Sharp Penalty Mapping Approach to Approximate Solution of Variational Inequalities [1]

Down with Penalty Functions !

E.A. Nurminski
nurminskiy.ea@dvfu.ru

Far Eastern Federal University, School of Natural Sciences,
Ajax St., Vladivostok, Russky Island, Russia

17th Baikal International Triennial School-Seminar
"Methods of Optimization and Their Applications"
July 31–August 6, 2017, Maksimikha Bay, Buryatia)

July 31, 2017

[1] This work is supported by Ministry of Science and Education of RF, project 1.7658.2017/6.7

The Outline

- Motivations, notations and basic preliminaries;
- Superposing feasibility and optimality;
- Oriented and penalty mappings;
- Main reduction result;
- Algorithmic news.

Transportation problems

Main problem: predict network load.
Mainstream model: noncooperative equilibrium.
Equilibrium: such network load pattern, that nobody gains from changes in its transportation plans.
Specifics:

- High dimensionality
- Strong nonlinearity

Classic flow equilibrium model (BMW, 1950+)

Setup:

- Transportation network: a directed graph $G = (V, E)$;
- SD-pairs: $W = S \times D$, supply–demand transportation requests, $S, D \subset V$;
- Demand pattern: $d : W \to \mathbb{R}_+$;
- $P_w, w \in W$ — set of routs for a transportation request w over the network G;
- Unknowns: $x = \{x_p, p \in P_w, w \in W\}$ — very large set of variables;

Equilibrium

no one route p wants to change its load as it negatively effects its terms of delivery.

Costs and Effects for BMW-model

For any $e \in E$ given vector $x = \{x_p, p \in P_w, w \in W\}$ calculate the edge load y_e:

$$y_e = \sum_{p \in P^e} x_p, \;\; P^e \text{ is a set of routes, going by the edge } e \in E.$$

determine the delay $\tau_e(\cdot)$ on this edge:

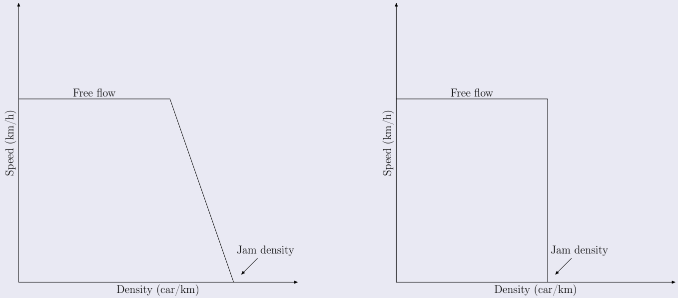
$$\tau_e(x) = \Phi_e(y_e) = \Phi_e(y_e(x)).$$

This delay takes place for averybody on this edge e, so the general situation can be described by the following picture.

Fundamental Diagrams

Delay-Flow dependences are collectively known under the name

Fundamental Diagrams



Féjer operators and processes

Definition

An operator $F_X : E \to E$ is called Féjer (with respect to a given nonempty set X) if for any $z \in X$

$$\|F_X(x) - z\| \leq \|x - z\|.$$

Let $\mathrm{Fix}(F_X)$ be a set of fixed points of operator F_X.

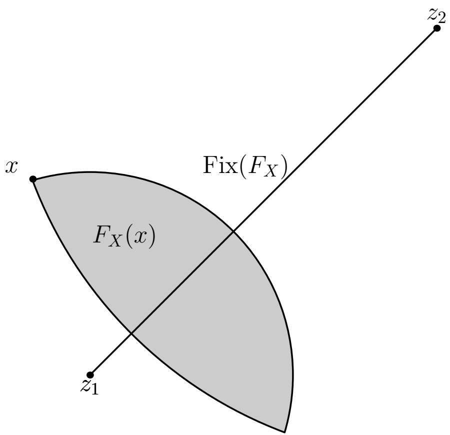
Theorem (Féjer, 1922)

$$\mathrm{Fix}(F_X) = \mathrm{co}(X)$$

References

- Féjer, L. (1922). Über die Lage der Nullstellen von Polynomen, die aus Minimumforderungen gewisser Art entspringen. Mathematische Annalen, 85(1), 41–48.
- Eremin, I. I. (2011). Methods for solving systems of linear and convex inequalities based on the Féjer principle. Proceedings of the Steklov Institute of Mathematics, 272(1), S36–S45.

