}_{FW}(\mathbf{x})$ and $\mathbf{d}_A(\mathbf{x})$, we consider two variants.

Away-step Frank-Wolfe (AFW) direction:

$$\mathbf{d}_{AFW}(\mathbf{x}) = \begin{cases} \mathbf{d}_{FW}(\mathbf{x})\,, & \text{if } \mathbf{d}_{FW}(\mathbf{x})^\top \nabla f(\mathbf{x}) \leq \mathbf{d}_A(\mathbf{x})^\top \nabla f(\mathbf{x})\,, \\ \dfrac{x_{\widehat{\jmath}}}{1 - x_{\widehat{\jmath}}} \mathbf{d}_A(\mathbf{x})\,, & \text{otherwise;} \end{cases}$$

and the Pairwise Frank-Wolfe (PFW) direction

$$\mathbf{d}_{PFW}(\mathbf{x}) = x_{\widehat{\jmath}} [\mathbf{d}_{FW}(\mathbf{x}) + \mathbf{d}_A(\mathbf{x})] = x_{\widehat{\jmath}} \left[\mathbf{e}_{\widehat{\imath}} - \mathbf{e}_{\widehat{\jmath}} \right] \,.$$

Two variants of search directions

Based upon $(\hat{\imath}, \hat{\jmath})$, i.e., on $\mathbf{d}_{FW}(\mathbf{x})$ and $\mathbf{d}_A(\mathbf{x})$, we consider two variants.

Away-step Frank-Wolfe (AFW) direction:

$$\mathbf{d}_{AFW}(\mathbf{x}) = \begin{cases} \mathbf{d}_{FW}(\mathbf{x}), & \text{if } \mathbf{d}_{FW}(\mathbf{x})^\top \nabla f(\mathbf{x}) \leq \mathbf{d}_A(\mathbf{x})^\top \nabla f(\mathbf{x}), \\ \dfrac{x_{\hat{\jmath}}}{1 - x_{\hat{\jmath}}} \mathbf{d}_A(\mathbf{x}), & \text{otherwise;} \end{cases}$$

and the Pairwise Frank-Wolfe (PFW) direction

$$\mathbf{d}_{PFW}(\mathbf{x}) = x_{\hat{\jmath}} [\mathbf{d}_{FW}(\mathbf{x}) + \mathbf{d}_A(\mathbf{x})] = x_{\hat{\jmath}} \left[\mathbf{e}_{\hat{\imath}} - \mathbf{e}_{\hat{\jmath}}\right] .$$

Both discontinuous in $\mathbf{x}$, as with gradient projection.

Exact line search

Have rescaled so that feasible step lengths are $\alpha \in [0, 1]$.

Exact line search

Have rescaled so that feasible step lengths are $\alpha \in [0, 1]$.

Exact line search: largest α minimizing φ,

Exact line search

Have rescaled so that feasible step lengths are $\alpha \in [0,1]$.

Exact line search: largest α minimizing φ,

$$\alpha_{exact}(\mathbf{d},\mathbf{x}) := \max Argmin \{\varphi(\alpha) : \alpha \in [0,1]\} \ ,$$

Exact line search

Have rescaled so that feasible step lengths are $\alpha \in [0, 1]$.

Exact line search: largest α minimizing φ,

$$\alpha_{exact}(\mathbf{d}, \mathbf{x}) := \max Argmin \{\varphi(\alpha) : \alpha \in [0, 1]\} \ ,$$

which satisfies, if $0 < \alpha_{exact}(\mathbf{d}, \mathbf{x}) < 1$,

$$\alpha_{exact}(\mathbf{d}, \mathbf{x}) = -\frac{\dot{\varphi}(0)}{\ddot{\varphi}(\tilde{\alpha})}$$

for some $\tilde{\alpha} \in (0, 1)$.

Exact line search

Have rescaled so that feasible step lengths are $\alpha \in [0, 1]$.

Exact line search: largest α minimizing φ,

$$\alpha_{exact}(\mathbf{d}, \mathbf{x}) := \max Argmin \{\varphi(\alpha) : \alpha \in [0, 1]\} \ ,$$

which satisfies, if $0 < \alpha_{exact}(\mathbf{d}, \mathbf{x}) < 1$,

$$\alpha_{exact}(\mathbf{d}, \mathbf{x}) = -\frac{\dot{\varphi}(0)}{\ddot{\varphi}(\tilde{\alpha})}$$

for some $\tilde{\alpha} \in (0, 1)$.

Happens only in strictly convex case $\dot{\varphi}(0) < 0 < \ddot{\varphi}(0)$.

Inexact line search à la Armijo; iteration

Select $\delta \in (0, 1)$ and $\gamma \in (0, \frac{1}{2})$; sufficient decrease:

Inexact line search à la Armijo; iteration

Select $\delta \in (0, 1)$ and $\gamma \in (0, \frac{1}{2})$; sufficient decrease:

$$m(\mathbf{d}, \mathbf{x}) := \min \left\{ m \in \mathbb{N} : f(\mathbf{x} + \delta^m \mathbf{d}) \leq f(\mathbf{x}) + \gamma \delta^m \mathbf{d}^\top \nabla f(\mathbf{x}) \right\}$$

