Active Set Expansion Strategies in MPRGP

Jakub Kruzik[1,2]
jakub.kruzik@ugn.cas.cz

15 June 2021
Seminar aplikovane matematiky, KAM

[1]Institute of Geonics of the Czech Academy of Sciences
[2]Dept. of Appl. Math., Faculty of Electr. Engin. and CS, VSB - TU Ostrava

Problem definition

$$\underset{\boldsymbol{x}}{\operatorname{argmin}} f(\boldsymbol{x}) \quad \text{s.t.} \quad x \in \Omega$$

$$\underset{\boldsymbol{x}}{\operatorname{argmin}} \frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{b} \quad \text{s.t.} \quad \boldsymbol{l} \leq \boldsymbol{x} \leq \boldsymbol{u},$$

where $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is SPS.

Active/Free Set and Gradient Splitting

Active/Free set:

$$\mathcal{A}(\boldsymbol{x}) = \{j : x_j = l_j \vee x_j = u_j\} \qquad \mathcal{F}(\boldsymbol{x}) = \{j : l_j < x_j < u_j\}$$

Active/Free Set and Gradient Splitting

Active/Free set:

$$\mathcal{A}(\boldsymbol{x}) = \{j : x_j = l_j \vee x_j = u_j\} \qquad \mathcal{F}(\boldsymbol{x}) = \{j : l_j < x_j < u_j\}$$

Gradient splitting:

$$\boldsymbol{g} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$$

$$g_j^f = \begin{cases} 0 & \text{if} \quad j \in \mathcal{A}, \\ g_j & \text{if} \quad j \in \mathcal{F}. \end{cases} \qquad g_j^c = \begin{cases} 0 & \text{if} \quad j \in \mathcal{F}, \\ \min(g_j, 0) & \text{if} \quad x_j = l_j, \\ \max(g_j, 0) & \text{if} \quad x_j = u_j. \end{cases}$$

Active/Free Set and Gradient Splitting

Active/Free set:

$$\mathcal{A}(\boldsymbol{x}) = \{j : x_j = l_j \vee x_j = u_j\} \qquad \mathcal{F}(\boldsymbol{x}) = \{j : l_j < x_j < u_j\}$$

Gradient splitting:

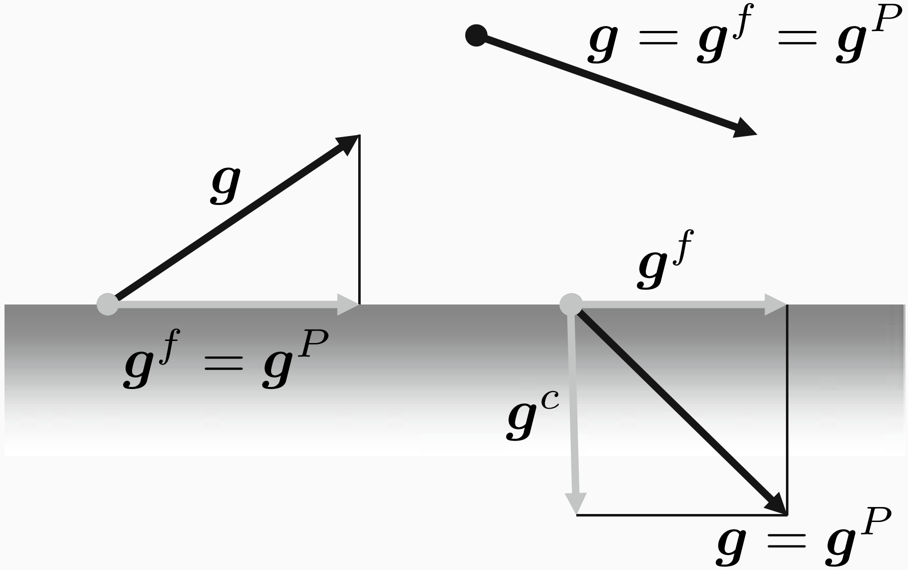
$$\boldsymbol{g} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$$

$$g_j^f = \begin{cases} 0 & \text{if} \quad j \in \mathcal{A}, \\ g_j & \text{if} \quad j \in \mathcal{F}. \end{cases} \qquad g_j^c = \begin{cases} 0 & \text{if} \quad j \in \mathcal{F}, \\ \min(g_j, 0) & \text{if} \quad x_j = l_j, \\ \max(g_j, 0) & \text{if} \quad x_j = u_j. \end{cases}$$

Projected gradient:

$$\boldsymbol{g}^P = \boldsymbol{g}^f + \boldsymbol{g}^c$$

Gradient Splitting - Example



Keeping x Feasible

Projection onto the feasible set Ω:

$$[P_\Omega(\boldsymbol{x})]_j = \min(u_j, \max(l_j, x_j)).$$

Reduce free gradient:

$$g_j^r = \begin{cases} 0 & \text{if} \quad j \in \mathcal{A}, \\ \min\left(\frac{x_j - l_j}{\overline{\alpha}}, g_j\right) & \text{if} \quad j \in \mathcal{F} \quad \text{and} \quad g_j > 0, \\ \max\left(\frac{x_j - u_j}{\overline{\alpha}}, g_j\right) & \text{if} \quad j \in \mathcal{F} \quad \text{and} \quad g_j \leq 0, \end{cases}$$

Therefore:

$$P_\Omega(\boldsymbol{x} - \overline{\alpha}\boldsymbol{g}^f) = \boldsymbol{x} - \overline{\alpha}\boldsymbol{g}^r$$

MPRGP - Modified Proportioning with Reduced Gradient Pojections

Input: $\boldsymbol{A}$, $\boldsymbol{x}_0 \in \Omega$, $\boldsymbol{b}$, $\Gamma > 0$, $\overline{\alpha} \in (0, 2||\boldsymbol{A}||^{-1}]$

