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

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

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

Splitting Algorithms, Modern Operator Theory, and Applications
(September 17–23, 2017, Oaxaca, Mexica)

September 19, 2017

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

# The Outline

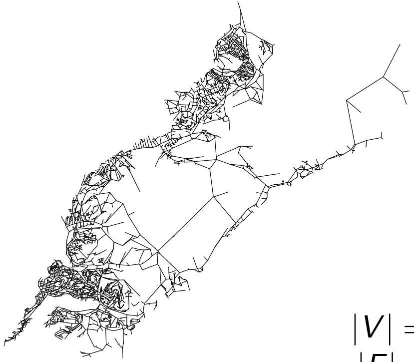
- Motivations, notations and basic preliminaries;
- Superposing feasibility and optimality;
- Oriented and sharp penalty mappings;
- Main reduction result;
- Implementation issues.

# Motivations: transportation studies

**Main problem:** Forcasting the network load in $(V, E)$ transportation network.
**Mainstream model:** noncooperative equilibrium. Dates back to 19 century.
**Equilibrium:** such network load pattern, that nobody gains from infinitesimal changes in its transportation plans.
**Mathematics:** $T(x^\star)(x - x^\star) \geq 0$ for any $x \in X$, where $T_p(x)$ is per-auto delay on the route $p$, $X$ is suply-demand balancing set.
**Specifics:**

- High dimensionality — exponential in $n = |V|$
- Computationaly intensive — requires roughly $n^4$ operation per one gradient computation;
- Strong nonlinearity — Delays/Density dependence is commonly approximated as $\tau(\rho) \sim \rho^k$, with $k \sim 4 - 7$.

# Vladivostok-2009 $(V, G)$



$|V| = 4290,$
$|E| = 5172$

# 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 $z \in X$

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

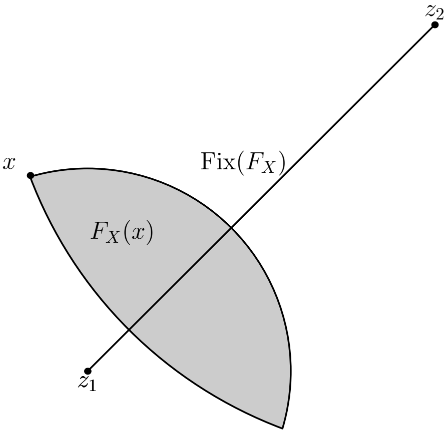
Let $\mathrm{Fix}(F_X)$ be a set of fixed points of operator $F_X$.

**Theorem (Féjer, 1922)**

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

- 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 Féjer operator $F_X$, $X = \{z_1, z_2\}$



# Locally strong Féjer operator

Féjer process (FP):

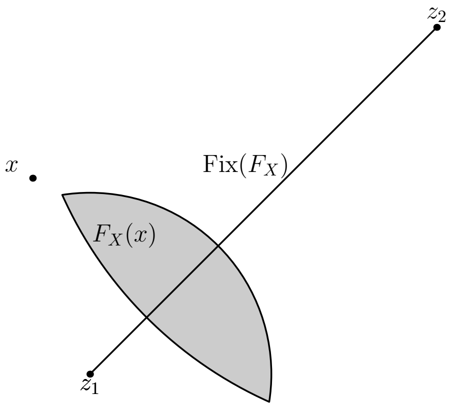
$$x^{k+1} = F_X(x^k), k = 1, 2, dots$$

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



# Convex feasibility

Define distance function $\text{dist}(x, X) = \min_{z \in X} \|z - x\|$.

**Theorem**

*Let the sequence $\{x^k, k = 1, 2, \dots\}$ is generated by the recurrent 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 $k \to \infty$.*

Follows from the theorem 2.16 in H.Bauschke, J.Borwein, SIAM Review, 38(3), 1996.

# Féjer processes with perturbations

FP with perturbations (PFP):

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

where $z^k \to 0$ is an *arbitrary* diminishing perturbations.
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 properties of $\{z^k, k = 0, 1, \dots\}$ one can make the sequence $\{x^k, k = 0, 1, \dots\}$ to converge to specific parts of $X$.
Of course, we are mostly interested in solutions of optimization problems and variational inequalities on $X$.