Inexact line search à la Armijo; iteration

Select $\delta \in (0,1)$ and $\gamma \in (0,\frac{1}{2})$; sufficient decrease:

$$m(\mathbf{d},\mathbf{x}) := \min\left\{m \in \mathbb{N} : f(\mathbf{x}+\delta^m\,\mathbf{d}) \leq f(\mathbf{x}) + \gamma\delta^m\,\mathbf{d}^\top \nabla f(\mathbf{x})\right\}$$

and put

$$\alpha_{Armijo}(\mathbf{d},\mathbf{x}) := \delta^{m(\mathbf{d},\mathbf{x})}\,.$$

Inexact line search à la Armijo; iteration

Select $\delta \in (0,1)$ and $\gamma \in (0,\frac{1}{2})$; sufficient decrease:

$$m(\mathbf{d},\mathbf{x}) := \min\left\{m \in \mathbb{N} : f(\mathbf{x} + \delta^m \mathbf{d}) \le f(\mathbf{x}) + \gamma\delta^m \mathbf{d}^\top \nabla f(\mathbf{x})\right\}$$

and put

$$\alpha_{Armijo}(\mathbf{d},\mathbf{x}) := \delta^{m(\mathbf{d},\mathbf{x})}\,.$$

Iteration for both line search variants:

Starting point $\mathbf{x}^0 \in \Delta$,

Inexact line search à la Armijo; iteration

Select $\delta \in (0,1)$ and $\gamma \in (0, \frac{1}{2})$; sufficient decrease:

$$m(\mathbf{d},\mathbf{x}) := \min \left\{ m \in \mathbb{N} : f(\mathbf{x} + \delta^m \mathbf{d}) \le f(\mathbf{x}) + \gamma \delta^m \mathbf{d}^\top \nabla f(\mathbf{x}) \right\}$$

and put

$$\alpha_{Armijo}(\mathbf{d},\mathbf{x}) := \delta^{m(\mathbf{d},\mathbf{x})} .$$

Iteration for both line search variants:

Starting point $\mathbf{x}^0 \in \Delta$,

$$\mathbf{x}^{k+1} = T(\mathbf{x}^k) := \mathbf{x}^k + \alpha(\mathbf{d}(\mathbf{x}^k), \mathbf{x}^k)\, \mathbf{d}(\mathbf{x}^k) , \quad k = 0, 1, \ldots$$

Inexact line search à la Armijo; iteration

Select $\delta \in (0,1)$ and $\gamma \in (0,\frac{1}{2})$; sufficient decrease:

$$m(\mathbf{d},\mathbf{x}) := \min\left\{m \in \mathbb{N} : f(\mathbf{x}+\delta^m\,\mathbf{d}) \leq f(\mathbf{x}) + \gamma\delta^m\,\mathbf{d}^\top \nabla f(\mathbf{x})\right\}$$

and put

$$\alpha_{Armijo}(\mathbf{d},\mathbf{x}) := \delta^{m(\mathbf{d},\mathbf{x})}\,.$$

Iteration for both line search variants:

Starting point $\mathbf{x}^0 \in \Delta$,

$$\mathbf{x}^{k+1} = T(\mathbf{x}^k) := \mathbf{x}^k + \alpha(\mathbf{d}(\mathbf{x}^k),\mathbf{x}^k)\,\mathbf{d}(\mathbf{x}^k)\,, \quad k = 0,1,\ldots$$

Have $T(\mathbf{x}) = \mathbf{x}$ if and only if $\mathbf{x}$ is stationary.

Inexact line search à la Armijo; iteration

Select $\delta \in (0,1)$ and $\gamma \in (0,\frac{1}{2})$; sufficient decrease:

$$m(\mathbf{d},\mathbf{x}) := \min\left\{m \in \mathbb{N} : f(\mathbf{x}+\delta^m\,\mathbf{d}) \le f(\mathbf{x}) + \gamma\delta^m\,\mathbf{d}^\top \nabla f(\mathbf{x})\right\}$$

and put

$$\alpha_{Armijo}(\mathbf{d},\mathbf{x}) := \delta^{m(\mathbf{d},\mathbf{x})}\,.$$

Iteration for both line search variants:

Starting point $\mathbf{x}^0 \in \Delta$,

$$\mathbf{x}^{k+1} = T(\mathbf{x}^k) := \mathbf{x}^k + \alpha(\mathbf{d}(\mathbf{x}^k),\mathbf{x}^k)\,\mathbf{d}(\mathbf{x}^k)\,,\quad k = 0,1,\ldots$$

Have $T(\mathbf{x}) = \mathbf{x}$ if and only if $\mathbf{x}$ is stationary.

Note that T is discontinuous in $\mathbf{x}$ if $\mathbf{d}$ is so.

Monotonicity and slope convergence

Proposition:

For $\mathbf{d} \in \{\mathbf{d}_{AFW}, \mathbf{d}_{PFW}\}$ *and* $\alpha \in \left\{\alpha_{exact}, \alpha_{Armijo}\right\}$, *we have*

Monotonicity and slope convergence

Proposition:

For $\mathbf{d} \in \{\mathbf{d}_{AFW}, \mathbf{d}_{PFW}\}$ *and* $\alpha \in \left\{\alpha_{exact}, \alpha_{Armijo}\right\}$*, we have*

(a) if $\mathbf{x}^{k+1} \neq \mathbf{x}^k$*, then* $f(\mathbf{x}^{k+1}) < f(\mathbf{x}^k)$*;*

Monotonicity and slope convergence

Proposition:

For $\mathbf{d} \in \{\mathbf{d}_{AFW}, \mathbf{d}_{PFW}\}$ *and* $\alpha \in \left\{\alpha_{exact}, \alpha_{Armijo}\right\}$, *we have*