Structure of a Fejer operator $F_X, \; X = \{z_1, z_2\}$



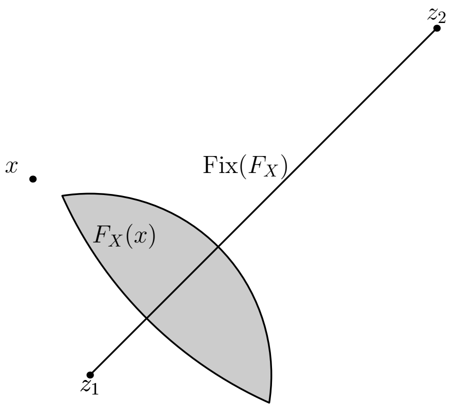
Locally strong Féjer operator

To ensure convergence of FP toward a goal set V stronger attraction properties are required.

Definition

A Féjer operator F_X is called locally strong Féjer if for any $\bar{x} \notin V$ there exists a neighborhood of zero U and $\alpha < 1$ such that $\|F_X(x) - v\| \leq \alpha\|x - v\|$ for any $v \in V$ and $x \in \bar{x} + U$.

Structure of a locally strong Féjer operator



Féjer processes

Féjer processes (FP) are defined by the recursive relationship

$$x^{k+1} = F_X(x^k), k = 0, 1, \dots$$

where x^0 is some starting point.
Define distance $\text{dist}(x, X) = \min_{z \in X} \|z - x\|$.

Theorem

Let the sequence $\{x^k, k = 1, 2, \dots\}$ is generated by the recursive correspondence $x^{k+1} = F_X(x^k), k = 0, 1, \dots$ with arbitrary x^0 and locally strong Féjer operator F_X. Then $dist(x^k, X) \to 0$ when $k \to \infty$.

Féjer processes with disturbances

FP with disturbances:

$$x^{k+1} = F_X(x^k + z^k), k = 0, 1, \ldots$$

where $z^k \to 0$ is an *arbitrary* diminishing disturbance. Major result:

Theorem

If $F_X(\cdot)$ is a locally strong Féjer operator with respect to X then $dist(x^k, X) \to 0$ when $k \to \infty$.

Assuming some additional conditions wrt $\{z^k, k = 0, 1, \ldots\}$ one can make the sequence $\{x^k, k = 0, 1, \ldots\}$ to converge to specific parts of X.

Use of disturbances: general idea

Selective Feasibility Problem: find $x^\star \in X_\star \subset X$
Examples: constrained optimization, VIP, etc

Split SFP into 2 problems:

1. General Feasibility: $x^\star \in X$
$$\text{solved by } x^{k+1} = F_X(x^k),\ \ k = 1, 2, \dots$$
2. Selective Feasibility: $x^\star \in X_\star$
$$\text{solved by } x^{k+1} = F_X(x^k + z^k),$$
$$z^k = \lambda_k G(x^k), \lambda_k \to 0$$

If $G(\cdot)$ in a certain way is "pointing toward" $X_\star$ then we might have a chance to converge to $X_\star$!

Attractants

Definition

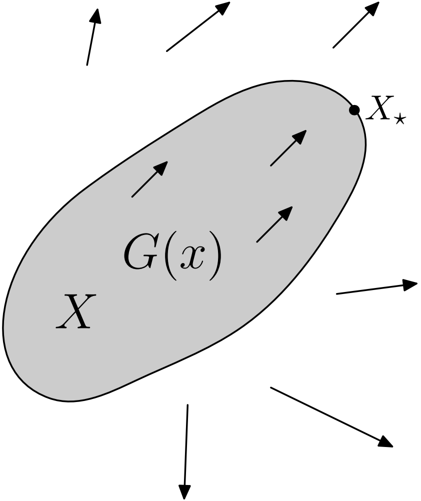
Set-valued mapping $D : E \to 2^E$ is called a strong locally restricted attractant of $X_\star \subset X$ if for each $x' \in X \setminus X_\star$ there exists a neighborhood of zero U such that,

$$g(z - x) \geq \delta > 0$$

for all $z \in X_\star, x \in x' + U, g \in D(x)$ and some $\delta > 0$.

