Low-Rank Matrix Completion (LRMC) using Nuclear Norm (NN) with Facial Reduction (FR)

Henry Wolkowicz

Dept. Combinatorics and Optimization, University of Waterloo, Canada

Sat. Nov. 5, 2016, 10:30-11:00

4th Alpen-Adria-Workshop on Optimization, Klagenfurt

** Motivation: No Slater CQ/Facial reduction

- Slater condition – existence of a strictly feasible solution – is at the heart of convex optimization.

- Without Slater: first-order optimality conditions may fail; dual problem may yield little information; small perturbations may result in infeasibility; many software packages can behave poorly.

- a pronounced phenomenon: Slater holds generically; but surprisingly many models from relaxations of hard nonconvex problems show loss of strict feasibility: QAP, GP, strengthened Max-Cut, Matrix completions, EDM, sensor network localization, SNL, Molecular Conformation, POP, . . .

- Here: SDP arising from NN for LRMC; Slater's condition holds but: FR on the optimal face

Main Refs:
(i): Huang and W.
(ii) "The many faces of degeneracy in conic optimization", Drusvyatskiy, W. '16 ;

** Facial Reduction/Preprocessing for LP

Primal-Dual Pair: A onto, $m \times n$, $\mathcal{P} = \{1, \ldots, n\}$

$$\text{(LP-P)} \quad \begin{array}{ll} \max & b^\top y \\ \text{s.t.} & A^\top y \leq c \end{array} \qquad \text{(LP-D)} \quad \begin{array}{ll} \min & c^\top x \\ \text{s.t.} & Ax = b, \\ & x \geq 0. \end{array}$$

Slater's CQ for (LP-D) / Theorem of alternative

Exactly One of (I), (II) is True:

(I) $\exists \hat{x}$ s.t. $A\hat{x} = b, \hat{x} > 0$ $\quad (0 < \hat{x} \in \operatorname{ri} F,$ feas. set)

Slater point

(II) $0 \neq z = A^\top y \geq 0, \; b^\top y = 0$ $\quad (\langle z, F \rangle = 0)$

exposing vector

LP Example, Preprocessing

$$\begin{array}{ll} \min & \begin{pmatrix} 2 & 6 & -1 & -2 & 7 \end{pmatrix} x \\ \text{s.t.} & \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & -1 & -1 & 0 & 1 \end{bmatrix} x = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\ & x \in \mathbb{R}^5_+ \end{array}$$

Sum the two constraints (multiply by: $y^T = (1\ 1)$):
get: $2x_1 + x_4 + x_5 = 0 \implies x_1 = x_4 = x_5 = 0$
i.e., equiv. simplified problem/smaller face; substitute:

$$x \in \text{face}(F) = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \left(\mathbb{R}^2_+\right),$$ and one redundant constraint

$\min 6v_1 - v_2$ s.t. $v_1 + v_2 = 1, \ \ v_1, v_2 \geq 0,$

Faces of Convex Sets/Conjugate Faces

Face of C, $\boxed{F \trianglelefteq C}$

- $F \subseteq C$ is a face of C if F contains any line segment in C whose relative interior intersects F.
- A convex cone $F \subseteq K$ is a face of a convex cone K, $F \trianglelefteq K$, if (simplified)

$$\boxed{x, y \in K \text{ and } x + y \in F \implies x, y \in F}$$

Polar (Dual) Cone/Conjugate Face

- polar cone $K^* := \{\phi : \langle \phi, k \rangle \geq 0, \ \forall k \in K\}$
- If $F \trianglelefteq K$, the conjugate face of F is

$$\boxed{F^c := F^\perp \cap K^* \trianglelefteq K^*}$$

Properties of Faces

General case

- A face of a face is a face
- intersection of a face with a face is a face.
- Let $C \subseteq K$, then face(C) denotes the minimal face (intersection of faces) containing C.

$F \trianglelefteq K$ is an exposed face if there exists $\phi \in K^*$ with

$$F = K \cap \phi^\perp$$

F^c is always exposed by $x \in \text{ri}\, F$.