(a) if $\mathbf{x}^{k+1} \neq \mathbf{x}^k$, *then* $f(\mathbf{x}^{k+1}) < f(\mathbf{x}^k)$;

(b) if $\mathbf{x}^{k+1} \neq \mathbf{x}^k$ *for all* $k \in \mathbb{N}$, *then*

$$\mathbf{d}(\mathbf{x}^k)^\top \nabla f(\mathbf{x}^k) \to 0 \quad \text{as } k \to \infty \,.$$

Monotonicity and slope convergence

Proposition:

For $\mathbf{d} \in \{\mathbf{d}_{AFW}, \mathbf{d}_{PFW}\}$ *and* $\alpha \in \left\{\alpha_{exact}, \alpha_{Armijo}\right\}$*, we have*

(a) if $\mathbf{x}^{k+1} \neq \mathbf{x}^k$*, then* $f(\mathbf{x}^{k+1}) < f(\mathbf{x}^k)$*;*

(b) if $\mathbf{x}^{k+1} \neq \mathbf{x}^k$ *for all* $k \in \mathbb{N}$*, then*

$$\mathbf{d}(\mathbf{x}^k)^\top \nabla f(\mathbf{x}^k) \to 0 \quad \textit{as } k \to \infty \, .$$

Key argument in proof: for $K = \max_{\mathbf{x} \in \Delta} \|\nabla^2 f(\mathbf{x})\|_{spec}$

Monotonicity and slope convergence

Proposition:

For $\mathbf{d} \in \{\mathbf{d}_{AFW}, \mathbf{d}_{PFW}\}$ *and* $\alpha \in \left\{\alpha_{exact}, \alpha_{Armijo}\right\}$, *we have*

(a) if $\mathbf{x}^{k+1} \neq \mathbf{x}^k$, *then* $f(\mathbf{x}^{k+1}) < f(\mathbf{x}^k)$*;*

(b) if $\mathbf{x}^{k+1} \neq \mathbf{x}^k$ *for all* $k \in \mathbb{N}$, *then*

$$\mathbf{d}(\mathbf{x}^k)^\top \nabla f(\mathbf{x}^k) \to 0 \quad \textit{as } k \to \infty\,.$$

Key argument in proof: for $K = \max_{\mathbf{x} \in \Delta} \|\nabla^2 f(\mathbf{x})\|_{spec}$ have

$$[\mathbf{d}(\mathbf{x}^k)^\top \nabla f(\mathbf{x}^k)]^2 = [\dot{\varphi}(0)]^2 \leq \frac{4K}{\eta}\,[f(\mathbf{x}^k) - f(\mathbf{x}^{k+1})]\,.$$

Global convergence

Theorem:

For $\mathbf{d} \in \{\mathbf{d}_{AFW}, \mathbf{d}_{PFW}\}$ *and* $\alpha \in \left\{\alpha_{exact}, \alpha_{Armijo}\right\}$, *we have that either* $\mathbf{x}^{\bar{k}}$ *is stationary for some* $\bar{k} \in \mathbb{N}$ *and iteration stops;*

Global convergence

Theorem:

For $\mathbf{d} \in \{\mathbf{d}_{AFW}, \mathbf{d}_{PFW}\}$ *and* $\alpha \in \left\{\alpha_{exact}, \alpha_{Armijo}\right\}$, *we have that*

either $\mathbf{x}^{\bar{k}}$ *is stationary for some* $\bar{k} \in \mathbb{N}$ *and iteration stops;*

or else any accumulation point $\mathbf{x}^*$ *of iterates* $(\mathbf{x}^k)_k$ *is stationary.*

Global convergence

Theorem:

For $\mathbf{d} \in \{\mathbf{d}_{AFW}, \mathbf{d}_{PFW}\}$ *and* $\alpha \in \left\{\alpha_{exact}, \alpha_{Armijo}\right\}$, *we have that*

either $\mathbf{x}^{\bar{k}}$ *is stationary for some* $\bar{k} \in \mathbb{N}$ *and iteration stops;*

or else any accumulation point $\mathbf{x}^*$ *of iterates* $(\mathbf{x}^k)_k$ *is stationary.*

Key arguments in proof: previous proposition; show that along subsequences stepsize $\alpha_k = 1$ eventually in difficult subcase.

Global convergence

Theorem:

For $\mathbf{d} \in \{\mathbf{d}_{AFW}, \mathbf{d}_{PFW}\}$ *and* $\alpha \in \left\{\alpha_{exact}, \alpha_{Armijo}\right\}$, *we have that*

either $\mathbf{x}^{\bar{k}}$ *is stationary for some* $\bar{k} \in \mathbb{N}$ *and iteration stops;*

or else any accumulation point $\mathbf{x}^*$ *of iterates* $(\mathbf{x}^k)_k$ *is stationary.*

Key arguments in proof: previous proposition; show that along subsequences stepsize $\alpha_k = 1$ eventually in difficult subcase.

Good proof of concept

Global convergence

Theorem:

For $\mathbf{d} \in \{\mathbf{d}_{AFW}, \mathbf{d}_{PFW}\}$ *and* $\alpha \in \left\{\alpha_{exact}, \alpha_{Armijo}\right\}$*, we have that*

either $\mathbf{x}^{\bar{k}}$ *is stationary for some* $\bar{k} \in \mathbb{N}$ *and iteration stops;*

or else any accumulation point $\mathbf{x}^*$ *of iterates* $(\mathbf{x}^k)_k$ *is stationary.*

Key arguments in proof: previous proposition; show that along subsequences stepsize $\alpha_k = 1$ eventually in difficult subcase.

