A Proximal Subgradient Method for Nonsmooth Sum-of-Ratios Optimization Problems

Minh N. Dao

One World Optimisation Seminar

March 28th, 2022

Joint work with Radu Boț (University of Vienna) and Guoyin Li (UNSW Sydney)

CRICOS Provider No. 00103D

Outline

Outline

Sum-of-ratios problems

Nonsmooth and nonconcave fractional maximization problem:

$$\max_{\mathbf{x}=(x_1,\ldots,x_m)\in S:=S_1\times\cdots\times S_m} F(\mathbf{x}) := h(x_1,\ldots,x_m) + \sum_{i=1}^{m}\frac{f_i(x_i)}{g_i(x_i)}. \qquad \text{(P)}$$

Here, for all $x_i \in S_i$, $f_i(x_i) \geq 0$ and $g_i(x_i) > 0$.

Classical sum-of-ratios optimization problem:

$$\max_{z\in C}\sum_{i=1}^{m}\frac{f_i(z)}{g_i(z)},$$

which can be rewritten as

$$\max_{x_1,\ldots,x_m\in C}\sum_{i=1}^{m}\frac{f_i(x_i)}{g_i(x_i)} \quad \text{s.t.} \quad x_1=\cdots=x_m.$$

A plausible alternative optimization formulation:

$$\max_{x_1,\ldots,x_m\in C} -\gamma\sum_{i=2}^{m}\|x_1-x_i\|^2+\sum_{i=1}^{m}\frac{f_i(x_i)}{g_i(x_i)}.$$

Sum-of-ratios problems

Nonsmooth and nonconcave fractional maximization problem:

$$\max_{\mathbf{x}=(x_1,\ldots,x_m)\in S:=S_1\times\cdots\times S_m} F(\mathbf{x}) := h(x_1,\ldots,x_m) + \sum_{i=1}^{m}\frac{f_i(x_i)}{g_i(x_i)}. \qquad \text{(P)}$$

Here, for all $x_i \in S_i$, $f_i(x_i) \geq 0$ and $g_i(x_i) > 0$.

Classical sum-of-ratios optimization problem:

$$\max_{z\in C}\sum_{i=1}^{m}\frac{f_i(z)}{g_i(z)},$$

which can be rewritten as

$$\max_{x_1,\ldots,x_m\in C}\sum_{i=1}^{m}\frac{f_i(x_i)}{g_i(x_i)} \quad \text{s.t.} \quad x_1=\cdots=x_m.$$

A plausible alternative optimization formulation:

$$\max_{x_1,\ldots,x_m\in C} -\gamma\sum_{i=2}^{m}\|x_1-x_i\|^2+\sum_{i=1}^{m}\frac{f_i(x_i)}{g_i(x_i)}.$$

Sum-of-ratios problems

Nonsmooth and nonconcave fractional maximization problem:

$$\max_{\mathbf{x}=(x_1,\ldots,x_m)\in S:=S_1\times\cdots\times S_m} F(\mathbf{x}) := h(x_1,\ldots,x_m) + \sum_{i=1}^{m} \frac{f_i(x_i)}{g_i(x_i)}. \qquad \text{(P)}$$

Here, for all $x_i \in S_i$, $f_i(x_i) \geq 0$ and $g_i(x_i) > 0$.

Classical sum-of-ratios optimization problem:

$$\max_{z\in C} \sum_{i=1}^{m} \frac{f_i(z)}{g_i(z)},$$

which can be rewritten as

$$\max_{x_1,\ldots,x_m\in C} \sum_{i=1}^{m} \frac{f_i(x_i)}{g_i(x_i)} \quad \text{s.t.} \quad x_1 = \cdots = x_m.$$

A plausible alternative optimization formulation:

$$\max_{x_1,\ldots,x_m\in C} -\gamma \sum_{i=2}^{m} \|x_1 - x_i\|^2 + \sum_{i=1}^{m} \frac{f_i(x_i)}{g_i(x_i)}.$$

Motivation

✍ Energy efficiency maximization problem:

$$\max_{\mathbf{x}\in\mathbb{R}^d} \sum_{i=1}^{d} w_i \frac{R_i(\mathbf{x})}{P_i(\mathbf{x})} \quad \text{s.t.} \quad 0 \leq \mathbf{x} \leq \mathbf{x}_{\max},$$

where $\mathbf{x}$ is the transmit power, R_i is the data rate of the ith user device, P_i is the power consumption to achieve the data rate R_i, and the coefficients w_i are to weight the user devices' energy efficiency.

✍ Sparse generalized eigenvalue problem:

$$\max_{\mathbf{x}\in\mathbb{R}^d} \frac{\mathbf{x}^\top A\mathbf{x}}{\mathbf{x}^\top B\mathbf{x}} - \lambda\phi(\mathbf{x}) \quad \text{s.t.} \quad \|\mathbf{x}\| = 1,$$

where $A \succeq 0$, $B \succ 0$, $\lambda > 0$, and ϕ is a regularization function which induces sparsity of the solution.

Motivation

✍ Energy efficiency maximization problem:

$$\max_{\mathbf{x}\in\mathbb{R}^d} \sum_{i=1}^{d} w_i \frac{R_i(\mathbf{x})}{P_i(\mathbf{x})} \quad \text{s.t.} \quad 0 \le \mathbf{x} \le \mathbf{x}_{\max},$$

where $\mathbf{x}$ is the transmit power, R_i is the data rate of the ith user device, P_i is the power consumption to achieve the data rate R_i, and the coefficients w_i are to weight the user devices' energy efficiency.

✍ Sparse generalized eigenvalue problem:

$$\max_{\mathbf{x}\in\mathbb{R}^d} \frac{\mathbf{x}^\top A\mathbf{x}}{\mathbf{x}^\top B\mathbf{x}} - \lambda\phi(\mathbf{x}) \quad \text{s.t.} \quad \|\mathbf{x}\| = 1,$$

where $A \succeq 0$, $B \succ 0$, $\lambda > 0$, and ϕ is a regularization function which induces sparsity of the solution.

Single-ratio problems

$$\min_{x \in S} \frac{f(x)}{g(x)} \tag{1}$$

✍ Dinkelbach-type methods:

- In each iteration n, find an optimal solution x_{n+1} of the subproblem
$$\min_{x \in S} \left(f(x) - \theta_n g(x)\right), \quad \text{where } \theta_n := \frac{f(x_n)}{g(x_n)}. \tag{2}$$
- Require that f and g are smooth, f is convex, and g is concave.
- Solving each (2) is typically as expensive and difficult as solving (1).

✍ Boț–Csetnek (2017):

- Solve subproblem (2) by proximal gradient methods.
- Require that f is a convex function and g is a smooth function which is either concave or convex.

✍ Boț–D–Li (2021):

- Solve subproblem (2) by extrapolated proximal subgradient methods.
- Only require that g is weakly convex.

Single-ratio problems

$$\min_{x \in S} \frac{f(x)}{g(x)} \tag{1}$$

✍ Dinkelbach-type methods:

- In each iteration n, find an optimal solution x_{n+1} of the subproblem
$$\min_{x \in S} (f(x) - \theta_n g(x)), \quad \text{where } \theta_n := \frac{f(x_n)}{g(x_n)}. \tag{2}$$
- Require that f and g are smooth, f is convex, and g is concave.
- Solving each (2) is typically as expensive and difficult as solving (1).

✍ Boţ–Csetnek (2017):

- Solve subproblem (2) by proximal gradient methods.
- Require that f is a convex function and g is a smooth function which is either concave or convex.

✍ Boţ–D–Li (2021):

- Solve subproblem (2) by extrapolated proximal subgradient methods.
- Only require that g is weakly convex.