$\boldsymbol{g}_0 = \boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{b}$, $\boldsymbol{p}_0 = \boldsymbol{g}_0^f$, $k = 0$

while $||\boldsymbol{g}_k^P||$ *is not small*:

    if $||\boldsymbol{g}_k^c|| \leq \Gamma||\boldsymbol{g}_k^f||$:

        $\alpha_k^{feas} = \max\{\alpha : \boldsymbol{x}_k - \alpha\boldsymbol{p}_k \in \Omega\}$

        $\alpha_k^{cg} = \boldsymbol{g}_k^T\boldsymbol{p}_k / \boldsymbol{p}_k^T\boldsymbol{A}\boldsymbol{p}_k$

        if $\alpha_k^{cg} \leq \alpha_k^{feas}$:

            CG step

        else:

            Expansion step

    else:

        Proportioning step

    $k = k + 1$

Output: $\boldsymbol{x}_k$

MPRGP - Modified Proportioning with Reduced Gradient Pojections

Input: $\boldsymbol{A}$, $\boldsymbol{x}_0 \in \Omega$, $\boldsymbol{b}$, $\Gamma > 0$, $\overline{\alpha} \in (0, 2||\boldsymbol{A}||^{-1}]$

$\boldsymbol{g}_0 = \boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{b}$, $\boldsymbol{p}_0 = \boldsymbol{g}_0^f$, $k = 0$

while $||\boldsymbol{g}_k^P||$ *is not small*:

  if $||\boldsymbol{g}_k^c|| \leq \Gamma ||\boldsymbol{g}_k^f||$:

    $\alpha_k^{feas} = \max\{\alpha : \boldsymbol{x}_k - \alpha \boldsymbol{p}_k \in \Omega\}$

    $\alpha_k^{cg} = \boldsymbol{g}_k^T \boldsymbol{p}_k / \boldsymbol{p}_k^T \boldsymbol{A} \boldsymbol{p}_k$

    if $\alpha_k^{cg} \leq \alpha_k^{feas}$:

      CG step

    else:

      Expansion step

  else:

    Proportioning step

  $k = k + 1$

Output: $\boldsymbol{x}_k$

CG step:

1. $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k^{cg} \boldsymbol{p}_k$
2. $\boldsymbol{g}_{k+1} = \boldsymbol{g}_k - \alpha_k^{cg} \boldsymbol{A} \boldsymbol{p}_k$
3. $\beta_k = \boldsymbol{p}_k^T \boldsymbol{A} \boldsymbol{g}_{k+1}^f / \boldsymbol{p}_k^T \boldsymbol{A} \boldsymbol{p}_k$
4. $\boldsymbol{p}_{k+1} = \boldsymbol{g}_{k+1}^f - \beta_k \boldsymbol{p}_k$

MPRGP - Modified Proportioning with Reduced Gradient Pojections

Input: $\boldsymbol{A}$, $\boldsymbol{x}_0 \in \Omega$, $\boldsymbol{b}$, $\Gamma > 0, \overline{\alpha} \in (0, 2||\boldsymbol{A}||^{-1}]$

$\boldsymbol{g}_0 = \boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{b}$, $\boldsymbol{p}_0 = \boldsymbol{g}_0^f$, $k = 0$

while $||\boldsymbol{g}_k^P||$ *is not small*:

  if $||\boldsymbol{g}_k^c|| \leq \Gamma ||\boldsymbol{g}_k^f||$:

    $\alpha_k^{feas} = \max\{\alpha : \boldsymbol{x}_k - \alpha \boldsymbol{p}_k \in \Omega\}$

    $\alpha_k^{cg} = \boldsymbol{g}_k^T \boldsymbol{p}_k / \boldsymbol{p}_k^T \boldsymbol{A} \boldsymbol{p}_k$

    if $\alpha_k^{cg} \leq \alpha_k^{feas}$:

      CG step

    else:

      Expansion step

  else:

    Proportioning step

  $k = k + 1$

Output: $\boldsymbol{x}_k$

CG step:

1. $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k^{cg} \boldsymbol{p}_k$
2. $\boldsymbol{g}_{k+1} = \boldsymbol{g}_k - \alpha_k^{cg} \boldsymbol{A} \boldsymbol{p}_k$
3. $\beta_k = \boldsymbol{p}_k^T \boldsymbol{A} \boldsymbol{g}_{k+1}^f / \boldsymbol{p}_k^T \boldsymbol{A} \boldsymbol{p}_k$
4. $\boldsymbol{p}_{k+1} = \boldsymbol{g}_{k+1}^f - \beta_k \boldsymbol{p}_k$

Expansion step:

1. $\boldsymbol{x}_{k+\frac{1}{2}} = \boldsymbol{x}_k - \alpha_k^{feas} \boldsymbol{p}_k$
2. $\boldsymbol{g}_{k+\frac{1}{2}} = \boldsymbol{g}_k - \alpha_k^{feas} \boldsymbol{A} \boldsymbol{p}_k$
3. $\boldsymbol{x}_{k+1} = P_\Omega(\boldsymbol{x}_{k+\frac{1}{2}} - \overline{\alpha} \boldsymbol{g}_{k+\frac{1}{2}}^f)$
4. $\boldsymbol{g}_{k+1} = \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{b}$
5. $\boldsymbol{p}_{k+1} = \boldsymbol{g}_{k+1}^f$

MPRGP - Modified Proportioning with Reduced Gradient Pojections

Input: $\boldsymbol{A}$, $\boldsymbol{x}_0 \in \Omega$, $\boldsymbol{b}$, $\Gamma > 0$, $\overline{\alpha} \in (0, 2||\boldsymbol{A}||^{-1}]$

```
g_0 = A x_0 - b, p_0 = g_0^f, k = 0
while ||g_k^P|| is not small:
    if ||g_k^c|| <= Γ||g_k^f||:
        α_k^feas = max{α : x_k - α p_k ∈ Ω}
        α_k^cg = g_k^T p_k / p_k^T A p_k
        if α_k^cg <= α_k^feas:
            CG step
        else:
            Expansion step
    else:
        Proportioning step
    k = k + 1
```