Good proof of concept but still not proved: iterates convergence

Global convergence

Theorem:

For $\mathbf{d} \in \{\mathbf{d}_{AFW}, \mathbf{d}_{PFW}\}$ *and* $\alpha \in \left\{\alpha_{exact}, \alpha_{Armijo}\right\}$*, we have that*

either $\mathbf{x}^{\bar{k}}$ *is stationary for some* $\bar{k} \in \mathbb{N}$ *and iteration stops;*

or else any accumulation point $\mathbf{x}^*$ *of iterates* $(\mathbf{x}^k)_k$ *is stationary.*

Key arguments in proof: previous proposition; show that along subsequences stepsize $\alpha_k = 1$ eventually in difficult subcase.

Good proof of concept but still not proved: iterates convergence (only one accumulation point $\mathbf{x}^*$).

Saddle points are avoided almost surely

Recently discussed: 2^{nd}-order optimality conditions (saddles);

Saddle points are avoided almost surely

Recently discussed: 2^{nd}-order optimality conditions (saddles);
related more general result here: any attracting $\mathbf{x}^*$ is local soln.!

Saddle points are avoided almost surely

Recently discussed: 2^{nd}-order optimality conditions (saddles); related more general result here: any attracting $\mathbf{x}^*$ is local soln.!

Theorem:

Suppose f is either convex or concave along lines

Saddle points are avoided almost surely

Recently discussed: 2^{nd}-order optimality conditions (saddles); related more general result here: any attracting $\mathbf{x}^*$ is local soln.!

Theorem:

Suppose f is either convex or concave along lines and that $f(\mathbf{x}^{k+1}) \leq f(\mathbf{x}^k)$ for all k.

Saddle points are avoided almost surely

Recently discussed: 2^{nd}-order optimality conditions (saddles);

related more general result here: any attracting $\mathbf{x}^*$ is local soln.!

Theorem:

Suppose f is either convex or concave along lines and that $f(\mathbf{x}^{k+1}) \leq f(\mathbf{x}^k)$ for all k. If all acc. points of $(\mathbf{x}^k)_k$ are stationary,

Saddle points are avoided almost surely

Recently discussed: 2^{nd}-order optimality conditions (saddles); related more general result here: any attracting $\mathbf{x}^*$ is local soln.!

Theorem:

Suppose f is either convex or concave along lines and that $f(\mathbf{x}^{k+1}) \leq f(\mathbf{x}^k)$ for all k. If all acc. points of $(\mathbf{x}^k)_k$ are stationary, the following statements are equivalent

Saddle points are avoided almost surely

Recently discussed: 2nd-order optimality conditions (saddles);

related more general result here: any attracting $\mathbf{x}^*$ is local soln.!

Theorem:

Suppose f *is either convex or concave along lines and that* $f(\mathbf{x}^{k+1}) \leq f(\mathbf{x}^k)$ *for all* k*. If all acc. points of* $(\mathbf{x}^k)_k$ *are stationary, the following statements are equivalent for any stationary* $\mathbf{x}^*$ *and neighbourhoods* $U \subseteq \Delta$ *or* $V \subseteq \Delta$ *of* $\mathbf{x}^*$*:*

Saddle points are avoided almost surely

Recently discussed: 2$^{\text{nd}}$-order optimality conditions (saddles); related more general result here: any attracting $\mathbf{x}^*$ is local soln.!

Theorem:

Suppose f is either convex or concave along lines and that $f(\mathbf{x}^{k+1}) \leq f(\mathbf{x}^k)$ for all k. If all acc. points of $(\mathbf{x}^k)_k$ are stationary, the following statements are equivalent for any stationary $\mathbf{x}^*$ and neighbourhoods $U \subseteq \boldsymbol{\Delta}$ or $V \subseteq \boldsymbol{\Delta}$ of $\mathbf{x}^*$:

(a) every sequence $\mathbf{x}^k \to \mathbf{x}^*$ as $k \to \infty$, if $\mathbf{x}^0 \in U$;

Saddle points are avoided almost surely

Recently discussed: 2nd-order optimality conditions (saddles);

related more general result here: any attracting $\mathbf{x}^*$ is local soln.!

Theorem:

Suppose f is either convex or concave along lines and that $f(\mathbf{x}^{k+1}) \leq f(\mathbf{x}^k)$ for all k. If all acc. points of $(\mathbf{x}^k)_k$ are stationary, the following statements are equivalent for any stationary $\mathbf{x}^*$ and neighbourhoods $U \subseteq \Delta$ or $V \subseteq \Delta$ of $\mathbf{x}^*$:

(a) every sequence $\mathbf{x}^k \to \mathbf{x}^*$ as $k \to \infty$, if $\mathbf{x}^0 \in U$;

(b) $f(\mathbf{x}) > f(\mathbf{x}^*)$ for all $\mathbf{x} \in V \setminus \{\mathbf{x}^*\}$ ... contains no KKT point.

Saddle points are avoided almost surely

Recently discussed: 2$^{\text{nd}}$-order optimality conditions (saddles); related more general result here: any attracting $\mathbf{x}^*$ is local soln.!