The SDP cone is facially exposed, all its faces are exposed. (In fact like $\mathbb{R}^n_+$: $\mathcal{S}^n_+$ is a proper closed convex cone, self-dual and facially exposed.)

Regularize abstract convex program (full generality)

in memorium: Jonathan Borwein 20 May 1951 - 2 Aug 2016,

jonborwein.org

(ACP) $\inf_x f(x)$ s.t. $g(x) \preceq_K 0, x \in \Omega$

(Borwein-W.'78-79)

$$(\mathrm{ACP}_R) \quad \inf_x f(x) \text{ s.t. } g(x) \preceq_{K^f} 0, x \in \Omega$$

where: K^f is the minimal face

Like LP, it is simple if we use the minimal face K^f.
For linear f, g: get a proper primal-dual pair!!
(dual of dual is primal)

* SDP Case/Replicating Cone/Faces

$X \in \mathcal{S}^n_+$ with spectral decomposition

$$X = [P\ Q] \begin{bmatrix} D_+ & 0 \\ 0 & 0 \end{bmatrix} [P\ Q]^T, \quad D_+ \in \mathbb{S}^r_{++} \quad (\mathrm{rank}\, X = r)$$

Then:

- $\mathrm{face}(X) = P\mathbb{S}^r_+ P^T = (QQ^T)^\perp \cap \mathcal{S}^n_+$
  ($Z = QQ^T$ *exposing vector/matrix* for face.)
- $\mathrm{face}(X)^c = Q\mathbb{S}^{n-r}_+ Q^T = (PP^T)^\perp \cap \mathcal{S}^n_+$
- $\mathrm{Range}(X) = \mathrm{Range}(P), \quad \mathrm{Null}(X) = \mathrm{Range}(Q)$

Range/Nullspace representations

$$\mathit{face}(X) = \{Y \in \mathcal{S}^n_+ : \mathrm{Range}(Y) \subseteq \mathrm{Range}(X)\}$$
$$\mathit{face}(X) = \{Y \in \mathcal{S}^n_+ : \mathrm{Null}(Y) \supseteq \mathrm{Null}(X)\}$$
$$\mathrm{ri}\, \mathit{face}(X) = \{Y \in \mathcal{S}^n_+ : \mathrm{Range}(Y) = \mathrm{Range}(X)\}$$

Semidefinite Programming, SDP, $\mathcal{S}^n_+$

$K = \mathcal{S}^n_+ = K^*$: nonpolyhedral, self-polar, facially exposed

$$\text{(SDP-P)} \quad v_P = \sup_{y \in \mathbb{R}^m} b^\top y \text{ s.t. } g(y) := \mathcal{A}^* y - c \preceq_{\mathcal{S}^n_+} 0$$

$$\text{(SDP-D)} \quad v_D = \inf_{x \in \mathcal{S}^n} \langle c, x \rangle \text{ s.t. } \mathcal{A} x = b, \; x \succeq_{\mathcal{S}^n_+} 0$$

where:

- PSD cone $\mathcal{S}^n_+ \subset \mathcal{S}^n$ symm. matrices
- $c \in \mathcal{S}^n$, $b \in \mathbb{R}^m$
- $\mathcal{A} : \mathcal{S}^n \to \mathbb{R}^m$ is an onto linear map, with adjoint $\mathcal{A}^*$
- $\mathcal{A}x = (\operatorname{trace} A_i x) = (\langle A_i, x \rangle) \in \mathbb{R}^m, \quad A_i \in \mathcal{S}^n$
  $\mathcal{A}^* y = \sum_{i=1}^m A_i y_i \in \mathcal{S}^n$

View One for FR in SDP

$(SDP_D) \quad \min\{\text{trace } CX \text{ s.t. } \mathcal{A}X = b, X \in \mathcal{S}^n_+\}$

Step 1: Let $0 \neq Z \succeq 0$ be an exposing vector.