# Use of perturbations: general idea

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

Split SFP into 2 problems:

1. General Feasibility: $x^\star \in X$
   solved by $x^{k+1} = F_X(x^k), \; k = 1, 2, \ldots$
2. Selective Feasibility: $x^\star \in X_\star$
   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,

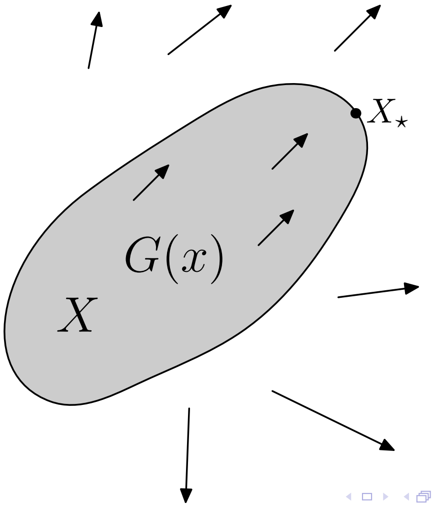
$$g(z - x) \geq \delta > 0$$

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

Examples:

- subdifferentials of convex functions, $X_\star = \mathrm{Argmin}\, f(x), x \in X$;
- strongly monotone operators of variational inequalities.

# Attractant mapping



# VIP superposing — general idea

Variational inequality problem

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

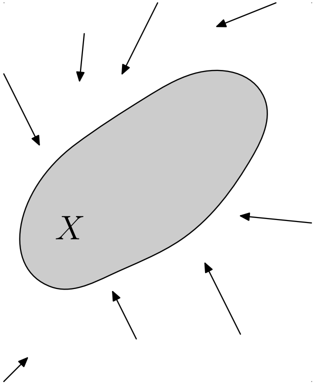
superposed as 2 problems:

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

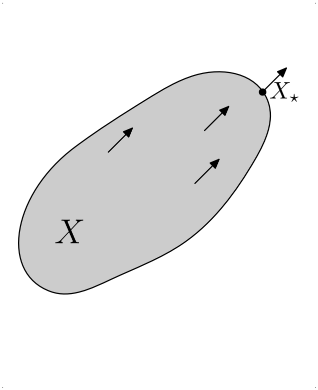
Resulting algorithms:

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

# VIP split view




Feasibility mapping




Optimality mapping

# VIP convergence result

Variational inequality problem

$$G(x^\star)(x - x^\star) \geq 0, \quad x \in X, \; X_\star - \text{solution set.}$$

PFP process:

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

**Theorem**

*Let $F_X$ — locally strong Féjer operator, $H_G$ — 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 $\mathrm{dist}(x^k, X_\star) \to 0$ when $\to \infty$.*

# Shortcomings

- stepsizeis do not adapt themself 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 $k \to \infty$, similar to penalty functions methods.

What can be done ?

- use sofisticated techniques for stepsize control ( quite computationaly expensive );
- apply smoothing techniques;
- look for approximate solutions;
- something else.

# VIP via Optimization

For inspiration we looked at the reduction of VIP to OP:

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

There is a number of 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 via Optimization, cntd

Problems:

- Most merit functions are implicitely defined and therefore are not, strictly speaking, computable;
- Merit functions do not inherit much of the structure of the original problem;
- Did I miss something ?

So why not try something else ?

# Variational and Pseudo-Variational Inequalities

Find $x^\star \in X$ such that: $\begin{array}{ll} G(x^\star)(x - x^\star) \geq 0 & \text{(VIP)} \\ G(x)(x - x^\star) \geq 0 & \text{(PIP)} \end{array}$ for all $x \in X$.

**Important**

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

It is assumed further on that:

- $G(x)$ is monotone,
- VIP and PIP have unique ( and therefore coinsiding) solutions

# Fixed points and PIP

Given PIP: $0 \leq G(x)(x - x^\star), \ \forall x \in X$, construct $\Phi_{G,X}(\cdot)$ or $\Phi_{G,X,\epsilon}(\cdot)$ such that either

**exact solution**

$x^k \to x^\star$ with $x^{k+1} = \Phi_{G,X}(x^k)$

or

**approximate solution**

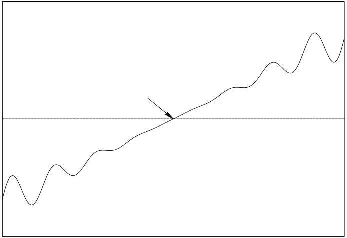
$x^k \to x^\star + \epsilon B$ with $x^{k+1} = \Phi_{G,X,\epsilon}(x^k)$

obtained, when $k \to \infty$. Notice that neither $\Phi_{G,X}(\cdot)$ nor $\Phi_{G,X,\epsilon}(\cdot)$ depend on "time" $k$.