Theorem:

Suppose f is either convex or concave along lines and that $f(\mathbf{x}^{k+1}) \leq f(\mathbf{x}^k)$ for all k. If all acc. points of $(\mathbf{x}^k)_k$ are stationary, the following statements are equivalent for any stationary $\mathbf{x}^*$ and neighbourhoods $U \subseteq \Delta$ or $V \subseteq \Delta$ of $\mathbf{x}^*$:

(a) every sequence $\mathbf{x}^k \to \mathbf{x}^*$ as $k \to \infty$, if $\mathbf{x}^0 \in U$;

(b) $f(\mathbf{x}) > f(\mathbf{x}^*)$ for all $\mathbf{x} \in V \setminus \{\mathbf{x}^*\}$ ... contains no KKT point.

So no negative curvature feasible direction of $\nabla^2 f(\mathbf{x}^*)$.

Saddle points are avoided almost surely

Recently discussed: 2nd-order optimality conditions (saddles); related more general result here: any attracting $\mathbf{x}^*$ is local soln.!

Theorem:

Suppose f *is either convex or concave along lines and that* $f(\mathbf{x}^{k+1}) \leq f(\mathbf{x}^k)$ *for all* k. *If all acc. points of* $(\mathbf{x}^k)_k$ *are stationary, the following statements are equivalent for any stationary* $\mathbf{x}^*$ *and neighbourhoods* $U \subseteq \Delta$ *or* $V \subseteq \Delta$ *of* $\mathbf{x}^*$*:*

(a) every sequence $\mathbf{x}^k \to \mathbf{x}^*$ *as* $k \to \infty$*, if* $\mathbf{x}^0 \in U$*;*

(b) $f(\mathbf{x}) > f(\mathbf{x}^*)$ *for all* $\mathbf{x} \in V \setminus \{\mathbf{x}^*\}$ *... contains no KKT point.*

So no negative curvature feasible direction of $\nabla^2 f(\mathbf{x}^*)$.

Implies that all strict local solutions are isolated KKT points.

Saddle points are avoided almost surely

Recently discussed: 2nd-order optimality conditions (saddles); related more general result here: any attracting $\mathbf{x}^*$ is local soln.!

Theorem:

Suppose f is either convex or concave along lines and that $f(\mathbf{x}^{k+1}) \leq f(\mathbf{x}^k)$ for all k. If all acc. points of $(\mathbf{x}^k)_k$ are stationary, the following statements are equivalent for any stationary $\mathbf{x}^*$ and neighbourhoods $U \subseteq \Delta$ or $V \subseteq \Delta$ of $\mathbf{x}^*$:

(a) every sequence $\mathbf{x}^k \to \mathbf{x}^*$ as $k \to \infty$, if $\mathbf{x}^0 \in U$;

(b) $f(\mathbf{x}) > f(\mathbf{x}^*)$ for all $\mathbf{x} \in V \setminus \{\mathbf{x}^*\}$ ... contains no KKT point.

So no negative curvature feasible direction of $\nabla^2 f(\mathbf{x}^*)$.

Implies that all strict local solutions are isolated KKT points.

Applies to all f considered here

Saddle points are avoided almost surely

Recently discussed: 2nd-order optimality conditions (saddles);

related more general result here: any attracting $\mathbf{x}^*$ is local soln.!

Theorem:

Suppose f is either convex or concave along lines and that $f(\mathbf{x}^{k+1}) \leq f(\mathbf{x}^k)$ for all k. If all acc. points of $(\mathbf{x}^k)_k$ are stationary, the following statements are equivalent for any stationary $\mathbf{x}^*$ and neighbourhoods $U \subseteq \Delta$ or $V \subseteq \Delta$ of $\mathbf{x}^*$:

(a) every sequence $\mathbf{x}^k \to \mathbf{x}^*$ as $k \to \infty$, if $\mathbf{x}^0 \in U$;

(b) $f(\mathbf{x}) > f(\mathbf{x}^*)$ for all $\mathbf{x} \in V \setminus \{\mathbf{x}^*\}$ ... contains no KKT point.

So no negative curvature feasible direction of $\nabla^2 f(\mathbf{x}^*)$.

Implies that all strict local solutions are isolated KKT points.

Applies to all f considered here, including nonconvex quadratic.

Convergence of iterates

Generic property for objective functions:
only finitely many stationary points.

Convergence of iterates

Generic property for objective functions:

only finitely many stationary points.

Then previous results imply only finitely many acc. points.

Convergence of iterates

Generic property for objective functions:

only finitely many stationary points.

Then previous results imply only finitely many acc. points.

Needed in:

Theorem:

For $\alpha \in \left\{\alpha_{exact}, \alpha_{Armijo}\right\}$ *and* $\mathbf{d} = \mathbf{d}_{AFW}$*, suppose that* $(\mathbf{x}^k)_k$ *has finitely many accumulation points. Then*

Convergence of iterates

Generic property for objective functions:

only finitely many stationary points.

Then previous results imply only finitely many acc. points.

Needed in:

Theorem:

For $\alpha \in \left\{\alpha_{exact}, \alpha_{Armijo}\right\}$ *and* $\mathbf{d} = \mathbf{d}_{AFW}$*, suppose that* $(\mathbf{x}^k)_k$ *has finitely many accumulation points. Then* $\mathbf{x}^k \to \mathbf{x}^*$ *as* $k \to \infty$*,*

Convergence of iterates

Generic property for objective functions:

only finitely many stationary points.