Single-ratio problems

$$\min_{x \in S} \frac{f(x)}{g(x)} \tag{1}$$

✍ Dinkelbach-type methods:

- In each iteration n, find an optimal solution x_{n+1} of the subproblem

$$\min_{x \in S} (f(x) - \theta_n g(x)), \quad \text{where } \theta_n := \frac{f(x_n)}{g(x_n)}. \tag{2}$$

- Require that f and g are smooth, f is convex, and g is concave.
- Solving each (2) is typically as expensive and difficult as solving (1).

✍ Boț–Csetnek (2017):

- Solve subproblem (2) by proximal gradient methods.
- Require that f is a convex function and g is a smooth function which is either concave or convex.

✍ Boț–D–Li (2021):

- Solve subproblem (2) by extrapolated proximal subgradient methods.
- Only require that g is weakly convex.

Single-ratio problems

$$\min_{x \in S} \frac{f(x)}{g(x)} \tag{1}$$

✍ Dinkelbach-type methods:

- In each iteration n, find an optimal solution x_{n+1} of the subproblem
$$\min_{x \in S} (f(x) - \theta_n g(x)), \quad \text{where } \theta_n := \frac{f(x_n)}{g(x_n)}. \tag{2}$$
- Require that f and g are smooth, f is convex, and g is concave.
- Solving each (2) is typically as expensive and difficult as solving (1).

✍ Boț–Csetnek (2017):

- Solve subproblem (2) by proximal gradient methods.
- Require that f is a convex function and g is a smooth function which is either concave or convex.

✍ Boț–D–Li (2021):

- Solve subproblem (2) by extrapolated proximal subgradient methods.
- Only require that g is weakly convex.

Outline

Standing assumptions

We assume throughout that S is closed but not necessarily convex, $-h$ is proper and lower semicontinuous, and, for each $i \in \{1, \dots, m\}$,

1. f_i is nonnegative and locally Lipschitz on an open set containing S_i and there exists $\alpha_i \geq 0$ such that, for all $x_i, z_i \in S_i$ and all $u \in \partial_L f_i(x_i)$,
$$\left\langle \frac{u}{2\sqrt{f_i(x_i)}}, z_i - x_i \right\rangle \leq \sqrt{f_i(z_i)} - \sqrt{f_i(x_i)} + \frac{\alpha_i}{2}\|z_i - x_i\|^2, \tag{A1}$$
whenever $f_i(x_i) > 0$;
2. g_i is locally Lipschitz on an open set containing S_i and there exists $\beta_i \geq 0$ such that, for all $x_i, z_i \in S_i$ and all $v \in \partial_L g_i(x_i)$,
$$\langle v, z_i - x_i \rangle \geq g_i(z_i) - g_i(x_i) - \frac{\beta_i}{2}\|z_i - x_i\|^2. \tag{A2}$$

✍ (A1) $\longleftarrow$ $\sqrt{f_i}$ is weakly convex or S_i is compact and f_i is differentiable with Lipschitz continuous gradient.

✍ (A2) $\longleftarrow$ g_i is differentiable with Lipschitz continuous gradient.

Standing assumptions

We assume throughout that S is closed but not necessarily convex, $-h$ is proper and lower semicontinuous, and, for each $i \in \{1, \ldots, m\}$,

1. f_i is nonnegative and locally Lipschitz on an open set containing S_i and there exists $\alpha_i \geq 0$ such that, for all $x_i, z_i \in S_i$ and all $u \in \partial_L f_i(x_i)$,
$$\left\langle \frac{u}{2\sqrt{f_i(x_i)}}, z_i - x_i \right\rangle \leq \sqrt{f_i(z_i)} - \sqrt{f_i(x_i)} + \frac{\alpha_i}{2}\|z_i - x_i\|^2, \tag{A1}$$
whenever $f_i(x_i) > 0$;
2. g_i is locally Lipschitz on an open set containing S_i and there exists $\beta_i \geq 0$ such that, for all $x_i, z_i \in S_i$ and all $v \in \partial_L g_i(x_i)$,
$$\langle v, z_i - x_i \rangle \geq g_i(z_i) - g_i(x_i) - \frac{\beta_i}{2}\|z_i - x_i\|^2. \tag{A2}$$

✍ (A1) $\longleftarrow$ $\sqrt{f_i}$ is weakly convex or S_i is compact and f_i is differentiable with Lipschitz continuous gradient.

✍ (A2) $\longleftarrow$ g_i is differentiable with Lipschitz continuous gradient.

Fractional vs. Non-fractional formulations

We consider

$$\max_{\mathbf{x}=(x_1,\ldots,x_m)\in S:=S_1\times\cdots\times S_m} h(x_1,\ldots,x_m)+\sum_{i=1}^{m}\frac{f_i(x_i)}{g_i(x_i)} \tag{P}$$

and

$$\max_{\substack{\mathbf{x}=(x_1,\ldots,x_m)\in S\\ \mathbf{y}=(y_1,\ldots,y_m)\in\mathbb{R}^m}} h(x_1,\ldots,x_m)+\sum_{i=1}^{m}\left[2y_i\sqrt{f_i(x_i)}-y_i^2 g_i(x_i)\right]. \tag{P$_1$}$$

Lemma

$\overline{\mathbf{x}}$ is a global solution for (P) if and only if $(\overline{\mathbf{x}},\overline{\mathbf{y}})$ is a global solution for (P$_1$), where $\overline{\mathbf{y}}=(\overline{y}_1,\ldots,\overline{y}_m)\in\mathbb{R}^m$ with $\overline{y}_i=\frac{\sqrt{f_i(\overline{x}_i)}}{g_i(\overline{x}_i)}$, in which case, both problems have the same optimal value.

Fractional vs. Non-fractional formulations

We consider

$$\max_{\mathbf{x}=(x_1,\ldots,x_m)\in S:=S_1\times\cdots\times S_m} h(x_1,\ldots,x_m)+\sum_{i=1}^{m}\frac{f_i(x_i)}{g_i(x_i)} \qquad \text{(P)}$$

and

$$\max_{\substack{\mathbf{x}=(x_1,\ldots,x_m)\in S\\ \mathbf{y}=(y_1,\ldots,y_m)\in\mathbb{R}^m}} h(x_1,\ldots,x_m)+\sum_{i=1}^{m}\left[2y_i\sqrt{f_i(x_i)}-y_i^2 g_i(x_i)\right]. \qquad (\mathrm{P}_1)$$

Lemma

$\overline{\mathbf{x}}$ *is a global solution for* (P) *if and only if* $(\overline{\mathbf{x}},\overline{\mathbf{y}})$ *is a global solution for* (P_1)*, where* $\overline{\mathbf{y}}=(\overline{y}_1,\ldots,\overline{y}_m)\in\mathbb{R}^m$ *with* $\overline{y}_i=\frac{\sqrt{f_i(\overline{x}_i)}}{g_i(\overline{x}_i)}$*, in which case, both problems have the same optimal value.*

Observation

$$\max_{\substack{\mathbf{x}=(x_1,\dots,x_m)\in S\\ \mathbf{y}=(y_1,\dots,y_m)\in\mathbb{R}^m}} H(\mathbf{x},\mathbf{y}) := h(x_1,\dots,x_m) + \sum_{i=1}^m \Big[\underbrace{2y_i\sqrt{f_i(x_i)} - y_i^2 g_i(x_i)}_{H_i(x_i,y_i)}\Big].$$

✍ $\max_{y_i\in\mathbb{R}} H(\mathbf{x},\mathbf{y}) \longrightarrow \max_{y_i\in\mathbb{R}} H_i(x_i,y_i) \longrightarrow$

$$y_{i,n} = \frac{\sqrt{f_i(x_{i,n})}}{g_i(x_{i,n})}.$$

✍ $\max_{x_i\in S_i} H(\mathbf{x},\mathbf{y}) \longrightarrow \max_{x_i\in S_i}\{h(x_1,\dots,x_m) + H_i(x_i,y_i)\} \longrightarrow$

$$\begin{aligned} x_{i,n+1} &\in \operatorname*{argmax}_{x_i\in S_i}\left\{h_{i,n+1}(x_i) - \tau_n\|x_i - x_{i,n} - \tfrac{1}{2\tau_n}w_{i,n}\|^2\right\} \\ &= \operatorname{prox}_{\frac{1}{2\tau_n}(-h_{i,n+1}+\iota_{S_i})}\left(x_{i,n} + \tfrac{1}{2\tau_n}w_{i,n}\right), \end{aligned}$$

where $h_{i,n+1}(x_i) := h(x_{1,n+1},\dots,x_{i-1,n+1},x_i,x_{i+1,n},\dots,x_{m,n})$ and $w_{i,n} \in \partial_L^x H_i(x_{i,n},y_{i,n})$.