$g_0 = Ax_0 - b$, $p_0 = g_0^f$, $k = 0$

while $||\boldsymbol{g}_k^P||$ *is not small*:

- if $||\boldsymbol{g}_k^c|| \leq \Gamma||\boldsymbol{g}_k^f||$:
  - $\alpha_k^{feas} = \max\{\alpha : \boldsymbol{x}_k - \alpha \boldsymbol{p}_k \in \Omega\}$
  - $\alpha_k^{cg} = \boldsymbol{g}_k^T \boldsymbol{p}_k / \boldsymbol{p}_k^T \boldsymbol{A} \boldsymbol{p}_k$
  - if $\alpha_k^{cg} \leq \alpha_k^{feas}$:
    - CG step
  - else:
    - Expansion step
- else:
  - Proportioning step
- $k = k + 1$

Output: $\boldsymbol{x}_k$

Proportioning step:

$\alpha_k = \boldsymbol{g}_k^T \boldsymbol{g}_k^c / (\boldsymbol{g}_k^c)^T \boldsymbol{A} \boldsymbol{g}_k^c$

$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k^c$

$\boldsymbol{g}_{k+1} = \boldsymbol{g}_k - \alpha_k \boldsymbol{A} \boldsymbol{g}_k^c$

$\boldsymbol{p}_{k+1} = \boldsymbol{g}_{k+1}^f$

MPRGP Operation Count

Step	**Hess. mult.**	Dot prod.	Vec. update	Grad. split.
CG	**1**	2	3	1
Expansion	**2**	1	5	2
Proportioning	**1**	1	3	1

MPRGP R-linear Convergence

[Z. Dostál, J. Schöberl, 2005]

Let $\boldsymbol{x}_k$ be generate by MRPGP, $\boldsymbol{x}_0 \in \Omega$, $\Gamma > 0$ and $\overline{\alpha} \in \left(0, 2||\boldsymbol{A}||^{-1}\right]$. Then

$$f(\boldsymbol{x}_{k+1}) - f(\widehat{\boldsymbol{x}}) \leq \eta \left(f(\boldsymbol{x}_k) - f(\widehat{\boldsymbol{x}})\right),$$

where $\widehat{\boldsymbol{x}}$ denotes the unique solution,

$$\eta = 1 - \frac{\widehat{\alpha}\lambda_{min}}{\vartheta\left(1 + \widehat{\Gamma}^2\right)},$$

$$\widehat{\Gamma} = \max\{\Gamma, \Gamma^{-1}\}, \qquad \vartheta = 2\max\{\overline{\alpha}||\boldsymbol{A}||, 1\}, \qquad \widehat{\alpha} = \min\{\overline{\alpha}, 2||\boldsymbol{A}||^{-1} - \overline{\alpha}\}$$

$$\eta^{opt} = 1 - \kappa\left(\boldsymbol{A}\right)^{-1}/4$$

for $\Gamma = 1$ and $\overline{\alpha} = ||\boldsymbol{A}||^{-1}$

Adaptive Step-length I.

$$
\begin{aligned}
f(\boldsymbol{x}) - f(\boldsymbol{x} - \bar{\alpha}\boldsymbol{d}) &= f(\boldsymbol{x}) - \frac{1}{2}(\boldsymbol{x} - \bar{\alpha}\boldsymbol{d})^T\boldsymbol{A}(\boldsymbol{x} - \bar{\alpha}\boldsymbol{d}) + (\boldsymbol{x} - \bar{\alpha}\boldsymbol{d})^T\boldsymbol{b} = \\
&= f(\boldsymbol{x}) - \frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{b} - \frac{1}{2}\bar{\alpha}^2\boldsymbol{d}^T\boldsymbol{A}\boldsymbol{d} + \bar{\alpha}\boldsymbol{d}^T\boldsymbol{A}\boldsymbol{x} - \bar{\alpha}\boldsymbol{d}^T\boldsymbol{b} = \\
&= -\frac{1}{2}\bar{\alpha}^2\boldsymbol{d}^T\boldsymbol{A}\boldsymbol{d} + \bar{\alpha}\boldsymbol{d}^T\boldsymbol{g} \geq 0,
\end{aligned}
$$

divide by $\bar{\alpha} > 0$

$$
\frac{1}{2}\bar{\alpha}\boldsymbol{d}^T\boldsymbol{A}\boldsymbol{d} \leq \boldsymbol{d}^T\boldsymbol{g}
$$

assuming $\boldsymbol{d}^T\boldsymbol{A}\boldsymbol{d} > 0$

$$
\bar{\alpha} \leq \frac{2\boldsymbol{d}^T\boldsymbol{g}}{\boldsymbol{d}^T\boldsymbol{A}\boldsymbol{d}}
$$

Adaptive Step-length II.

$$0 < \bar{\alpha} \leq 2 \left\| \boldsymbol{A} \right\|^{-1} \leq 2 \left\| \boldsymbol{A} \right\|^{-1} \frac{\boldsymbol{d}^T \boldsymbol{g}}{\boldsymbol{d}^T \boldsymbol{d}} \leq \frac{2 \boldsymbol{d}^T \boldsymbol{g}}{\boldsymbol{d}^T \boldsymbol{A} \boldsymbol{d}}.$$

- *fixed* $\bar{\alpha} = \alpha_u \left\| \boldsymbol{A} \right\|^{-1}$,
- *opt* $\bar{\alpha} = \alpha_u \frac{\boldsymbol{d}^T \boldsymbol{g}}{\boldsymbol{d}^T \boldsymbol{A} \boldsymbol{d}}$,
- *optapprox* $\bar{\alpha} = \alpha_u \left\| \boldsymbol{A} \right\|^{-1} \frac{\boldsymbol{d}^T \boldsymbol{g}}{\boldsymbol{d}^T \boldsymbol{d}}$,

where $\alpha_u \in (0, 2]$.