Then previous results imply only finitely many acc. points.

Needed in:

Theorem:

For $\alpha \in \left\{\alpha_{exact}, \alpha_{Armijo}\right\}$ *and* $\mathbf{d} = \mathbf{d}_{AFW}$*, suppose that* $(\mathbf{x}^k)_k$ *has finitely many accumulation points. Then* $\mathbf{x}^k \to \mathbf{x}^*$ *as* $k \to \infty$*, only one such accumulation point.*

Convergence of iterates

Generic property for objective functions:

only finitely many stationary points.

Then previous results imply only finitely many acc. points.

Needed in:

Theorem:

For $\alpha \in \left\{\alpha_{exact}, \alpha_{Armijo}\right\}$ *and* $\mathbf{d} = \mathbf{d}_{AFW}$*, suppose that* $(\mathbf{x}^k)_k$ *has finitely many accumulation points. Then* $\mathbf{x}^k \to \mathbf{x}^*$ *as* $k \to \infty$*, only one such accumulation point.*

Proof would be easy if T were continuous

Convergence of iterates

Generic property for objective functions:

only finitely many stationary points.

Then previous results imply only finitely many acc. points.

Needed in:

Theorem:

For $\alpha \in \left\{\alpha_{exact}, \alpha_{Armijo}\right\}$ *and* $\mathbf{d} = \mathbf{d}_{AFW}$*, suppose that* $(\mathbf{x}^k)_k$ *has finitely many accumulation points. Then* $\mathbf{x}^k \to \mathbf{x}^*$ *as* $k \to \infty$*, only one such accumulation point.*

Proof would be easy if T were continuous, since a finite connected set is a singleton

Convergence of iterates

Generic property for objective functions:

only finitely many stationary points.

Then previous results imply only finitely many acc. points.

Needed in:

Theorem:

For $\alpha \in \left\{\alpha_{exact}, \alpha_{Armijo}\right\}$ *and* $\mathbf{d} = \mathbf{d}_{AFW}$*, suppose that* $(\mathbf{x}^k)_k$ *has finitely many accumulation points. Then* $\mathbf{x}^k \to \mathbf{x}^*$ *as* $k \to \infty$*, only one such accumulation point.*

Proof would be easy if T were continuous, since a finite connected set is a singleton; but T isn't continuous !

Support identification in finite time

Let $S_k = S(\mathbf{x}^k)$ be supports of iterates $\mathbf{x}^k$

Support identification in finite time

Let $S_k = S(\mathbf{x}^k)$ be supports of iterates $\mathbf{x}^k \to \mathbf{x}^*$.

Support identification in finite time

Let $S_k = S(\mathbf{x}^k)$ be supports of iterates $\mathbf{x}^k \to \mathbf{x}^*$. By continuity,

$$S(\mathbf{x}^*) \subseteq S(\mathbf{x}^k) \quad \text{for all large enough } k\,.$$

Support identification in finite time

Let $S_k = S(\mathbf{x}^k)$ be supports of iterates $\mathbf{x}^k \to \mathbf{x}^*$. By continuity,

$$S(\mathbf{x}^*) \subseteq S(\mathbf{x}^k) \quad \text{for all large enough } k\,.$$

Furthermore, stationarity of $\mathbf{x}^*$ implies

$$S(\mathbf{x}^*) \subseteq S_0^* := \left\{ i : \nabla_i f(\mathbf{x}^*) = \nabla f(\mathbf{x}^*)^\top \mathbf{x}^* \right\} \,,$$

Support identification in finite time

Let $S_k = S(\mathbf{x}^k)$ be supports of iterates $\mathbf{x}^k \to \mathbf{x}^*$. By continuity,

$$S(\mathbf{x}^*) \subseteq S(\mathbf{x}^k) \quad \text{for all large enough } k\,.$$

Furthermore, stationarity of $\mathbf{x}^*$ implies

$$S(\mathbf{x}^*) \subseteq S_0^* := \left\{ i : \nabla_i f(\mathbf{x}^*) = \nabla f(\mathbf{x}^*)^\top \mathbf{x}^* \right\}\,,$$

and strict complementarity is equivalent to $S(\mathbf{x}^*) = S_0^*$.

Support identification in finite time

Let $S_k = S(\mathbf{x}^k)$ be supports of iterates $\mathbf{x}^k \to \mathbf{x}^*$. By continuity,

$$S(\mathbf{x}^*) \subseteq S(\mathbf{x}^k) \quad \text{for all large enough } k\,.$$

Furthermore, stationarity of $\mathbf{x}^*$ implies

$$S(\mathbf{x}^*) \subseteq S_0^* := \left\{ i : \nabla_i f(\mathbf{x}^*) = \nabla f(\mathbf{x}^*)^\top \mathbf{x}^* \right\}\,,$$

and strict complementarity is equivalent to $S(\mathbf{x}^*) = S_0^*$.

Theorem:

For $\mathbf{d} \in \{\mathbf{d}_{AFW}, \mathbf{d}_{PFW}\}$ *and* $\alpha \in \left\{\alpha_{exact}, \alpha_{Armijo}\right\}$*, suppose that iterates* $\mathbf{x}^k \to \mathbf{x}^*$ *as* $k \to \infty$*.*

Support identification in finite time

Let $S_k = S(\mathbf{x}^k)$ be supports of iterates $\mathbf{x}^k \to \mathbf{x}^*$. By continuity,

$$S(\mathbf{x}^*) \subseteq S(\mathbf{x}^k) \quad \text{for all large enough } k\,.$$

Furthermore, stationarity of $\mathbf{x}^*$ implies

$$S(\mathbf{x}^*) \subseteq S_0^* := \left\{ i : \nabla_i f(\mathbf{x}^*) = \nabla f(\mathbf{x}^*)^\top \mathbf{x}^* \right\}\,,$$

and strict complementarity is equivalent to $S(\mathbf{x}^*) = S_0^*$.