$\operatorname{prox}_{\gamma h}(z) := \operatorname{argmin}_x(h(x) + \frac{1}{2\gamma}\|x-z\|^2)$.

The *indicator function* ι_C: $\iota_C(x) := 0$ if $x\in C$ and $\iota_C(x) := +\infty$ otherwise.

Observation

$$\max_{\substack{\mathbf{x}=(x_1,\dots,x_m)\in S\\ \mathbf{y}=(y_1,\dots,y_m)\in\mathbb{R}^m}} H(\mathbf{x},\mathbf{y}) := h(x_1,\dots,x_m) + \sum_{i=1}^m \Big[\underbrace{2y_i\sqrt{f_i(x_i)} - y_i^2 g_i(x_i)}_{H_i(x_i,y_i)}\Big].$$

✍ $\max_{y_i\in\mathbb{R}} H(\mathbf{x},\mathbf{y}) \longrightarrow \max_{y_i\in\mathbb{R}} H_i(x_i,y_i) \longrightarrow$

$$y_{i,n} = \frac{\sqrt{f_i(x_{i,n})}}{g_i(x_{i,n})}.$$

✍ $\max_{x_i\in S_i} H(\mathbf{x},\mathbf{y}) \longrightarrow \max_{x_i\in S_i} \{h(x_1,\dots,x_m) + H_i(x_i,y_i)\} \longrightarrow$

$$x_{i,n+1} \in \operatorname*{argmax}_{x_i\in S_i} \left\{ h_{i,n+1}(x_i) - \tau_n \|x_i - x_{i,n} - \tfrac{1}{2\tau_n} w_{i,n}\|^2 \right\}$$
$$= \operatorname{prox}_{\frac{1}{2\tau_n}(-h_{i,n+1}+\iota_{S_i})}\big(x_{i,n} + \tfrac{1}{2\tau_n} w_{i,n}\big),$$

where $h_{i,n+1}(x_i) := h(x_{1,n+1},\dots,x_{i-1,n+1},x_i,x_{i+1,n},\dots,x_{m,n})$ and $w_{i,n} \in \partial_L^x H_i(x_{i,n},y_{i,n})$.

$\operatorname{prox}_{\gamma h}(z) := \operatorname{argmin}_x(h(x) + \frac{1}{2\gamma}\|x-z\|^2)$.
The *indicator function* ι_C: $\iota_C(x) := 0$ if $x\in C$ and $\iota_C(x) := +\infty$ otherwise.

Observation

$$\max_{\substack{\mathbf{x}=(x_1,\dots,x_m)\in S\\ \mathbf{y}=(y_1,\dots,y_m)\in\mathbb{R}^m}} H(\mathbf{x},\mathbf{y}) := h(x_1,\dots,x_m) + \sum_{i=1}^m \underbrace{\left[2y_i\sqrt{f_i(x_i)} - y_i^2 g_i(x_i)\right]}_{H_i(x_i,y_i)}.$$

✍ $\max_{y_i\in\mathbb{R}} H(\mathbf{x},\mathbf{y}) \longrightarrow \max_{y_i\in\mathbb{R}} H_i(x_i,y_i) \longrightarrow$

$$y_{i,n} = \frac{\sqrt{f_i(x_{i,n})}}{g_i(x_{i,n})}.$$

✍ $\max_{x_i\in S_i} H(\mathbf{x},\mathbf{y}) \longrightarrow \max_{x_i\in S_i} \{h(x_1,\dots,x_m) + H_i(x_i,y_i)\} \longrightarrow$

$$\begin{aligned} x_{i,n+1} &\in \operatorname*{argmax}_{x_i\in S_i}\left\{h_{i,n+1}(x_i) - \tau_n\|x_i - x_{i,n} - \tfrac{1}{2\tau_n}w_{i,n}\|^2\right\} \\ &= \operatorname{prox}_{\frac{1}{2\tau_n}(-h_{i,n+1}+\iota_{S_i})}\left(x_{i,n} + \tfrac{1}{2\tau_n}w_{i,n}\right), \end{aligned}$$

where $h_{i,n+1}(x_i) := h(x_{1,n+1},\dots,x_{i-1,n+1},x_i,x_{i+1,n},\dots,x_{m,n})$ and $w_{i,n} \in \partial_L^x H_i(x_{i,n},y_{i,n})$.

$\operatorname{prox}_{\gamma h}(z) := \operatorname{argmin}_x(h(x) + \frac{1}{2\gamma}\|x-z\|^2)$.

The *indicator function* ι_C: $\iota_C(x) := 0$ if $x\in C$ and $\iota_C(x) := +\infty$ otherwise.

Inertial proximal subgradient method

① Let $n = 0$, $\mathbf{x}_{-1} = \mathbf{x}_0 = (x_{1,0}, \ldots, x_{m,0}) \in S$, $\delta > 0$, and $\overline{\nu} \in [0, \delta/2)$.

② Set $\mathbf{y}_n = (y_{1,n}, \ldots, y_{m,n})$ with

$$y_{i,n} = \frac{\sqrt{f_i(x_{i,n})}}{g_i(x_{i,n})},$$

Choose $\tau_n \geq \delta + \max_{i=1,\ldots,m}\{\frac{1}{2}(2y_{i,n}\alpha_i + y_{i,n}^2\beta_i)\}$ and $\nu_n \in [0, \overline{\nu}/\tau_n]$. For each $i \in \{1, \ldots, m\}$, let $z_{i,n} = x_{i,n} + \nu_n(x_{i,n} - x_{i,n-1})$, $u_{i,n} \in \partial_L f_i(x_{i,n})$, $v_{i,n} \in \partial_L g_i(x_{i,n})$, and

$$w_{i,n} = \begin{cases} y_{i,n}\frac{u_{i,n}}{\sqrt{f_i(x_{i,n})}} - y_{i,n}^2 v_{i,n} & \text{if } f_i(x_{i,n}) > 0, \\ 0 & \text{if } f_i(x_{i,n}) = 0. \end{cases}$$

Denote $h_{i,n+1}(x_i) := h(x_{1,n+1}, \ldots, x_{i-1,n+1}, x_i, x_{i+1,n}, \ldots, x_{m,n})$ and find

$$x_{i,n+1} \in \operatorname*{argmax}_{x_i \in S_i}\left\{h_{i,n+1}(x_i) - \tau_n\left\|x_i - z_{i,n} - \frac{1}{2\tau_n}w_{i,n}\right\|^2\right\}.$$

Update $\mathbf{x}_{n+1} = (x_{1,n+1}, \ldots, x_{m,n+1})$.

③ If a termination criterion is not met, let $n = n + 1$ and go to Step 2.

Subsequential convergence

From now on, suppose that F is bounded from above on S and that the set $\{\mathbf{x} \in S : F(\mathbf{x}) \geq F(\mathbf{x}_0)\}$ is bounded.