Examples of such attractants are sub-differentials of convex functions and strongly monotone operators of variational inequalities.

Attractant mapping



VIP superposing — general idea

Variational inequality problem

$$G(x^\star)(x - x^\star) \geq 0, \quad x \in X$$

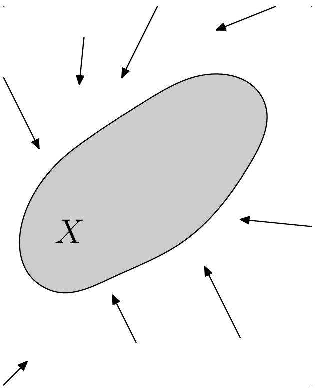
superposed as 2 problems:

1. Feasibility: $x^\star \in X$
   $F_X(\cdot)$ — projection, penalty functions, ...
2. Optimality: $G(\cdot)$ — VIP operator, gradient, ...

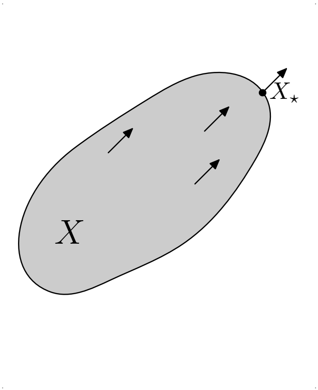
Resulting algorithms:

$$x^{k+1} = F_X(x^k + \lambda_k G(x^k)), k = 1, 2, \ldots$$

VIP split view



Feasibility mapping



Optimality mapping

VIP superposing — convergence conditions

Variational inequality problem

$$G(x^\star)(x - x^\star) \geq 0, \quad x \in X$$

Resulting algorithms:

$$x^{k+1} = F_X(x^k + \lambda_k G(x^k)), k = 1, 2, \ldots$$

Theorem

Let F_X — locally strong Féjer operator, G — a strong locally restricted attractant of $X_\star \subset X$ and $\lambda_k \to 0$ when $k \to \infty$, $\sum_k \lambda_k = \infty$. Then $dist(x^k, X_\star) \to 0$ when $\to \infty$.

Shortcomings

- stepsize does not adapt itself to the concrete problem;
- convergence rate is of the order of $O(1/k)$;
- disbalance between feasibility and optimality increases when $\lambda_k \to 0$ as $k \to \infty$.

What can be done ?

- different ideas for stepsize regulation (quite computationaly expensive);
- smoothing techniques;
- approximate solutions;
- something else.

VIP via Optimization

$$G(x)(x-z) \geq 0, x \in X, \forall z \in X \rightleftarrows \min F(x), x \in X$$

Merit and gap functions:

- $F(x) = \max G(x)(x-z), z \in X$ Auslender, 1976
- "Saddle" function $L(x,z) = (f(x) - f(z) + (G(x) - f'(x))(x-z)$ Aucmuty, 1989 Larsson-Ptriksson, 1994
- $F(x) = -\min_{z \in x - X}\{G(x)z + \frac{1}{2}zHz\}, z \in X$, Fukushima, 1992, 1996
- $F(x) = \phi_\alpha(x) - \phi_\beta(x), \phi_\alpha(x) = \max_{z \in x - X}\{G(x)z + \frac{1}{2\alpha}zHz\}$ Peng, 1997, see also Konnov-Penyagina.

VIP and PVIP

Find $x^\star \in X$ such that:
$G(x^\star)(x - x^\star) \geq 0$ **VIP** $\quad G(x)(x - x^\star) \geq 0$ **(PVIP)**
for all $x \in X$.

Important

If G is *monotone*, then any solution of **PVIP** is a solution of **VIP**.