add constraint trace $ZX = 0$. (Equivalently $ZX = 0$)

from spectral decomposition of Z, with Range P = Null Z:

substitute: $X = P\mathbb{S}^{t_1}_+ P^T, \qquad t_1 = \text{nullity}(Z)$

We get the equivalent smaller problem

$$(SDP_{D1}) \quad \begin{array}{ll} \min & \text{trace}(P^TCP)R \\ \text{s.t.} & \text{trace}(P^TA_iP)R = b_i, i = 1, \ldots, m \\ & R \in \mathbb{S}^{t_1}_+ \end{array}$$

Remove/delete redundant linear constraints; repeat Step 1.

minimum number of steps: singularity degree (ref. Jos Sturm)

View Two for FR in SDP

Lemma: Using exposing vectors

Let

$$Z_i \succeq 0, F_i = \mathcal{S}_+^n \cap Z_i^\perp, i = 1, \ldots, m.$$

Then

$$\cap_{i=1}^m F_i = \mathcal{S}_+^n \cap \left(\sum_{i=1}^m Z_i \right)^\perp$$

intersection of faces is exposed by sum of exposing vectors □

** FR - Motivation/Application; EDM, SNL

Highly (implicit) degenerate/low-rank problem

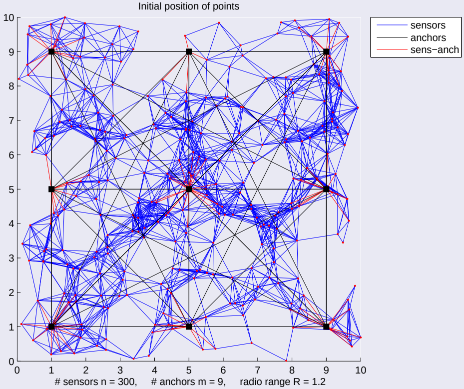
- high (implicit) degeneracy translates to low rank solutions
- take advantage of degeneracy; fast, high accuracy solutions

SNL - a Fundamental Problem of Distance Geometry; easy to describe - dates back to Grasssmann 1886

- r : embedding dimension
- n ad hoc wireless sensors $p_1, \ldots, p_n \in \mathbb{R}^r$ to locate in $\mathbb{R}^r$;
- m of the sensors $p_{n-m+1}, \ldots, p_n$ are anchors (positions known, using e.g. GPS)
- pairwise distances $D_{ij} = \|p_i - p_j\|^2, ij \in E$, are known within radio range $R > 0$
- $$P^\top = \begin{bmatrix} p_1 & \ldots & p_n \end{bmatrix} = \begin{bmatrix} X^\top & A^\top \end{bmatrix} \in \mathbb{R}^{r \times n}$$

Sensor Localization Problem/Partial EDM

Sensors ◦ and Anchors ■



Popular Techniques; SDP Relax.; Highly Degen.

$$\mathcal{K}: \mathcal{S}^n \to \mathcal{S}^n \qquad \mathcal{K}(B) = \operatorname{diag}(B)e^T + e\operatorname{diag}(B)^T - 2B$$

Nearest, Weighted, SDP Approx. (relax/discard $\operatorname{rank} B$)

- $\min_{B \succeq 0} \|H \circ (\mathcal{K}(B) - D)\|$

  $\operatorname{rank} B = r; \qquad H_{ij} = \begin{cases} 1/\sqrt{D_{ij}} & \text{if } ij \in E, \\ H_{ij} = 0 & \text{otherwise} \end{cases}$
- with rank constraint: a non-convex, NP-hard program
- SDP relaxation is convex
  BUT: expensive/low accuracy/implicitly highly degenerate

cliques restrict ranks of feasible B to get FR

$$B = VRV^T$$

Random Noisless Problems, Krislock W. **2010 laptop**

- 2.16 GHz Intel Core 2 Duo, 2 GB of RAM
- Dimension $r = 2$
- Square region: $[0, 1] \times [0, 1]$
- $m = 9$ anchors
- Using only Rigid Clique Union and Rigid Node Absorption
- Error measure: Root Mean Square Deviation

$$\text{RMSD} = \left(\frac{1}{n} \sum_{i=1}^{n} \|p_i - p_i^{\text{true}}\|^2 \right)^{1/2}$$