Theorem (Boț–D–Li)

1. $\forall n \in \mathbb{N}$, $F(\mathbf{x}_n) - \overline{\nu}\|\mathbf{x}_n - \mathbf{x}_{n-1}\|^2 \leq F(\mathbf{x}_{n+1}) - (\delta - \overline{\nu})\|\mathbf{x}_{n+1} - \mathbf{x}_n\|^2$.
2. *The sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ is bounded and $\sum_{n=0}^{+\infty} \|\mathbf{x}_{n+1} - \mathbf{x}_n\|^2 < +\infty$. Moreover, the sequence $(F(\mathbf{x}_n))_{n\in\mathbb{N}}$ is convergent.*
3. *If $\limsup_{n\to+\infty} \tau_n = \overline{\tau} < +\infty$ and h is continuous on $S \cap \operatorname{dom} h$, then every cluster point $\overline{\mathbf{x}}$ of $(\mathbf{x}_n)_{n\in\mathbb{N}}$ is a lifted coordinate-wise stationary point for* (P)*: For each $i \in \{1, \ldots, m\}$,*

$$0 \in \partial_L^{x_i}(-h + \iota_S)(\overline{\mathbf{x}}) + \frac{-g_i(\overline{x}_i)\partial_L f_i(\overline{x}_i) + f_i(\overline{x}_i)\partial_L g_i(\overline{x}_i)}{g_i(\overline{x}_i)^2},$$

where $\partial_L^{x_i}$ denotes the subdifferential w.r.t. x_i-variable

✍ The assumption on continuity of h can be removed if $m = 1$.

Convergence to a stationary point

Every cluster point $\bar{\mathbf{x}}$ of $(\mathbf{x}_n)_{n\in\mathbb{N}}$ is a stationary point for (P), i.e.,

$$0 \in \partial_L(-F + \iota_S)(\bar{\mathbf{x}})$$

provided that $\liminf_{n\to+\infty} \tau_n = \bar{\tau} > 0$ and one of the following holds:

1. $m = 1$, $-h + \iota_S$ and g_1 are regular on S, and f_1 is strictly differentiable on an open set containing S;
2. h is strictly differentiable on an open set containing S, ι_S is regular on S, and for each $i \in \{1, \dots, m\}$, f_i is strictly differentiable on an open set containing S_i and g_i is regular on S_i;
3. h is strictly differentiable on an open set containing S and, for each $i \in \{1, \dots, m\}$, f_i and g_i are strictly differentiable on an open set containing S_i.

Outline

Kurdyka–Łojasiewicz (KL) property

A proper lower semicontinuous function h is said to satisfy the KL property at $\overline{x} \in \operatorname{dom} \partial_L h$ if there exist a neighborhood U of $\overline{x}$, $\eta \in (0, +\infty]$, and a continuous concave function $\varphi\colon [0, \eta) \to \mathbb{R}_+$ such that

1. $\varphi(0) = 0$ and φ is continuously differentiable on $(0, \eta)$ with $\varphi' > 0$;
2. for all $x \in U$ with $h(\overline{x}) < h(x) < h(\overline{x}) + \eta$,

$$\varphi'(h(x) - h(\overline{x}))\operatorname{dist}(0, \partial_L h(x)) \geq 1. \tag{3}$$

If $\varphi(s) = \gamma s^{1-\alpha}$ for some $\gamma > 0$ and $\alpha \in [0, 1)$, then (3) becomes

$$\operatorname{dist}(0, \partial_L h(x)) \geq c|h(x) - h(\overline{x})|^{\alpha}, \tag{4}$$

and we say that h has the KL property with exponent α at $\overline{x}$.

- Semialgebraic functions: KL with some exponent $\alpha \in [0, 1)$.
- Piecewise linear quadratic functions: KL with exponent $1/2$.

Kurdyka–Łojasiewicz (KL) property

A proper lower semicontinuous function h is said to satisfy the KL property at $\overline{x} \in \operatorname{dom} \partial_L h$ if there exist a neighborhood U of $\overline{x}$, $\eta \in (0, +\infty]$, and a continuous concave function $\varphi \colon [0, \eta) \to \mathbb{R}_+$ such that

1. $\varphi(0) = 0$ and φ is continuously differentiable on $(0, \eta)$ with $\varphi' > 0$;
2. for all $x \in U$ with $h(\overline{x}) < h(x) < h(\overline{x}) + \eta$,

$$\varphi'(h(x) - h(\overline{x}))\, \operatorname{dist}(0, \partial_L h(x)) \geq 1. \tag{3}$$

If $\varphi(s) = \gamma s^{1-\alpha}$ for some $\gamma > 0$ and $\alpha \in [0, 1)$, then (3) becomes

$$\operatorname{dist}(0, \partial_L h(x)) \geq c|h(x) - h(\overline{x})|^{\alpha}, \tag{4}$$

and we say that h has the KL property with exponent α at $\overline{x}$.

- Semialgebraic functions: KL with some exponent $\alpha \in [0, 1)$.
- Piecewise linear quadratic functions: KL with exponent $1/2$.

Kurdyka–Łojasiewicz (KL) property

A proper lower semicontinuous function h is said to satisfy the KL property at $\overline{x} \in \operatorname{dom} \partial_L h$ if there exist a neighborhood U of $\overline{x}$, $\eta \in (0, +\infty]$, and a continuous concave function $\varphi\colon [0, \eta) \to \mathbb{R}_+$ such that

1. $\varphi(0) = 0$ and φ is continuously differentiable on $(0, \eta)$ with $\varphi' > 0$;
2. for all $x \in U$ with $h(\overline{x}) < h(x) < h(\overline{x}) + \eta$,

$$\varphi'(h(x) - h(\overline{x}))\,\operatorname{dist}(0, \partial_L h(x)) \geq 1. \tag{3}$$

If $\varphi(s) = \gamma s^{1-\alpha}$ for some $\gamma > 0$ and $\alpha \in [0, 1)$, then (3) becomes

$$\operatorname{dist}(0, \partial_L h(x)) \geq c|h(x) - h(\overline{x})|^{\alpha}, \tag{4}$$

and we say that h has the KL property with exponent α at $\overline{x}$.

- Semialgebraic functions: KL with some exponent $\alpha \in [0, 1)$.
- Piecewise linear quadratic functions: KL with exponent $1/2$.

Global convergence with KL property

Attouch–Bolte (2009), Attouch–Bolte–Svaiter (2013):

1. $\forall n \in \mathbb{N},\ h(x_{n+1}) + \alpha\|x_{n+1} - x_n\|^2 \leq h(x_n)$.
2. $\forall n \in \mathbb{N},\ \operatorname{dist}(0, \partial_L h(x_{n+1})) \leq \beta\|x_{n+1} - x_n\|$.
3. There exist a subsequence $(x_{k_n})_{n\in\mathbb{N}}$ and $\widetilde{x}$ such that $x_{k_n} \to \widetilde{x}$ and $h(x_{k_n}) \to h(\widetilde{x})$ as $n \to +\infty$.

Ochs (2019):

1. $\forall n \in \mathbb{N},\ h(x_{n+1}, u_{n+1}) + \alpha_n \Delta_n^2 \leq h(x_n, u_n)$.
2. $\forall n \in \mathbb{N},\ \beta_n \operatorname{dist}(0, \partial_L h(x_n, u_n)) \leq \beta \sum_{i\in I} \lambda_i \Delta_{n-i} + \varepsilon_n \quad (\sum_{i\in I} \lambda_i = 1)$.
3. There exist a subsequence $((x_{k_n}, u_{k_n}))_{n\in\mathbb{N}}$ and $(\widetilde{x}, \widetilde{u})$ such that $(x_{k_n}, u_{k_n}) \to (\widetilde{x}, \widetilde{u})$ and $h(x_{k_n}, u_{k_n}) \to h(\widetilde{x}, \widetilde{u})$ as $n \to +\infty$.
4. $\inf_{n\in\mathbb{N}} \alpha_n > 0$, $\inf_{n\in\mathbb{N}} \alpha_n\beta_n > 0$, $\sum_{n=1}^{+\infty} \beta_n = +\infty$, & $\sum_{n=1}^{+\infty} \varepsilon_n < +\infty$.
5. $\exists j \in \mathbb{Z},\ c \in \mathbb{R}$: $\forall n \in \mathbb{N},\ \|x_{n+1} - x_n\| \leq c\Delta_{n+j}$.