Theorem:

For $\mathbf{d} \in \{\mathbf{d}_{AFW}, \mathbf{d}_{PFW}\}$ *and* $\alpha \in \left\{\alpha_{exact}, \alpha_{Armijo}\right\}$*, suppose that iterates* $\mathbf{x}^k \to \mathbf{x}^*$ *as* $k \to \infty$*. Then there is a finite* $\bar{k}$ *such that*

$$S(\mathbf{x}^*) \subseteq S(\mathbf{x}^k) \subseteq S_0^* \quad \textit{for all } k \geq \bar{k}\,.$$

Support identification in finite time

Let $S_k = S(\mathbf{x}^k)$ be supports of iterates $\mathbf{x}^k \to \mathbf{x}^*$. By continuity,

$$S(\mathbf{x}^*) \subseteq S(\mathbf{x}^k) \quad \text{for all large enough } k\,.$$

Furthermore, stationarity of $\mathbf{x}^*$ implies

$$S(\mathbf{x}^*) \subseteq S_0^* := \left\{ i : \nabla_i f(\mathbf{x}^*) = \nabla f(\mathbf{x}^*)^\top \mathbf{x}^* \right\}\,,$$

and strict complementarity is equivalent to $S(\mathbf{x}^*) = S_0^*$.

Theorem:

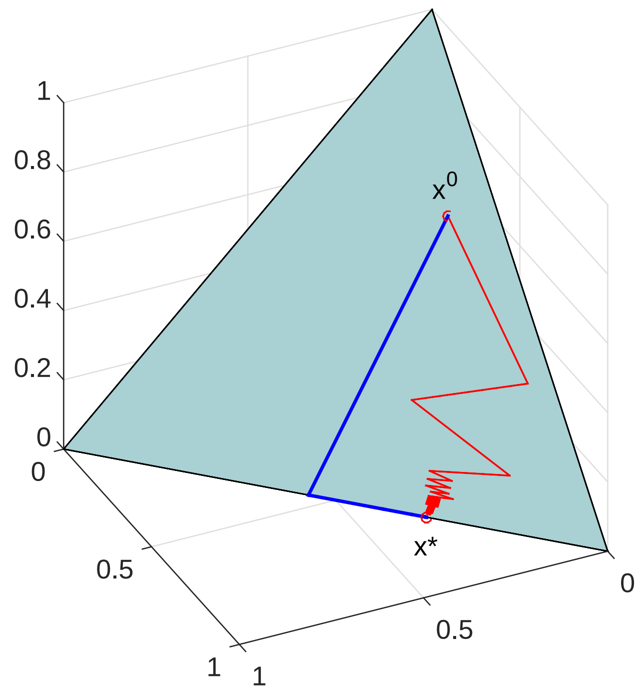
For $\mathbf{d} \in \{\mathbf{d}_{AFW}, \mathbf{d}_{PFW}\}$ and $\alpha \in \left\{\alpha_{exact}, \alpha_{Armijo}\right\}$, suppose that iterates $\mathbf{x}^k \to \mathbf{x}^*$ as $k \to \infty$. Then there is a finite $\bar{k}$ such that

$$S(\mathbf{x}^*) \subseteq S(\mathbf{x}^k) \subseteq S_0^* \quad \textit{for all } k \geq \bar{k}\,.$$

So under strict complementarity $S(\mathbf{x}^k) = S_0^* = S(\mathbf{x}^*)$ if $k \geq \bar{k}$.

Not true for classical Frank-Wolfe !

Example: $f(\mathbf{x}) = \mathbf{x}^\top Q\mathbf{x}$ (convex!) quadratic with α_{Armijo} where

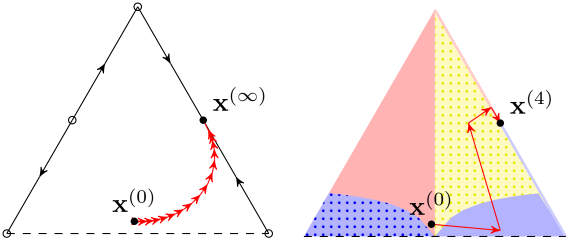


$$Q = \begin{bmatrix} 6 & 0 & 6 \\ 0 & 3 & 3 \\ 6 & 3 & 10 \end{bmatrix}.$$

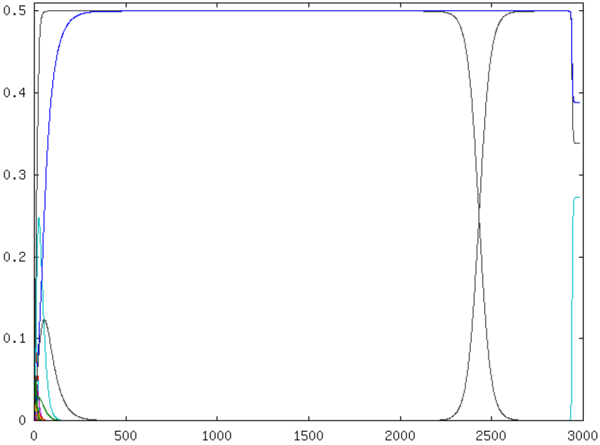
Cristofari & al. (2017), arXiv:1703.07761