Results - N Huge/High Accuracy

Large-Scale Problems (results from **2010**)

# sensors	# anchors	radio range	RMSD	Time
20000	9	.025	$5e{-}16$	25s
40000	9	.02	$8e{-}16$	1m 23s
60000	9	.015	$5e{-}16$	3m 13s
100000	9	.01	$6e{-}16$	9m 8s

Size of SDPs Solved: $N = \binom{n}{2}$ (# vrbls)

$\mathcal{E}_n(\text{density of } \mathcal{G}) = \pi R^2$; $M = \mathcal{E}_n(|E|) = \pi R^2 N$ (# constraints)

Size of SDP Problems:

$M = [3,078,915 \quad 12,315,351 \quad 27,709,309 \quad 76,969,790]$

$N = 10^9 \, [0.2000 \quad 0.8000 \quad 1.8000 \quad 5.0000]$

*** FR for LRMC (Huang-W.'16)

Intractable (nonconvex) minimum rank completion

Given partial $m \times n$ real matrix $Z \in \mathbb{R}^{m \times n}$.

$$(LRMC) \quad \begin{array}{ll} \min & \operatorname{rank}(M) \\ \text{s.t.} & \|M_{\hat{E}} - Z_{\hat{E}}\| \leq \delta, \end{array}$$

$\hat{E}$ sampled indices; $Z_{\hat{E}} \in \mathbb{R}^{\hat{E}}$; $\delta > 0$ tuning parameter

convex nuclear norm relaxation

$$\begin{array}{ll} \min & \|M\|_* \\ \text{s.t.} & \|M_{\hat{E}} - Z_{\hat{E}}\| \leq \delta, \end{array}$$

where $\|M\|_* = \sum_i \sigma_i(M)$ (sum of singular values)

SDP Equivalent to Nuclear Norm Minimization

Trace minimization

$$\begin{array}{ll} \min & \frac{1}{2}\|Y\|_* = \frac{1}{2}\operatorname{trace}(Y) \\ \text{s.t.} & \|Y_{\bar{E}} - Q_{\bar{E}}\| \le \delta \\ & Y \in \mathbb{S}_+^{m+n}, \end{array}$$

$Q = \begin{bmatrix} 0 & Z \\ Z^T & 0 \end{bmatrix} \in \mathbb{S}_+^{m+n}$ and $\bar{E}$ indices in Y corresponding to $\hat{E}$

Noiseless case: strict feasibility trivially holds

$$Y_{\bar{E}} = Q_{\bar{E}}$$

choose diagonal of Y sufficiently large, positive.
(strict feas. holds for dual as well)

Why consider this here?

one can exploit the structure at the optimum and efficiently apply FR(identify active set)

Associated Undirected Weighted Graph $G = (V, E, W)$

node set $V = \{1, \ldots, m, m+1, \ldots, m+n\}$ Let:

$$
\begin{aligned}
E_{1,m} &:= \{ij \in V \times V : i < j \leq m\} \\
E_{m+1,m+n} &:= \{ij \in V \times V : m+1 \leq i < j \leq m+n\}
\end{aligned}
$$

edge set

$$E := \bar{E} \cup E_{1,m} \cup E_{m+1,m+n} \qquad \text{(include diagonal blocks)}$$

weights for all $ij \in E$

$$w_{ij} := \begin{cases} Z_{i(j-m)}, & \forall ij \in \bar{E} \\ 0, & \text{otherwise.} \end{cases}$$

Corresponding adjacency matrix A; cliques C

nontrivial cliques of interest (after row/col perms) corresp. to full (specified) submatrix X in Z; $C = \{i_1, \ldots, i_k\}$ with cardinalities

$$|C \cap \{1, \ldots, m\}| = p \neq 0, \quad |C \cap \{m+1, \ldots, m+n\}| = q \neq 0.$$

Exposing Vector for Low-Rank Completions

Clique - X; generically rank r by lsc of rank

$$X \equiv \{Z_{i(j-m)} : ij \in C\}, \quad \text{specified } p \times q \text{ submatrix.}$$