Global convergence with KL property

Attouch–Bolte (2009), Attouch–Bolte–Svaiter (2013):

1. $\forall n \in \mathbb{N},\ h(x_{n+1}) + \alpha\|x_{n+1} - x_n\|^2 \leq h(x_n)$.
2. $\forall n \in \mathbb{N},\ \operatorname{dist}(0, \partial_L h(x_{n+1})) \leq \beta\|x_{n+1} - x_n\|$.
3. There exist a subsequence $(x_{k_n})_{n\in\mathbb{N}}$ and $\widetilde{x}$ such that $x_{k_n} \to \widetilde{x}$ and $h(x_{k_n}) \to h(\widetilde{x})$ as $n \to +\infty$.

Ochs (2019):

1. $\forall n \in \mathbb{N},\ h(x_{n+1}, u_{n+1}) + \alpha_n \Delta_n^2 \leq h(x_n, u_n)$.
2. $\forall n \in \mathbb{N},\ \beta_n \operatorname{dist}(0, \partial_L h(x_n, u_n)) \leq \beta \sum_{i\in I} \lambda_i \Delta_{n-i} + \varepsilon_n \quad (\sum_{i\in I} \lambda_i = 1)$.
3. There exist a subsequence $((x_{k_n}, u_{k_n}))_{n\in\mathbb{N}}$ and $(\widetilde{x}, \widetilde{u})$ such that $(x_{k_n}, u_{k_n}) \to (\widetilde{x}, \widetilde{u})$ and $h(x_{k_n}, u_{k_n}) \to h(\widetilde{x}, \widetilde{u})$ as $n \to +\infty$.
4. $\inf_{n\in\mathbb{N}} \alpha_n > 0$, $\inf_{n\in\mathbb{N}} \alpha_n\beta_n > 0$, $\sum_{n=1}^{+\infty} \beta_n = +\infty$, & $\sum_{n=1}^{+\infty} \varepsilon_n < +\infty$.
5. $\exists j \in \mathbb{Z},\ c \in \mathbb{R}$: $\forall n \in \mathbb{N},\ \|x_{n+1} - x_n\| \leq c\Delta_{n+j}$.

$\mathcal{H}, \mathcal{K}$: finite-dimensional real Hilbert spaces,
$h\colon \mathcal{K} \to (-\infty, +\infty]$: a proper lower semicontinuous function,
$\forall n \in \mathbb{N}$, $x_n \in \mathcal{H}$, $z_n \in \mathcal{K}$, $\alpha_n, \beta_n > 0$, $\Delta_n, \varepsilon_n \geq 0$.

Let $\underline{\imath} \leq \bar{\imath}$ be integers and, for each $i \in I := \{\underline{\imath}, \underline{\imath}+1, \ldots, \bar{\imath}\}$, let $\lambda_i \geq 0$ with $\sum_{i\in I} \lambda_i = 1$. We set $\Delta_k = 0$ for $k < 0$ and assume that

1. (*Sufficient decrease*) $\forall n \in \mathbb{N}$, $h(z_{n+1}) + \alpha_n \Delta_n^2 \leq h(z_n)$.
2. (*Relative error*) $\forall n \in \mathbb{N}$, $\beta_n \operatorname{dist}(0, \partial_L h(z_n)) \leq \sum_{i\in I} \lambda_i \Delta_{n-i} + \varepsilon_n$.
3. (*Continuity*) There exist a subsequence $(z_{k_n})_{n\in\mathbb{N}}$ and $\widetilde{z}$ such that $z_{k_n} \to \widetilde{z}$ and $h(z_{k_n}) \to h(\widetilde{z})$ as $n \to +\infty$.
4. (*Parameter*) $\inf_{n\in\mathbb{N}} \alpha_n > 0$, $\inf_{n\in\mathbb{N}} \alpha_n\beta_n > 0$, and $\sum_{n=1}^{+\infty} \varepsilon_n < +\infty$.
5. (*Distance*) $\exists j \in \mathbb{Z}$, $c \in \mathbb{R}$: $\forall n \in \mathbb{N}$, $\|x_{n+1} - x_n\| \leq c\Delta_{n+j}$.

A unified analysis framework

$\mathcal{H}, \mathcal{K}$: finite-dimensional real Hilbert spaces,
$h\colon \mathcal{K} \to (-\infty, +\infty]$: a proper lower semicontinuous function,
$\forall n \in \mathbb{N}$, $x_n \in \mathcal{H}$, $z_n \in \mathcal{K}$, $\alpha_n, \beta_n > 0$, $\Delta_n, \varepsilon_n \geq 0$.

Let $\underline{\imath} \leq \bar{\imath}$ be integers and, for each $i \in I := \{\underline{\imath}, \underline{\imath}+1, \dots, \bar{\imath}\}$, let $\lambda_i \geq 0$ with $\sum_{i \in I} \lambda_i = 1$. We set $\Delta_k = 0$ for $k < 0$ and assume that

1. (*Sufficient decrease*) $\forall n \in \mathbb{N}$, $h(z_{n+1}) + \alpha_n \Delta_n^2 \leq h(z_n)$.
2. (*Relative error*) $\forall n \in \mathbb{N}$, $\beta_n \operatorname{dist}(0, \partial_L h(z_n)) \leq \sum_{i \in I} \lambda_i \Delta_{n-i} + \varepsilon_n$.
3. (*Continuity*) There exist a subsequence $(z_{k_n})_{n \in \mathbb{N}}$ and $\widetilde{z}$ such that $z_{k_n} \to \widetilde{z}$ and $h(z_{k_n}) \to h(\widetilde{z})$ as $n \to +\infty$.
4. (*Parameter*) $\inf_{n \in \mathbb{N}} \alpha_n > 0$, $\inf_{n \in \mathbb{N}} \alpha_n \beta_n > 0$, and $\sum_{n=1}^{+\infty} \varepsilon_n < +\infty$.
5. (*Distance*) $\exists j \in \mathbb{Z}$, $c \in \mathbb{R}$: $\forall n \in \mathbb{N}$, $\|x_{n+1} - x_n\| \leq c \Delta_{n+j}$.

Abstract convergence

Theorem (Boț–D–Li 2021)

Suppose that $(z_n)_{n\in\mathbb{N}}$ is a bounded sequence with set of cluster points Ω, and h is constant on Ω and satisfies the KL property at each point of Ω. Set $\Omega_0 := \{\overline{z} \in \Omega : h(z_n) \to h(\overline{z})$ as $n \to +\infty\}$ and $\overline{h} := h(z)$ for $z \in \Omega_0$. Then

1. $\sum_{n=0}^{+\infty} \|x_{n+1} - x_n\| < +\infty$, *and the sequence $(x_n)_{n\in\mathbb{N}}$ is convergent.*
2. *If* $\inf_{n\in\mathbb{N}} \beta_n > 0$, *then* $\forall \overline{z} \in \Omega_0$, $0 \in \partial_L h(\overline{z})$.
3. *Suppose further that h satisfies the KL property with exponent $\alpha \leq 1/2$ at every point of Ω, that $\underline{\imath} \leq 1$, and that*

$$\inf_{n\in\mathbb{N}} \beta_n > 0 \quad \text{and} \quad \varepsilon_n = O(h(z_{n-\overline{\imath}}) - h(z_{n+1-\underline{\imath}})) \text{ as } n \to +\infty.$$

Then $\exists \gamma > 0$, $\rho \in (0,1)$, *and* $\overline{x} \in \mathcal{H}$ *such that*

$$\forall n \in \mathbb{N}, \quad h(z_n) - \overline{h} \leq \gamma\rho^n \text{ and } \|x_n - \overline{x}\| \leq \gamma\rho^{\frac{n}{2}}.$$