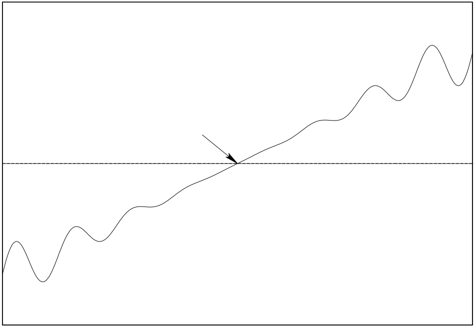
Assume that::

- $G(x)$ is monotone,
- **VIP** and **PVIP** have unique (and therefore conisiding) solutions

Oriented mappings

Let $\mathcal{C}(E)$ is the space of convex compacts of E, and $G : X \to \mathcal{C}(E)$.

$$(g(x) - g(x^\star))(x - x^\star) \geq 0$$
$$\text{for all } x \in X \text{ and } g(x) \in G(x), g(x^\star) \in G(x^\star)$$



Simple example

Strongly oriented mappings

Definition

A set-valued mapping $G : E \to \mathcal{C}(E)$ is called strongly oriented toward $\bar{x}$ on a set X if for any $\epsilon > 0$ there is $\gamma_\epsilon > 0$ such that

$$g_x(x - \bar{x}) \geq \gamma_\epsilon$$

for any $g_x \in G(x)$ and all $x \in X \setminus \{\bar{x} + \epsilon B\}$.

If G is oriented (strongly oriented) toward $\bar{x}$ at all points $x \in X$ then we will call it oriented (strongly oriented) toward $\bar{x}$ on X.

Note: if $\bar{x} = x^\star$, a solution of **PVIP**, then G is oriented toward $x^\star$ on X by definition and the other way around.

Composition of oriented and feasibility mappings

Let F_X — feasibility, $G(x)$ — oriented "optimality" mappings and

$$G(x, \epsilon) = \epsilon G(x) + P_X(x).$$

Under rather common conditions

- $\mathrm{Fix}(G(\cdot, \epsilon_k)) \to x^\star \in X_\star$ when $\epsilon_k \to +0, \ \sum_k \epsilon_k = \infty$.
- $\mathrm{Fix}(G(\cdot, \epsilon)) \subset X_\star + \gamma_\epsilon B$ with $\gamma_\epsilon \sim O(\epsilon)$.

Long-range orientation

To ensure the desirable global behavior of iteration methods we need an additional technical assumption.

Definition

A mapping $G : E \to E$ is called long-range oriented toward a set X if there exists $\rho_G \geq 0$ such that for any $\bar{x} \in X$

$$G(x)(x - \bar{x}) > 0 \text{ for all } x \text{ such that } \|x\| \geq \rho_G \quad (1)$$

.

We will call ρ_G the radius of long-range orientation of G toward X.

Penalty: modified polar

Definition

The set $K_X(x) = \{p : p(x - y) \geq 0 \text{ for all } y \in X\}$ we will call the polar cone of X at a point x.

Enforced polarity:

Definition

Let $\epsilon \geq 0$ and $x \notin X + \epsilon B$. The set

$$K_X^\epsilon(x) = \{p : p(x - y) \geq 0 \text{ for all } y \in X + \epsilon B\}$$

will be called ϵ-strong polar cone of X at x.

Globalization

Define a composite upper semicontinuos mapping for the whole E:

$$\tilde{K}^{\epsilon}_X(x) = \begin{cases} \{0\} & \text{if } x \in X \\ K_X(x) & \text{if } x \in \operatorname{cl}\{\{X + \epsilon B\} \setminus X\} \\ K^{\epsilon}_X(x) & \text{if } x \in \rho_F B \setminus \{X + \epsilon B\} \end{cases}$$

Sharp penalty mapping

Now we define a sharp penalty mapping for X as

$$P_X^\epsilon(x) = \begin{cases} \tilde{K}_X^\epsilon(x) \cap p : \|p\| = 1 & x \notin \text{int}\{X\} \\ \{0\} & \text{otherwise.} \end{cases}$$

A key lemma

Lemma

Let $X \subset E$ is closed and bounded, G is monotone and long-range oriented toward X with the radius of orientability ρ_G and strongly oriented toward solution $x^\star$ of **PVIP** *on X with the constants $\gamma_\epsilon > 0$ for $\epsilon > 0$, satisfies conditions of the slide (24) and $P_X^\epsilon(\cdot)$ is a sharp penalty of the slide 29. Then for any sufficiently small $\epsilon > 0$ there exists $\lambda_\epsilon > 0$ and $\delta_\epsilon > 0$ such that for all $\lambda \geq \lambda_\epsilon$ a penalized mapping $G_\lambda(x) = G(x) + \lambda P_X^\epsilon(x)$ satisfies the inequality $g_x(x - x^\star) \geq \delta_\epsilon$ for all $x \in \rho_G B \setminus \{x^\star + \epsilon B\}$ and any $g_x \in G_\lambda(x)$.*