A fresh reference

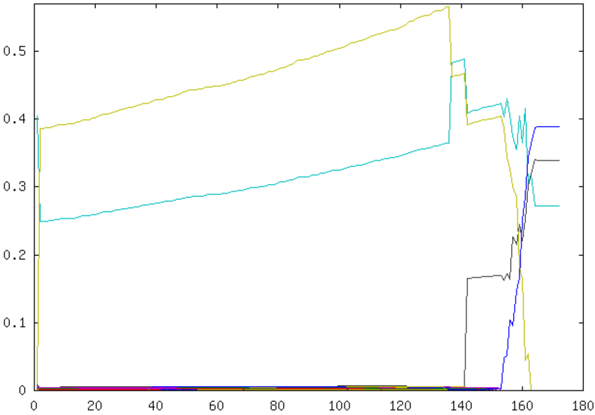
[B./Rinaldi/Rota Bulò '18] First-order methods for the impatient: support identification in finite time with convergent Frank-Wolfe variants *Optimization online* 2018/07/6694 (3 July 2018).



$\mathbf{x}^{(\infty)}$
$\mathbf{x}^{(0)}$
$\mathbf{x}^{(4)}$
$\mathbf{x}^{(0)}$



[Rota Bulò/Pelillo '17]



[Rota Bulò/Pelillo '17]

Computational Results

Experiments: image segmentation

Berkeley database; objects = pixels, Gaussian similarity

$$a_{ij} = \exp(-\|\mathbf{c}(i) - \mathbf{c}(j)\|/\sigma^2)$$

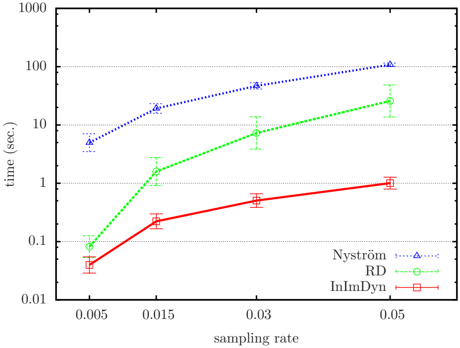
where $\mathbf{c}(i) \in \mathbb{R}^3$ is (Lab) color code and $\sigma > 0$.

Different sampling rates: different $n \approx 200, 600, 1200, 2000$.

Comparison with Nyström method [Fowlkes *et al*:04] (spectral clustering; number of clusters determined by preprocessing) and with RD: stopping criterion: $t \leq t_{\max}$ or Nash error function

$$\varepsilon(\mathbf{x}^t) = \sum_{i \in N} \left[\min\left\{x_i, \left(A\mathbf{x}^t\right)_i - \pi(\mathbf{x}^t)\right\}\right]^2 \leq 10^{-10}.$$

Results: recall/precision identical, but *Inf.Imm.Dyn* much faster.


1000
100
10
1
0.1
0.01
time (sec.)
0.005
0.015
0.03
0.05
sampling rate
Nyström
RD
InImDyn

Experiments: Region-based image matching

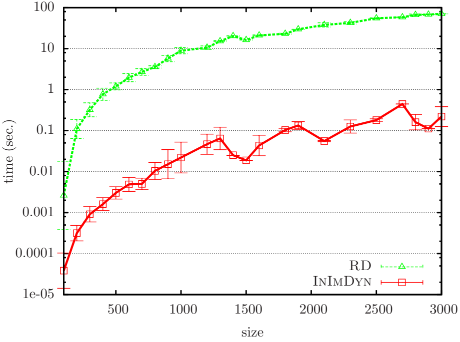
Data from [Todorovic/Ahuja:08] lead to subproblems of tree matching (≥ 100 per image), similarities are adjacencies in *association graph* $G = (V_{\text{ass}}, E_{\text{ass}})$ of two trees $T_i = (V_i, E_i)$, $i = 1, 2$:

$$V_{\text{ass}} = V_1 \times V_2 \,, \quad E_{\text{ass}} = \Big\{ \{(i,h),(j,k)\} : d_{T_1}(i,j) = d_{T_2}(h,k) \Big\} \,,$$

where $d_T(i,j)$ is tree distance of two vertices i, j in T.

This yields instances with n up to 3000, grouped according to their size across images, to obtain error bars.

Compared to *RD*, *Inf.Imm.Dyn* is orders of magnitude faster.


100
10
1
0.1
0.01
0.001
0.0001
1e-05
time (sec.)
500
1000
1500
2000
2500
3000
size
RD
InImDyn

Convergence of iterates

Generic property for objective functions:

only finitely many stationary points.

Then previous results imply only finitely many acc. points.

Needed in:

Theorem:

For $\alpha \in \left\{\alpha_{exact}, \alpha_{Armijo}\right\}$ *and* $\mathbf{d} \in \{\mathbf{d}_{AFW}, \mathbf{d}_{PFW}\}$*, suppose that* $(\mathbf{x}^k)_k$ *has finitely many accumulation points. Then* $\mathbf{x}^k \to \mathbf{x}^*$ *as* $k \to \infty$*, only one such accumulation point.*

Proof would be easy if T were continuous, since a finite connected set is a singleton; but T isn't continuous !