Global and linear convergence

Theorem (Boț–D–Li)

Suppose that, for each $i \in \{1, \dots, m\}$, f_i and g_i are continuously differentiable on an open set containing S_i, and $\nabla \frac{f_i}{g_i}$ is Lipschitz continuous on S_i. Suppose further that $\limsup_{n\to+\infty} \tau_n = \overline{\tau} < +\infty$, that either $m = 1$ or h is differentiable on an open set containing S with Lipschitz continuous gradient on S, and that

$$G(\mathbf{x}, \mathbf{u}) = -F(\mathbf{x}) + \iota_S(\mathbf{x}) + \overline{\nu}\|\mathbf{x} - \mathbf{u}\|^2$$

satisfies KL property at $(\overline{\mathbf{x}}, \overline{\mathbf{x}})$ for all $\overline{\mathbf{x}} \in \operatorname{dom} \partial_L(-F + \iota_S)$. Then

1. $\sum_{n=0}^{+\infty} \|\mathbf{x}_{n+1} - \mathbf{x}_n\| < +\infty$, *and the sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ converges to a stationary point $\mathbf{x}^*$ for (P), i.e., $0 \in \partial_L(-F + \iota_S)(\mathbf{x}^*)$.*
2. *If G satisfies the KL property with exponent $\alpha \le 1/2$ at $(\overline{\mathbf{x}}, \overline{\mathbf{x}})$ for all $\overline{\mathbf{x}} \in \operatorname{dom} \partial_L(-F + \iota_S)$, then $\exists \gamma > 0$, $\rho \in (0, 1)$ such that*

$$\forall n \in \mathbb{N} \quad \|\mathbf{x}_n - \mathbf{x}^*\| \le \gamma\rho^{\frac{n}{2}} \quad \text{and} \quad \|F(\mathbf{x}_n) - F(\mathbf{x}^*)\| \le \gamma\rho^n.$$

Global and linear convergence

Theorem (Boț–D–Li)

Suppose that, for each $i \in \{1, \dots, m\}$, f_i *and* g_i *are continuously differentiable on an open set containing* S_i, *and* $\nabla \frac{f_i}{g_i}$ *is Lipschitz continuous on* S_i. *Suppose further that* $\limsup_{n\to+\infty} \tau_n = \overline{\tau} < +\infty$, *that either* $m = 1$ *or* h *is differentiable on an open set containing* S *with Lipschitz continuous gradient on* S, *and that*

$$G(\mathbf{x}, \mathbf{u}) = -F(\mathbf{x}) + \iota_S(\mathbf{x}) + \overline{\nu}\|\mathbf{x} - \mathbf{u}\|^2$$

satisfies KL property at $(\overline{\mathbf{x}}, \overline{\mathbf{x}})$ *for all* $\overline{\mathbf{x}} \in \operatorname{dom} \partial_L(-F + \iota_S)$. *Then*

1. $\sum_{n=0}^{+\infty} \|\mathbf{x}_{n+1} - \mathbf{x}_n\| < +\infty$, *and the sequence* $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *converges to a stationary point* $\mathbf{x}^*$ *for* (P), *i.e.,* $0 \in \partial_L(-F + \iota_S)(\mathbf{x}^*)$.
2. *If* G *satisfies the KL property with exponent* $\alpha \le 1/2$ *at* $(\overline{\mathbf{x}}, \overline{\mathbf{x}})$ *for all* $\overline{\mathbf{x}} \in \operatorname{dom} \partial_L(-F + \iota_S)$, *then* $\exists \gamma > 0$, $\rho \in (0, 1)$ *such that*

$$\forall n \in \mathbb{N} \quad \|\mathbf{x}_n - \mathbf{x}^*\| \le \gamma \rho^{\frac{n}{2}} \quad \text{and} \quad \|F(\mathbf{x}_n) - F(\mathbf{x}^*)\| \le \gamma \rho^n.$$

$$\max_{\mathbf{x}\in\mathbb{R}^d} \frac{\mathbf{x}^\top A\mathbf{x}}{\mathbf{x}^\top B\mathbf{x}} - \lambda\|\mathbf{x}\|_0 \quad \text{s.t.} \quad \|\mathbf{x}\| = 1, \tag{GEP}$$

where A is positive semidefinite, B is positive definite, and $\lambda > 0$.

The corresponding merit function takes the form

$$\widehat{\Phi}_{GEP}(\mathbf{x},\mathbf{u}) = \frac{\mathbf{x}^\top A_0\mathbf{x}}{\mathbf{x}^\top B\mathbf{x}} + \lambda\|\mathbf{x}\|_0 + \iota_\Lambda(\mathbf{x}) + \rho\|\mathbf{x}-\mathbf{u}\|^2,$$

where $A_0 = -A$ is symmetric, $\Lambda = \{\mathbf{x}\in\mathbb{R}^d : \|\mathbf{x}\| = 1\}$, and $\rho \geq 0$.

Theorem (Boț–D–Li)

Consider $\Phi(\mathbf{x}) = \frac{\mathbf{x}^\top A_0\mathbf{x}}{\mathbf{x}^\top B\mathbf{x}} + \lambda\|\mathbf{x}\|_0 + \iota_\Lambda(\mathbf{x})$, where A, B are symmetric matrices with B positive definite, and $\lambda > 0$. Then

1. *Φ is a KL function with exponent $1/2$.*
2. *For all $\rho \geq 0$, $\widehat{\Phi}_{GEP}$ satisfies the KL property with exponent $1/2$ at $(\overline{\mathbf{x}},\overline{\mathbf{x}})$ for all $\overline{\mathbf{x}} \in \operatorname{dom}\partial_L\Phi$.*

$$\max_{\mathbf{x}\in\mathbb{R}^d} \frac{\mathbf{x}^\top A\mathbf{x}}{\mathbf{x}^\top B\mathbf{x}} - \lambda\|\mathbf{x}\|_0 \quad \text{s.t.} \quad \|\mathbf{x}\| = 1, \tag{GEP}$$

where A is positive semidefinite, B is positive definite, and $\lambda > 0$.

The corresponding merit function takes the form

$$\widehat{\Phi}_{GEP}(\mathbf{x},\mathbf{u}) = \frac{\mathbf{x}^\top A_0\mathbf{x}}{\mathbf{x}^\top B\mathbf{x}} + \lambda\|\mathbf{x}\|_0 + \iota_\Lambda(\mathbf{x}) + \rho\|\mathbf{x}-\mathbf{u}\|^2,$$

where $A_0 = -A$ is symmetric, $\Lambda = \{\mathbf{x}\in\mathbb{R}^d : \|\mathbf{x}\| = 1\}$, and $\rho \geq 0$.

Theorem (Boț–D–Li)

Consider $\Phi(\mathbf{x}) = \frac{\mathbf{x}^\top A_0\mathbf{x}}{\mathbf{x}^\top B\mathbf{x}} + \lambda\|\mathbf{x}\|_0 + \iota_\Lambda(\mathbf{x})$, where A, B are symmetric matrices with B positive definite, and $\lambda > 0$. Then

1. *Φ is a KL function with exponent $1/2$.*
2. *For all $\rho \geq 0$, $\widehat{\Phi}_{GEP}$ satisfies the KL property with exponent $1/2$ at $(\bar{\mathbf{x}},\bar{\mathbf{x}})$ for all $\bar{\mathbf{x}} \in \operatorname{dom}\partial_L\Phi$.*

$$\max_{\mathbf{x}\in\mathbb{R}^d} \frac{\mathbf{x}^\top A\mathbf{x}}{\mathbf{x}^\top B\mathbf{x}} - \lambda\|\mathbf{x}\|_0 \quad \text{s.t.} \quad \|\mathbf{x}\| = 1, \qquad \text{(GEP)}$$

where A is positive semidefinite, B is positive definite, and $\lambda > 0$.