Wlog

$$Z = \begin{bmatrix} Z_1 & Z_2 \\ X & Z_3 \end{bmatrix},$$

full rank factorization $X = \bar{P}\bar{Q}^T$ using SVD

$$X = \bar{P}\bar{Q}^T = U_X \Sigma_X V_X^T,\ \Sigma_X \in \mathbb{S}_{++}^{r_X}, \quad \bar{P} = U_X \Sigma_X^{1/2},\ \bar{Q} = V_X \Sigma_X^{1/2}.$$

$$C_X = \{i, \dots, m, m+1, \dots, m+k\}, \qquad r < \max\{p, q\},$$

target rank r

rewrite optimality conditions for SDP as:

$$0 \preceq Y = \begin{bmatrix} U \\ P \\ Q \\ V \end{bmatrix} D \begin{bmatrix} U \\ P \\ Q \\ V \end{bmatrix}^T = \left[\begin{array}{c|cc|c} UDU^T & UDP^T & UDQ^T & UDV^T \\ \hline PDU^T & \boxed{PDP^T} & PDQ^T & PDV^T \\ QDU^T & QDP^T & \boxed{QDQ^T} & QDV^T \\ \hline VDU^T & VDP^T & VDQ^T & VDV^T \end{array}\right]$$

Using exposing vectors

Lemma (Basic FR)

Let $r < \min\{p, q\}$ and $X = PDQ^T = \bar{P}\bar{Q}^T$ as above. We find a pair of exposing vectors using

$$\mathrm{FR}(\bar{P}, \bar{Q}): \ \bar{P}\bar{P}^T + \boxed{\bar{U}\bar{U}^T} \succ 0, \ \bar{P}^T\bar{U} = 0,$$

$$\bar{Q}\bar{Q}^T + \boxed{\bar{V}\bar{V}^T} \succ 0, \ \bar{Q}^T\bar{V} = 0.$$

Fill out the exposing vectors with zeros to get Z_U, Z_V, resp. Then

$$Z = Z_U + Z_V \qquad \text{is an exposing vector for opt. } Y$$

Add them all up!

Numerics LRMC/average over 5 instances

largest: $mn = 50,000,000$ elements

Table: noiseless: $r = 2$; $m \times n$ size $\uparrow$.

Specifications			Time (s)	Rank	Residual (%Z)
m	n	mean(p)			
700	2000	0.30	9.00	2.0	4.4605e-14
1000	5000	0.30	28.76	2.0	3.0297e-13
1400	9000	0.30	77.59	2.0	7.8674e-14
1900	14000	0.30	192.14	2.0	6.7292e-14
2500	20000	0.30	727.99	2.0	4.2753e-10

Table: noiseless: $r = 4$; $m \times n$ size $\uparrow$.

Specifications			Time (s)	Rank	Residual (%Z)
m	n	mean(p)			
700	2000	0.36	12.80	4.0	1.5217e-12
1000	5000	0.36	49.66	4.0	1.0910e-12
1400	9000	0.36	131.53	4.0	6.0304e-13
1900	14000	0.36	291.22	4.0	3.4847e-11
2500	20000	0.36	798.70	4.0	7.2256e-08

Wrong Rank; Refinement Step with Dual Multiplier

δ_0 best target value; $b = Z_{\hat{E}}$ known entries; equiv. NN problem:

$$\begin{array}{ll} \min & \operatorname{trace}(R) \\ \text{s.t.} & \|(VRV^T)_{\hat{E}} - b)\| \leq \delta_0 \\ & R \succeq 0. \end{array}$$

Flip the problem/parametric approach, (DKVW, '14)

Refinement step:

$$\begin{array}{lll} \varphi(\tau) := & \min & \| \left(\hat{V} R \hat{V}^T\right)_{\hat{E}} - b\| + \gamma\|R\|_F \\ & \text{s.t.} & \operatorname{trace}(R) \leq \tau \\ & & R \succeq 0. \end{array}$$

- regularizing term $+\gamma\|R\|_F$ added to underdet. prob.
- starting value τ from unconstrained prob.
- shrink the trace of R to reduce the resulting rank.
- stop when dual multiplier > 0.