Adaptive Step-length II.

$$0 < \bar{\alpha} \leq 2 \left\| \boldsymbol{A} \right\|^{-1} \leq 2 \left\| \boldsymbol{A} \right\|^{-1} \frac{\boldsymbol{d}^T \boldsymbol{g}}{\boldsymbol{d}^T \boldsymbol{d}} \leq \frac{2 \boldsymbol{d}^T \boldsymbol{g}}{\boldsymbol{d}^T \boldsymbol{A} \boldsymbol{d}}.$$

- *fixed* $\bar{\alpha} = \alpha_u \left\| \boldsymbol{A} \right\|^{-1}$,
- *opt* $\bar{\alpha} = \alpha_u \frac{\boldsymbol{d}^T \boldsymbol{g}}{\boldsymbol{d}^T \boldsymbol{A} \boldsymbol{d}}$,
- *optapprox* $\bar{\alpha} = \alpha_u \left\| \boldsymbol{A} \right\|^{-1} \frac{\boldsymbol{d}^T \boldsymbol{g}}{\boldsymbol{d}^T \boldsymbol{d}}$,

where $\alpha_u \in (0, 2]$.

The step needs to be projeted

$$\boldsymbol{x}_{k+1} = P_\Omega(\boldsymbol{x}_{k+\frac{1}{2}} - \overline{\alpha} \widetilde{\boldsymbol{d}}).$$

Decoupled descent direction $\widetilde{\boldsymbol{d}} = \boldsymbol{g}^f$ or $\widetilde{\boldsymbol{d}} = \boldsymbol{g}^r$ and vector for computing the step-length $\boldsymbol{d} = \boldsymbol{g}^f$ or $\boldsymbol{d} = \boldsymbol{g}^r$.

MPRGP with Projected CG Step Expansion

Input: $\boldsymbol{A}$, $\boldsymbol{x}_0 \in \Omega$, $\boldsymbol{b}$, $\Gamma > 0$, $\overline{\alpha} \in (0, 2||\boldsymbol{A}||^{-1}]$

$\boldsymbol{g}_0 = \boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{b}$, $\boldsymbol{p}_0 = \boldsymbol{g}_0^f$, $k = 0$

while $||\boldsymbol{g}_k^P||$ *is not small*:

    if $||\boldsymbol{g}_k^c|| \leq \Gamma||\boldsymbol{g}_k^f||$:

        $\alpha_k^{feas} = \max\{\alpha : \boldsymbol{x}_k - \alpha\boldsymbol{p}_k \in \Omega\}$

        $\alpha_k^{cg} = \boldsymbol{g}_k^T\boldsymbol{p}_k / \boldsymbol{p}_k^T\boldsymbol{A}\boldsymbol{p}_k$

        $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k^{cg}\boldsymbol{p}_k$

        Projected CG

    else:

        Proportioning step

    $k = k + 1$

Output: $\boldsymbol{x}_k$

Projected CG:

if $\alpha_k^{cg} \leq \alpha^{feas}$:

    $\boldsymbol{g}_{k+1} = \boldsymbol{g}_k - \alpha_k^{cg}\boldsymbol{A}\boldsymbol{p}_k$

    $\beta_k = \boldsymbol{p}_k^T\boldsymbol{A}\boldsymbol{g}_{k+1}^f / \boldsymbol{p}_k^T\boldsymbol{A}\boldsymbol{p}_k$

    $\boldsymbol{p}_{k+1} = \boldsymbol{g}_{k+1}^f - \beta_k\boldsymbol{p}_k$

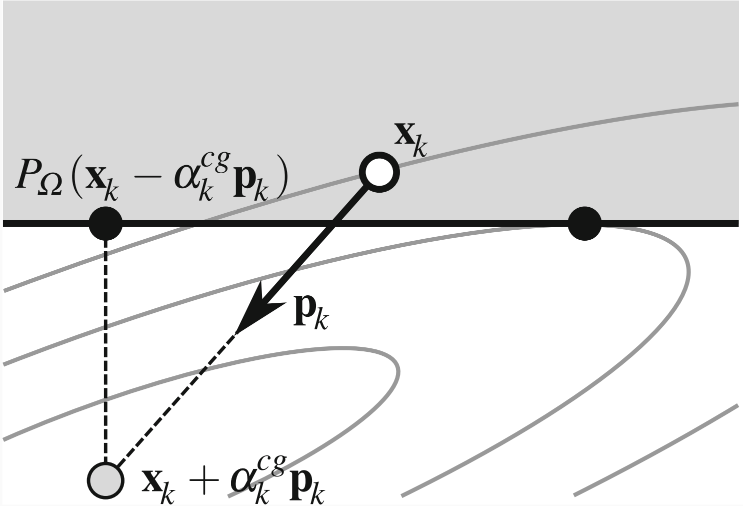
else:

    $\boldsymbol{x}_{k+1} = P_\Omega(\boldsymbol{x}_{k+1})$

    $\boldsymbol{g}_{k+1} = \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{b}$

    $\boldsymbol{p}_{k+1} = \boldsymbol{g}_{k+1}^f$

Projected CG and Convergence



$$f(P_\Omega(\boldsymbol{x}_k - \alpha_k^{cg}\boldsymbol{p}_k)) > f(\boldsymbol{x}_k)$$

[Z. Dostál, 2009]

MPRGP Operation Count

Step	Hess. mult.	Dot prod.	Vec. update	Grad. split.
CG	1	2	3	1
Expansion	2	1	5	2
Proportioning	1	1	3	1
Expansion-optapprox	2	3	5	2
Expansion-opt	3	3	5	2
Expansion-projCG	2	1	3	1

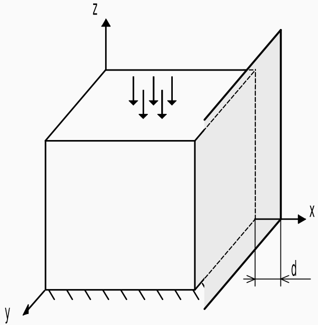
Numerical Experiments

Benchmarks

- $\Gamma = 1$
- $||\boldsymbol{A}||$ estimated by power method
  - 5 – 50 iterations (10^{-4} relative error)
  - not included in results