The corresponding merit function takes the form

$$\widehat{\Phi}_{GEP}(\mathbf{x},\mathbf{u}) = \frac{\mathbf{x}^\top A_0\mathbf{x}}{\mathbf{x}^\top B\mathbf{x}} + \lambda\|\mathbf{x}\|_0 + \iota_\Lambda(\mathbf{x}) + \rho\|\mathbf{x}-\mathbf{u}\|^2,$$

where $A_0 = -A$ is symmetric, $\Lambda = \{\mathbf{x}\in\mathbb{R}^d : \|\mathbf{x}\| = 1\}$, and $\rho \geq 0$.

Theorem (Boț–D–Li)

Consider $\Phi(\mathbf{x}) = \frac{\mathbf{x}^\top A_0\mathbf{x}}{\mathbf{x}^\top B\mathbf{x}} + \lambda\|\mathbf{x}\|_0 + \iota_\Lambda(\mathbf{x})$, *where* A, B *are symmetric matrices with* B *positive definite, and* $\lambda > 0$. *Then*

1. Φ *is a KL function with exponent* $1/2$.
2. *For all* $\rho \geq 0$, $\widehat{\Phi}_{GEP}$ *satisfies the KL property with exponent* $1/2$ *at* $(\overline{\mathbf{x}},\overline{\mathbf{x}})$ *for all* $\overline{\mathbf{x}} \in \operatorname{dom}\partial_L\Phi$.

$$\max_{x\in\mathbb{R}^d} \frac{\mathbf{x}^\top A\mathbf{x}}{\mathbf{x}^\top B\mathbf{x}} \quad \text{s.t.} \quad \|\mathbf{x}\| = 1,\ \|\mathbf{x}\|_0 \leq r, \tag{GEPS}$$

where A is positive semidefinite, B is positive definite, and $r > 0$.

The corresponding merit function takes the form

$$\widehat{\Phi}_{GEPS}(\mathbf{x},\mathbf{u}) = \frac{\mathbf{x}^\top A_0\mathbf{x}}{\mathbf{x}^\top B\mathbf{x}} + \iota_{\Lambda\cap C_r}(\mathbf{x}) + \rho\|\mathbf{x}-\mathbf{u}\|^2,$$

where $A_0 = -A$ is symmetric, $C_r = \{\mathbf{x}\in\mathbb{R}^d : \|\mathbf{x}\|_0 \leq r\}$, and $\rho \geq 0$.

Theorem (Boț–D–Li)

Consider $\Phi(\mathbf{x}) = \frac{\mathbf{x}^\top A_0\mathbf{x}}{\mathbf{x}^\top B\mathbf{x}} + \iota_{\Lambda\cap C_r}(\mathbf{x})$, where A, B are symmetric matrices with B positive definite, and $r > 0$. Then

1. *Φ is a KL function with exponent $1/2$.*
2. *For all $\rho \geq 0$, $\widehat{\Phi}_{GEPS}$ satisfies the KL property with exponent $1/2$ at $(\overline{\mathbf{x}},\overline{\mathbf{x}})$ for all $\overline{\mathbf{x}} \in \operatorname{dom}\partial_L\Phi$.*

$$\max_{x\in\mathbb{R}^d} \frac{\mathbf{x}^\top A\mathbf{x}}{\mathbf{x}^\top B\mathbf{x}} \quad \text{s.t.} \quad \|\mathbf{x}\| = 1,\ \|\mathbf{x}\|_0 \leq r, \tag{GEPS}$$

where A is positive semidefinite, B is positive definite, and $r > 0$.

The corresponding merit function takes the form

$$\widehat{\Phi}_{GEPS}(\mathbf{x},\mathbf{u}) = \frac{\mathbf{x}^\top A_0\mathbf{x}}{\mathbf{x}^\top B\mathbf{x}} + \iota_{\Lambda\cap C_r}(\mathbf{x}) + \rho\|\mathbf{x}-\mathbf{u}\|^2,$$

where $A_0 = -A$ is symmetric, $C_r = \{\mathbf{x}\in\mathbb{R}^d : \|\mathbf{x}\|_0 \leq r\}$, and $\rho \geq 0$.

Theorem (Boț–D–Li)

Consider $\Phi(\mathbf{x}) = \frac{\mathbf{x}^\top A_0\mathbf{x}}{\mathbf{x}^\top B\mathbf{x}} + \iota_{\Lambda\cap C_r}(\mathbf{x})$, where A, B are symmetric matrices with B positive definite, and $r > 0$. Then

1. *Φ is a KL function with exponent $1/2$.*
2. *For all $\rho \geq 0$, $\widehat{\Phi}_{GEPS}$ satisfies the KL property with exponent $1/2$ at $(\overline{\mathbf{x}},\overline{\mathbf{x}})$ for all $\overline{\mathbf{x}} \in \operatorname{dom}\partial_L\Phi$.*

$$\max_{x\in\mathbb{R}^d} \frac{\mathbf{x}^\top A\mathbf{x}}{\mathbf{x}^\top B\mathbf{x}} \quad \text{s.t.} \quad \|\mathbf{x}\| = 1,\ \|\mathbf{x}\|_0 \le r, \tag{GEPS}$$

where A is positive semidefinite, B is positive definite, and $r > 0$.

The corresponding merit function takes the form

$$\widehat{\Phi}_{GEPS}(\mathbf{x},\mathbf{u}) = \frac{\mathbf{x}^\top A_0\mathbf{x}}{\mathbf{x}^\top B\mathbf{x}} + \iota_{\Lambda\cap C_r}(\mathbf{x}) + \rho\|\mathbf{x}-\mathbf{u}\|^2,$$

where $A_0 = -A$ is symmetric, $C_r = \{\mathbf{x}\in\mathbb{R}^d : \|\mathbf{x}\|_0 \le r\}$, and $\rho \ge 0$.

Theorem (Boţ–D–Li)

Consider $\Phi(\mathbf{x}) = \frac{\mathbf{x}^\top A_0\mathbf{x}}{\mathbf{x}^\top B\mathbf{x}} + \iota_{\Lambda\cap C_r}(\mathbf{x})$, *where* A, B *are symmetric matrices with* B *positive definite, and* $r > 0$. *Then*

1. Φ *is a KL function with exponent* $1/2$.
2. *For all* $\rho \ge 0$, $\widehat{\Phi}_{GEPS}$ *satisfies the KL property with exponent* $1/2$ *at* $(\bar{\mathbf{x}},\bar{\mathbf{x}})$ *for all* $\bar{\mathbf{x}} \in \operatorname{dom}\partial_L\Phi$.

Outline

Sparse generalized eigenvalue problem

Consider p observations $\mathbf{z}_1, \ldots, \mathbf{z}_p \in \mathbb{R}^d$, each of which belongs to one of two distinct classes.

Let $I_k \subseteq \{1, \ldots, p\}$ contain the indices of the observations in class k, with $p_k = |I_k|$, $k = 1, 2$, and $p_1 + p_2 = p$. Let $\widehat{\boldsymbol{\mu}}_k = \frac{1}{p_k} \sum_{i \in I_k} \mathbf{z}_i$, for $k = 1, 2$.

The so-called within-class and between-class covariance matrices are

$$V_w = \frac{1}{p} \sum_{k=1}^{2} \sum_{i \in I_k} (\mathbf{z}_i - \widehat{\boldsymbol{\mu}}_k)(\mathbf{z}_i - \widehat{\boldsymbol{\mu}}_k)^\top \quad \text{and} \quad V_b = \frac{1}{p} \sum_{k=1}^{2} p_k \widehat{\boldsymbol{\mu}}_k \widehat{\boldsymbol{\mu}}_k^\top.$$

The classification problem using sparse Fisher discriminant analysis:

$$\max_{\mathbf{x} \in \mathbb{R}^d} \frac{\mathbf{x}^\top V_b \mathbf{x}}{\mathbf{x}^\top V_w \mathbf{x}} - \lambda \phi(\mathbf{x}) \quad \text{s.t.} \quad \|\mathbf{x}\| = 1, \tag{SFDA}$$

where ϕ is a regularization function inducing sparsity, and $\lambda > 0$.