# Oriented mappings

## Definition

$G : X \to \mathcal{C}(E)$ is called a mapping oriented toward $x^\star$ if $g(x - x^\star) \geq 0$ for all $x \in X$ and all $g \in G(x)$.



Simple example: oriened but not monotone.

# Strongly oriented mappings

## Definition

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

$$g(x - x^\star) \geq \gamma_\epsilon$$

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

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

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

# 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 $\bar{x} \in X$

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

.

We will call $\rho_G$ the radius of long-range orientation of $G$ toward $X$.

# Composition of oriented and feasibility mappings

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

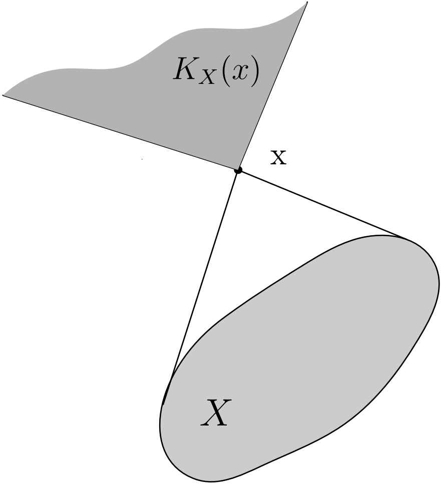
$$G(x, \epsilon) = \epsilon G(x) + P_X(x).$$

Under rather common conditions

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

# Feasibility mapping: modified polar

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 $x$.

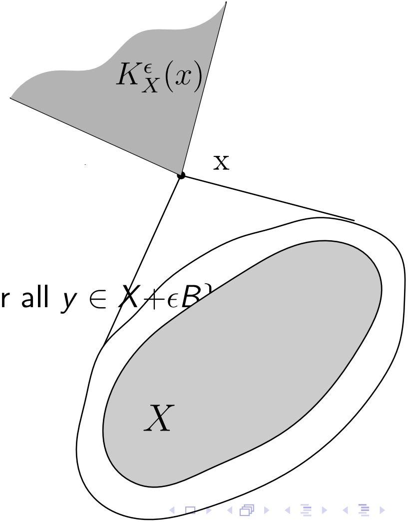


# Feasibility mapping: extended modified polar

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 $\epsilon$-strong polar cone of $X$ at $x$.



# Algorithmic details for 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 $X = \{x : h(x) \leq 0\}$ and $x \notin X$ then $h(x) > 0$ and

$$0 < h(x) - h(y) \leq g_h(x - y)$$

  for any $y \in X$ and any $g_h \in \partial h(x)$, which means that $g_h \in K_X(x)$.
- Combining projection, Minkowski and subdifferentail calculus ...

# Feasibility mapping: globalization and penaly

Define a composite upper semicontinuos mapping for the whole $E$:

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

and define a sharp penalty mapping for $X$ as

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

# A key result for iteration methods

Consider PIP

$$0 \leq G(x)(x - x^\star), \forall x \in X$$

and assume that the $G$ is localy strong oriented map for its solution set, and that the sharp penalty map $P^\epsilon_X$ is constructed. Then the following holds.

## Pseudo-lemma

*If* [the list of prerequisits] *is satisfied then for any* $\epsilon > 0$ *there exists* $\lambda_\epsilon > 0$ *such that the penalized PIP-operator* $G_\lambda(\cdot) = G(\cdot) + \lambda P^\epsilon_X(\cdot)$ *is a localy strong attractor of* $x^\star + \epsilon B$ *for any* $\lambda > \lambda_\epsilon$.

# The list of prerequisits

Assume that;

1. $X \subset E$ is closed and bounded,
2. $G$ is monotone, long-range oriented toward $X$ with the radius of orientability $\rho_G$,
3. $G$ is strongly oriented toward solution $x^\star$ of PIP on $X$ with the constants $\gamma_\epsilon > 0$ for $\epsilon > 0$,
4. $P^\epsilon_X(\cdot)$ is a sharp penalty as defined early.

# 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 $\lambda_\epsilon$ which guarantees

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

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

# Iteration algorithm

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

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

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