Benchmarks

- $\Gamma = 1$
- $||\boldsymbol{A}||$ estimated by power method
  - $5 - 50$ iterations (10^{-4} relative error)
  - not included in results



3D Linear Elasticity Contact Problem

- FETI dual formulation:

$$\underset{\boldsymbol{\lambda}}{\operatorname{argmin}} \frac{1}{2}\boldsymbol{\lambda}^T\boldsymbol{F}\boldsymbol{\lambda} - \boldsymbol{\lambda}^T\boldsymbol{d} \text{ s.t. } \boldsymbol{\lambda}_I \geq \boldsymbol{o} \text{ and } \boldsymbol{G}\boldsymbol{\lambda} = \boldsymbol{e}$$

- SMALBE outer solver
- $81,812,703$ (undecomposed) degrees of freedom over $1,000$ subdomains
- relative tolerance 10^{-6}

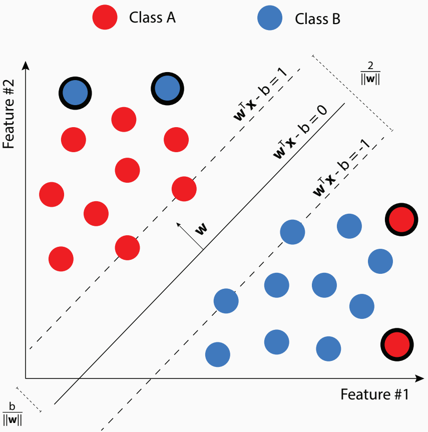
Benchmarks

Support Vector Machines (SVMs) Classification

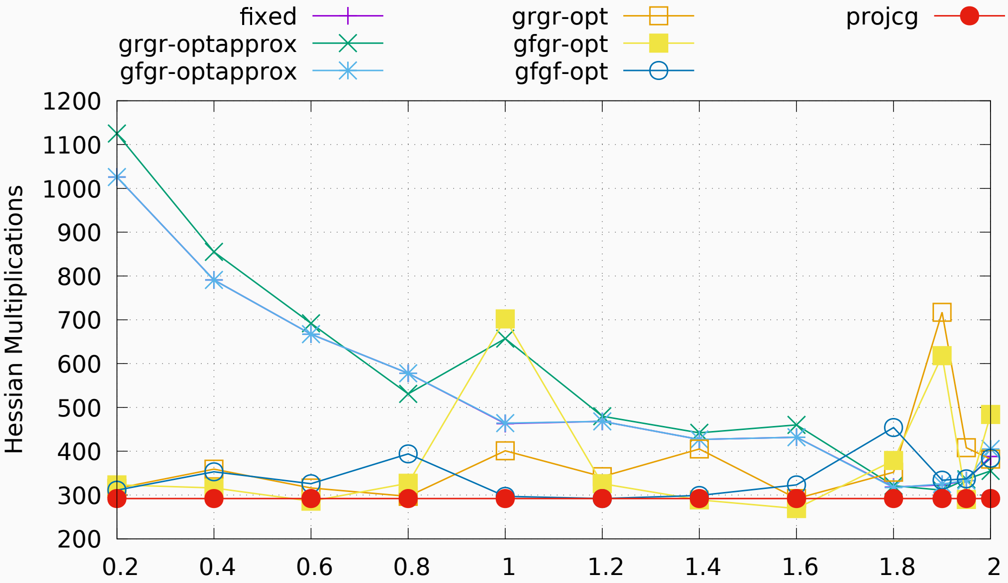
$$\underset{\boldsymbol{\lambda}}{\text{argmin}} \ \frac{1}{2}\boldsymbol{\lambda}^T \boldsymbol{Y}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{Y}\boldsymbol{\lambda} - \boldsymbol{\lambda}^T\boldsymbol{e} \quad \text{s.t.} \quad \boldsymbol{o} \leq \boldsymbol{\lambda} \leq C\boldsymbol{e}$$

- $l1$-loss function, dual formulation
- LIBSVM datasets
- relative tolerance 10^{-1}

Dataset	# samples	# features
Australian	690	14
Diabetes	678	8
Ionosphere	351	34



Benchmark: Linear Elasticity Contact Problem, I.



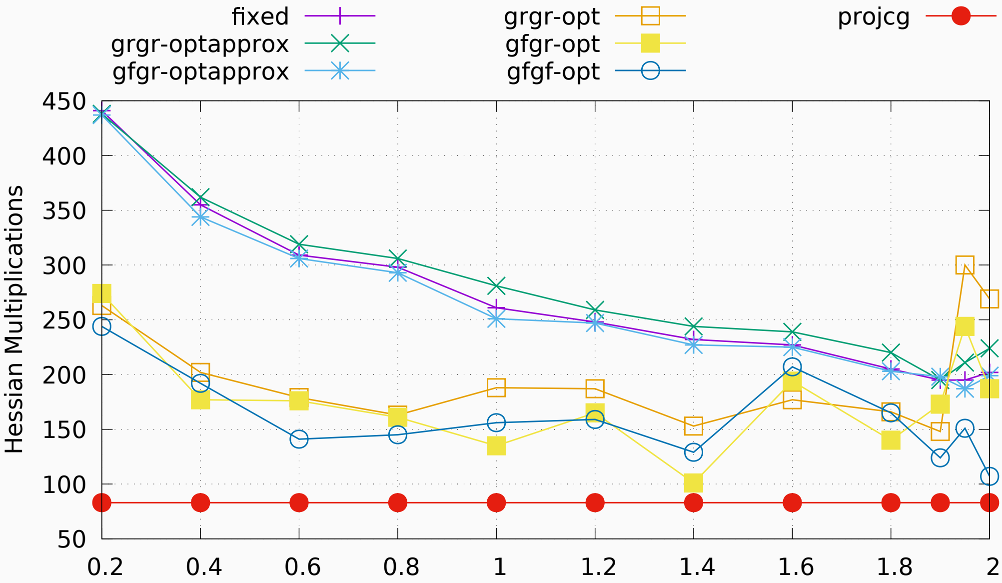
Benchmark: Linear Elasticity Contact Problem, II.