- Truncated Rayleigh flow method (TRFM) proposed by Tan *et al.* (2018): Local linear convergence for $\phi = \iota_{C_r}$ with $C_r := \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}_0\| \le r\}$.
- Our method: Global linear convergence for $\phi = \iota_{C_r}$ or $\phi = \|\cdot\|_0$.

Sparse generalized eigenvalue problem

Consider p observations $\mathbf{z}_1, \ldots, \mathbf{z}_p \in \mathbb{R}^d$, each of which belongs to one of two distinct classes.

Let $I_k \subseteq \{1, \ldots, p\}$ contain the indices of the observations in class k, with $p_k = |I_k|$, $k = 1, 2$, and $p_1 + p_2 = p$. Let $\widehat{\boldsymbol{\mu}}_k = \frac{1}{p_k} \sum_{i \in I_k} \mathbf{z}_i$, for $k = 1, 2$.

The so-called within-class and between-class covariance matrices are

$$V_w = \frac{1}{p} \sum_{k=1}^{2} \sum_{i \in I_k} (\mathbf{z}_i - \widehat{\boldsymbol{\mu}}_k)(\mathbf{z}_i - \widehat{\boldsymbol{\mu}}_k)^\top \quad \text{and} \quad V_b = \frac{1}{p} \sum_{k=1}^{2} p_k \widehat{\boldsymbol{\mu}}_k \widehat{\boldsymbol{\mu}}_k^\top.$$

The classification problem using sparse Fisher discriminant analysis:

$$\max_{\mathbf{x} \in \mathbb{R}^d} \frac{\mathbf{x}^\top V_b \mathbf{x}}{\mathbf{x}^\top V_w \mathbf{x}} - \lambda \phi(\mathbf{x}) \quad \text{s.t.} \quad \|\mathbf{x}\| = 1, \qquad \text{(SFDA)}$$

where ϕ is a regularization function inducing sparsity, and $\lambda > 0$.

- Truncated Rayleigh flow method (TRFM) proposed by Tan *et al.* (2018): Local linear convergence for $\phi = \iota_{C_r}$ with $C_r := \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}_0\| \le r\}$.
- Our method: Global linear convergence for $\phi = \iota_{C_r}$ or $\phi = \|\cdot\|_0$.

Sparse generalized eigenvalue problem

Consider p observations $\mathbf{z}_1, \ldots, \mathbf{z}_p \in \mathbb{R}^d$, each of which belongs to one of two distinct classes.

Let $I_k \subseteq \{1, \ldots, p\}$ contain the indices of the observations in class k, with $p_k = |I_k|$, $k = 1, 2$, and $p_1 + p_2 = p$. Let $\widehat{\boldsymbol{\mu}}_k = \frac{1}{p_k} \sum_{i \in I_k} \mathbf{z}_i$, for $k = 1, 2$.

The so-called within-class and between-class covariance matrices are

$$V_w = \frac{1}{p} \sum_{k=1}^{2} \sum_{i \in I_k} (\mathbf{z}_i - \widehat{\boldsymbol{\mu}}_k)(\mathbf{z}_i - \widehat{\boldsymbol{\mu}}_k)^\top \quad \text{and} \quad V_b = \frac{1}{p} \sum_{k=1}^{2} p_k \widehat{\boldsymbol{\mu}}_k \widehat{\boldsymbol{\mu}}_k^\top.$$

The classification problem using sparse Fisher discriminant analysis:

$$\max_{\mathbf{x} \in \mathbb{R}^d} \frac{\mathbf{x}^\top V_b \mathbf{x}}{\mathbf{x}^\top V_w \mathbf{x}} - \lambda \phi(\mathbf{x}) \quad \text{s.t.} \quad \|\mathbf{x}\| = 1, \qquad \text{(SFDA)}$$

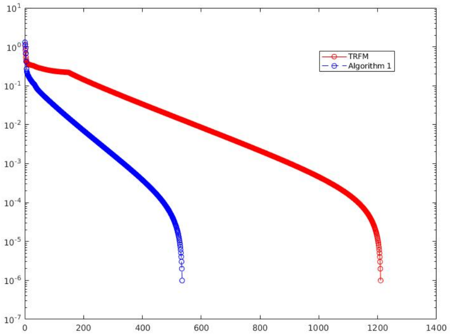
where ϕ is a regularization function inducing sparsity, and $\lambda > 0$.

- Truncated Rayleigh flow method (TRFM) proposed by Tan *et al.* (2018): Local linear convergence for $\phi = \iota_{C_r}$ with $C_r := \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}_0\| \leq r\}$.
- Our method: Global linear convergence for $\phi = \iota_{C_r}$ or $\phi = \|\cdot\|_0$.

Sparsity constraint

Adopting the same setting as in Tan *et al.* (2018), we run TRFM and our method (Algorithm 1) for 50 trials.

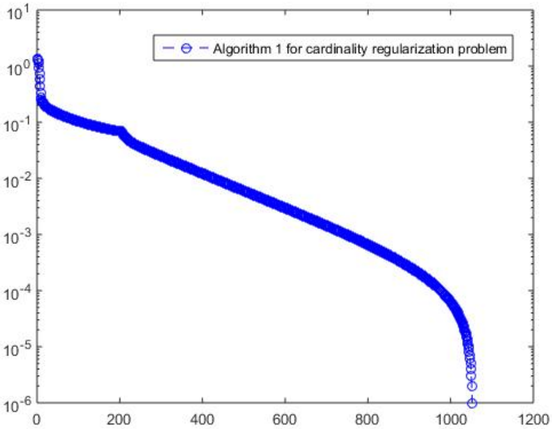
	Sparsity level of computed sol.	Objective value of the computed sol.	CPU time	Number of iterations
TRFM	26	11.5051	6.1950	1202
Our method	23	12.5158	3.9684	564



Euclidean distance between $\mathbf{x}_n$ and $\mathbf{x}^*$ in every iteration

Cardinality regularization

Sparsity level of computed sol.	Objective value of the computed sol.	CPU time	Number of Iterations
22	12.3854	4.7104	1053



Euclidean distance between $\mathbf{x}_n$ and $\mathbf{x}^*$ in every iteration

Some key references

- H. Attouch and J. Bolte, On the convergence of the proximal algorithm for nonsmooth functions involving analytic features, *Math. Program. Ser. B* **116**(1–2), 5–16 (2009).
- H. Attouch, J. Bolte, and B.F. Svaiter, Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods, *Math. Program. Ser. A* **137**(1–2), 91–129 (2013).
- R.I. Boț and E.R. Csetnek, Proximal-gradient algorithms for fractional programming, *Optimization* **66**(8), 1383–1396 (2017).
- R.I. Boț, MND, and G. Li, Extrapolated proximal subgradient algorithms for nonconvex and nonsmooth fractional programs, *Math. Oper. Res.*, DOI: 10.1287/moor.2021.1214.
- R.I. Boț, MND, and G. Li, Inertial proximal block coordinate method for a class of nonsmooth and nonconvex sum-of-ratios optimization problems, arXiv:2011.09782.
- G. Li and T.K. Pong, Calculus of the exponent of Kurdyka–Łojasiewicz inequality and its applications to linear convergence of first-order methods, *Found. Comput. Math.* **18**(5), 1199–1232 (2018).
- P. Ochs, Unifying abstract inexact convergence theorems and block coordinate variable metric iPiano, *SIAM J. Optim.* **29**(1), 541–570 (2019).
- K.M. Tan, Z. Wang, H. Liu, and T. Zhang, Sparse generalized eigenvalue problem: Optimal statistical rates via truncated Rayleigh flow, *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80**(5), 1057–1086 (2018).

THANK YOU VERY MUCH!