The idea of the proof

Define the following subsets of E:

$$X_\epsilon^{(1)} = X \setminus \{x^\star + \epsilon B\},$$

$$X_\epsilon^{(2)} = \{\{X + \epsilon B\} \setminus X\} \setminus \{x^\star + \epsilon B\},$$

$$X_\epsilon^{(3)} = \rho_G B \setminus \{\{X + \epsilon B\} \setminus \{x^\star + \epsilon B\}\}.$$

which cover $\rho_G B \setminus \{x^\star + \epsilon B\}$ and show that there is λ_ϵ which guarantees

$$g_x(x - x^\star) \geq \delta_\epsilon > 0$$

in each of these subsets for any $g_x \in G_\lambda(x)$.

Algorithmic details: polar cone

The most common ways:

- by projection onto set X:

$$x - \Pi_X(x) \in K_X(x)$$

  where $\Pi_X(x) \in X$ is the orthogonal projection of x on X,
- by subdifferential calculus if $X = \{x : h(x) \leq 0\}$. Under Slater condition $h(y) < 0$ for all $y \in \text{int}\{X\}$ and

$$0 < h(x) - h(y) \leq g_h(x)(x - y) \text{ for any } y \in \text{int}\{X\}.$$

  By continuity $0 < h(x) - h(y) \leq g_h(x)(x - y)$ for all $y \in X$ which means that $g_h \in K_X(x)$.

Algorithmic details: Minkowski projection

Find some $x^c \in \text{int}\{X\}$ and use it to compute Minkowski function

$$\mu_X(x, x^c) = \inf_{\theta \geq 0}\{\theta : x^c + (x - x^c)\theta^{-1} \in X\} > 1 \text{ for } x \notin X.$$

Then by construction $\bar{x} = x^c + (x - x^c)\mu_X(x, x^c)^{-1} \in \partial X$, i.e. $h(\bar{x}) = 0$ and for any $g_h \in \partial h(\bar{x})$ the inequality $g_h \bar{x} \geq g_h y$ holds for any $y \in X$.
By taking $y = x^c$ obtain $g_h \bar{x} \geq g_h x^c$ and therefore

$$g_h \bar{x} = g_h x^c + g_h(x - x^c)\mu_X(x, x^c)^{-1} = \mu_X(x, x^c)^{-1} g_h x + (1 - \mu_X($$
$$\mu_X(x, x^c)^{-1} g_h x + (1 - \mu_X(x, x^c)^{-1}) g_h \bar{x}.$$

Hence $g_h x \geq g_h \bar{x} \geq g_h y$ for any $y \in X$, which means that $g_h \in K_X(x)$.

What about $X + \epsilon B$?

Easy:
It can be approximated from above (included into) by the relaxed inequality $X + \epsilon B \subset \{x : h(x) \leq L\epsilon\}$ where L is a Lipschitz constant in an appropriate neighborhood of X.

Iteration algorithm

After construction of the mapping G_λ, oriented toward solution $x^\star$ of VIP on the whole space E except ϵ-neighborhood of $x^\star$ we can use it in an iterative manner like

$$x^{k+1} = x^k - \theta_k f^k, \; f^k \in G_\lambda(x^k), \; k = 0, 1, \ldots,$$

where $\{\theta_k\}$ is a certain prescribed sequence of step-size multipliers.
The hope is that the sequence of $\{x^k\}, k = 0, 1, \ldots$ will converge to at least the set $X_\epsilon = x^\star + \epsilon B$ of approximate solutions.

Bird's eye view of the algorithm

Taking everything granted and computable execute the major loop of the algorithm:

while *The limit is not reached* **do**

Generate a next approximate solution x_{k+1}:

$$x^{k+1} = \begin{cases} x^k - \theta_k f^k, \ f^k \in G_\lambda(x^k), & \text{if } \|x^k\| \leq 2\rho_G \\ x^0 & \text{otherwise.} \end{cases}$$

Increment iteration counter $k \longrightarrow k+1$;

end

Complete: accept $\{x^k\}, k = 0, 1, \ldots$ as an approximate solution of VIP.

Of course the main remaining problem is to prove that it really works. But it is a different story ...