exp. type	α_u	outer it.	#Hess. mult.	#CG	#Exp.	#Prop.
fixed	1.8	10	318	158	74	2
	1.9	10	323	144	83	3
	1.95	7	333	70	127	2
grgr-optapprox	1.9	10	311	162	68	3
	1.8	10	321	145	82	2
	1.95	10	336	125	99	3
gfgr-optapprox	1.8	10	317	159	73	2
	1.9	10	325	140	86	3
	1.95	8	333	95	114	3

Benchmark: Linear Elasticity Contact Problem, III.

exp. type	α_u	outer it.	#Hess. mult.	#CG	#Exp.	#Prop.
fixed	1.8	10	318	158	74	2
	1.9	10	323	144	83	3
	1.95	7	333	70	127	2
grgr-opt	1.6	8	292	120	54	2
	0.8	9	297	132	51	3
	0.2	8	315	113	64	2
gfgr-opt	1.6	9	269	129	43	2
	0.6	12	286	139	44	3
	1.4	9	289	154	41	3
gfgf-opt	1.2	10	292	151	43	2
	1.0	11	297	171	37	4
	1.4	9	299	152	45	3
projcg	-	10	292	171	53	5

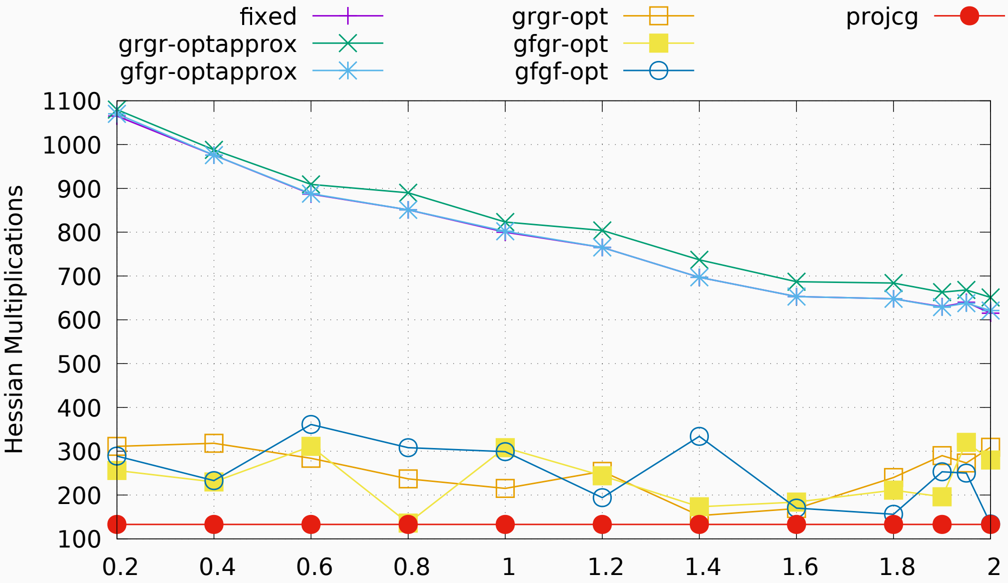
Benchmark: SVM Classification, Australian Dataset, I.



Benchmark: SVM Classification, Australian Dataset, II.

exp. type	α_u	#Hess. mult.	#CG	#Exp.	#Prop.
fixed	1.9	195	8	92	2
	1.95	195	8	92	2
	2.0	202	10	95	1
grgr-opt	1.9	148	28	39	2
	1.4	153	27	41	2
	0.8	163	16	48	2
gfgr-opt	1.4	101	20	26	2
	1.0	135	15	39	2
	1.8	140	25	37	3
gfgf-opt	2.0	107	23	27	2
	1.9	124	28	31	2
	1.4	129	28	33	1
projcg	-	83	16	32	2

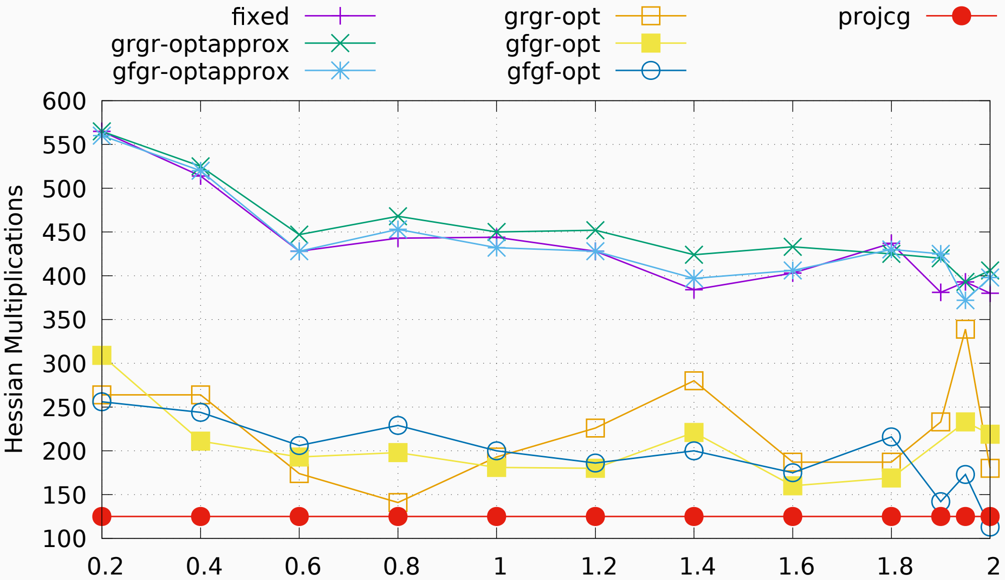
Benchmark: SVM Classification, Diabetes Dataset, I.



Benchmark: SVM Classification, Diabetes Dataset, II.

exp. type	α_u	#Hess. mult.	#CG	#Exp.	#Prop.
fixed	2.0	615	1	306	1
	1.9	630	2	313	1
	1.95	640	2	318	1
grgr-opt	1.4	153	22	43	1
	1.6	169	26	47	1
	1.0	215	20	64	2
gfgr-opt	0.8	136	9	41	3
	1.4	173	25	48	3
	1.6	184	25	52	2
gfgf-opt	2.0	134	11	38	8
	1.8	156	18	44	5
	1.6	170	27	47	1
projcg	-	133	13	58	3

Benchmark: SVM Classification, Ionosphere Dataset, II.

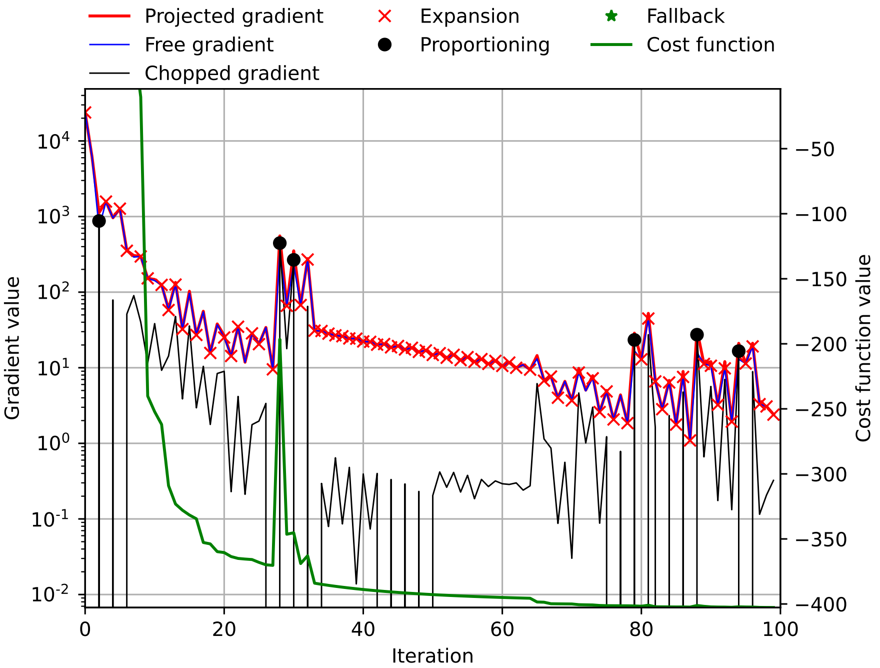


Benchmark: SVM Classification, Ionosphere Dataset, II.

exp. type	α_u	#Hess. mult.	#CG	#Exp.	#Prop.
fixed	2	380	14	182	1
	1.9	381	21	179	1
	1.4	384	26	178	1
grgr-opt	0.8	141	8	43	3
	0.6	174	11	53	3
	2	180	23	50	6
gfgr-opt	1.6	160	30	42	3
	1.8	169	30	44	6
	1.2	180	18	52	5
gfgf-opt	2	113	22	29	3
	1.9	142	26	37	4
	1.95	173	28	46	6
projcg	-	125	14	54	2

SVM - tolerance $1e-4$

Dataset	Exp. Type	Hess. mult.	CG	Exp,.	Prop.	Cost inc.	Fall.
australian	fixed	4567	1134	1704	24	423	0
australian	projCG	3571	565	1455	95	127	0
diabetes	fixed	1108	124	491	1	18	0
diabetes	projCG	**1439**	113	627	71	109	0
ionosphere	fixed	628	149	237	4	8	0
ionosphere	projCG	265	104	78	4	6	0



Projected gradient
Free gradient
Chopped gradient
Expansion
Proportioning
Fallback
Cost function
Gradient value
Cost function value
Iteration
10^4
10^3
10^2
10^1
10^0
10^-1
10^-2
−50
−100
−150
−200
−250
−300
−350
−400
0
20
40
60
80
100

MPRGP with Projected CG and Fallback

Input: $\boldsymbol{A}$, $\boldsymbol{x}_0 \in \Omega$, $\boldsymbol{b}$, $\Gamma > 0$, $\overline{\alpha} \in (0, 2||\boldsymbol{A}||^{-1}]$

$\boldsymbol{g}_0 = \boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{b}$, $\boldsymbol{p}_0 = \boldsymbol{g}_0^f$, $k = 0$

while $||\boldsymbol{g}_k^P||$ *is not small*:

  if $||\boldsymbol{g}_k^c|| \leq \Gamma ||\boldsymbol{g}_k^f||$:

    $\alpha_k^{feas} = \max\{\alpha : \boldsymbol{x}_k - \alpha \boldsymbol{p}_k \in \Omega\}$

    $\alpha_k^{cg} = \boldsymbol{g}_k^T \boldsymbol{p}_k / \boldsymbol{p}_k^T \boldsymbol{A} \boldsymbol{p}_k$

    $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k^{cg} \boldsymbol{p}_k$

    Projected CG with fallback

  else:

    Proportioning step

  $k = k + 1$

Output: $\boldsymbol{x}_k$

---

if $\alpha_k^{cg} \leq \alpha^{feas}$:

  $\boldsymbol{g}_{k+1} = \boldsymbol{g}_k - \alpha_k^{cg} \boldsymbol{A} \boldsymbol{p}_k$

  $\beta_k = \boldsymbol{p}_k^T \boldsymbol{A} \boldsymbol{g}_{k+1}^f / \boldsymbol{p}_k^T \boldsymbol{A} \boldsymbol{p}_k$

  $\boldsymbol{p}_{k+1} = \boldsymbol{g}_{k+1}^f - \beta_k \boldsymbol{p}_k$

else:

  $\boldsymbol{x}_{k+1} = P_\Omega(\boldsymbol{x}_{k+1})$

  $\boldsymbol{g}_{k+1} = \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{b}$

  if *fallback*:

    $\boldsymbol{x}_{k+\frac{1}{2}} = \boldsymbol{x}_k - \alpha_k^{feas} \boldsymbol{p}_k$

    $\boldsymbol{g}_{k+\frac{1}{2}} = \boldsymbol{g}_k - \alpha_k^{feas} \boldsymbol{A} \boldsymbol{p}_k$

    $\boldsymbol{x}_{k+1} = P_\Omega(\boldsymbol{x}_{k+\frac{1}{2}} - \overline{\alpha} \boldsymbol{g}_{k+\frac{1}{2}}^f)$

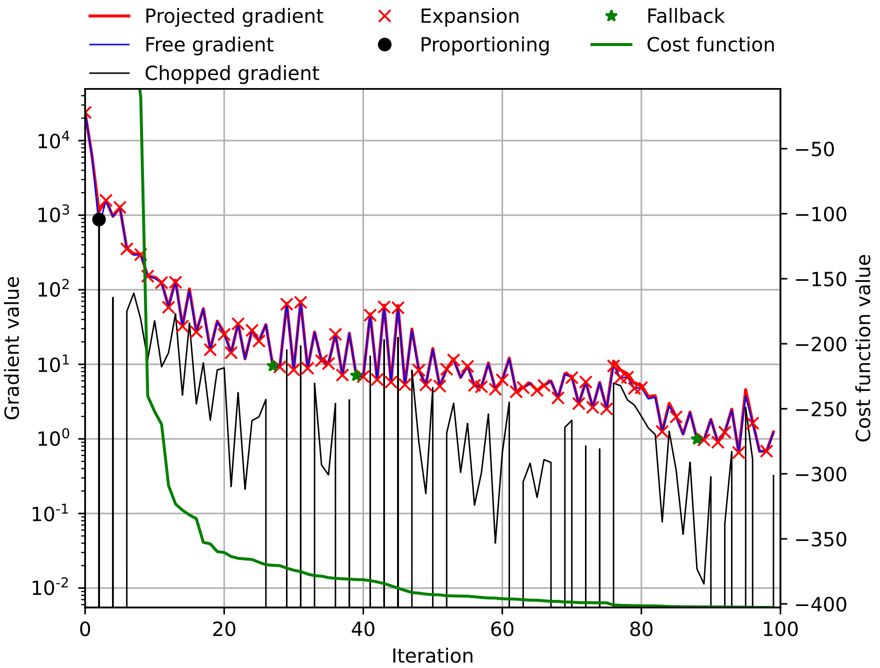
    $\boldsymbol{g}_{k+1} = \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{b}$

  $\boldsymbol{p}_{k+1} = \boldsymbol{g}_{k+1}^f$

Fallback Variants

Fallback 1:

$$f(\boldsymbol{x}_{k+1}) > f(\boldsymbol{x}_k)$$



Projected gradient
Expansion
Fallback
Free gradient
Proportioning
Cost function
Chopped gradient
Gradient value
Cost function value
Iteration
10^4
10^3
10^2
10^1
10^0
10^{-1}
10^{-2}
−50
−100
−150
−200
−250
−300
−350
−400
0
20
40
60
80
100

SVM - tolerance $1e-4$

Dataset	Exp. Type	Hess. mult.	CG	Exp,.	Prop.	Cost inc.	Fall.
australian	fixed	4567	1134	1704	24	423	0
australian	projCG	3571	565	1455	95	127	0
australian	fallback 1	3298	1015	1014	36	218	218
diabetes	fixed	1108	124	491	1	18	0
diabetes	projCG	1439	113	627	71	109	0
diabetes	fallback 1	292	84	96	1	14	14
ionosphere	fixed	628	149	237	4	8	0
ionosphere	projCG	265	104	78	4	6	0
ionosphere	fallback 1	**320**	109	89	5	27	27

Fallback Variants

Fallback 1:

$$f(\boldsymbol{x}_{k+1}) > f(\boldsymbol{x}_k)$$

Fallback 2:

$$f(\boldsymbol{x}_{k+1}) > f(\boldsymbol{x}_k) \quad \wedge \quad ||\boldsymbol{g}_{k+1}^c|| > \Gamma||\boldsymbol{g}_{k+1}^f||$$

SVM - tolerance $1e-4$

Dataset	Exp. Type	Hess. Mult.	CG	Exp.	Prop.	Cost inc.	Fall.
australian	fixed	4567	1134	1704	24	423	0
australian	projCG	3571	565	1455	95	127	0
australian	fallback 1	3298	1015	1014	36	218	218
australian	fallback 2	2160	747	662	32	74	56
diabetes	fixed	1108	124	491	1	18	0
diabetes	projCG	1439	113	627	71	109	0
diabetes	fallback 1	292	84	96	1	14	14
diabetes	fallback 2	292	84	96	1	14	14
ionosphere	fixed	628	149	237	4	8	0
ionosphere	projCG	265	104	78	4	6	0
ionosphere	fallback 1	320	109	89	5	27	27
ionosphere	fallback 2	277	104	78	5	13	11

Conclusion and Outlook

- *fixed* $\alpha_u \approx 1.9$
- *optapprox* is not useful
- *opt* can have a huge performance benefit
  - $\boldsymbol{g}^f$ for both direction and step-length
  - $\alpha_u \approx 1.9$
- Projected CG
  - nearly best performace
  - no additional parameter
  - expansion may be too agressive $\Rightarrow$ may need some fallback

Thank you for your attention!

Any questions?

Jakub Kruzik[1,2]
jakub.kruzik@ugn.cas.cz

http://permon.vsb.cz

[1] Institute of Geonics of the Czech Academy of Sciences
[2] Dept. of Appl. Math., Faculty of Electr. Engin. and CS, VSB - TU Ostrava