Preliminary Numerics LRMC/avg. 5 instances

largest: $mn = 1,800,000$ elements

Table: noisy: $r = 3$; $m \times n$ size $\uparrow$; density p.

Specifications				Recover (%Z)	Time (s)		Rank		Residual (%Z)	
m	n	% noise	mean(p)		initial	refine	initial	refine	initial	refine
700	1000	0.0e+00	0.18	86.90	1.87	4.85	3.20	3.00	3.76e-07	3.66e-07
700	1000	1.0e-01	0.18	86.75	2.81	31.12	4.25	3.75	2.07e+03	9.78e+01
700	1000	2.0e-01	0.18	86.75	2.77	31.60	4.25	3.50	2.88e+03	2.70e+02
700	1000	3.0e-01	0.18	86.75	2.69	32.35	4.00	3.25	2.96e+03	1.59e+02
700	1000	4.0e-01	0.18	86.90	2.61	35.53	4.00	3.60	6.09e+04	4.10e+02
700	1000	1.0e-03	0.33	100.00	4.93	13.04	3.00	3.00	2.43e-03	2.43e-03
700	1000	1.0e-03	0.30	100.00	4.62	14.36	3.80	3.00	1.77e-02	1.77e-02
700	1000	1.0e-03	0.26	99.69	4.19	16.04	3.00	3.00	6.94e-02	6.94e-02
700	1000	1.0e-03	0.22	97.77	3.81	13.91	3.40	3.00	9.74e-01	8.42e-01
700	1000	1.0e-03	0.18	86.75	2.93	13.23	4.75	3.00	3.54e+00	1.65e+00
900	2000	1.0e-04	0.18	96.81	8.01	26.45	4.60	3.00	9.72e-02	9.71e-02
900	2000	1.0e-04	0.16	92.60	6.04	18.93	4.80	3.00	2.10e+00	7.20e-01
900	2000	1.0e-04	0.16	89.45	5.41	31.18	4.00	3.25	5.72e+01	7.57e-01
900	2000	1.0e-04	0.15	83.53	4.71	18.10	5.00	3.00	2.61e-01	2.60e-01
900	2000	1.0e-04	0.14	74.94	7.21	28.49	4.00	3.00	4.05e+01	5.70e+00

** Conclusion

Preprocessing

- Though strict feasibility holds generically, failure appears in many applications. Loss of strict feasibility is directly related to ill-posedness and difficulty in numerical methods.
- Preprocessing based on structure can both *regularize* and simplify the problem. In many cases one gets an optimal solution without the need of any SDP solver.

Exploit structure at optimum

For low-rank matrix completion the structure at the optimum can be exploited to apply FR even though strict feasibility holds. (Related to active set identification.)

J.M. Borwein and H. Wolkowicz, *Characterization of optimality for the abstract convex program with finite-dimensional range*, J. Austral. Math. Soc. Ser. A **30** (1980/81), no. 4, 390–411. MR 83i:90156

Y-L. Cheung, S. Schurr, and H. Wolkowicz, *Preprocessing and regularization for degenerate semidefinite programs*, Computational and Analytical Mathematics, In Honor of Jonathan Borwein's 60th Birthday (D.H. Bailey, H.H. Bauschke, P. Borwein, F. Garvan, M. Thera, J. Vanderwerff, and H. Wolkowicz, eds.), Springer Proceedings in Mathematics & Statistics, vol. 50, Springer, 2013, pp. 225–276.

D. Drusvyatskiy, N. Krislock, Y-L. Cheung Voronin, and H. Wolkowicz, *Noisy sensor network localization: robust facial reduction and the Pareto frontier*, Tech. report, University of Waterloo, Waterloo, Ontario, 2014, arXiv:1410.6852, 20 pages.

D. Drusvyatskiy, G. Pataki, and H. Wolkowicz, *Coordinate shadows of semidefinite and Euclidean distance matrices*, SIAM J. Optim. **25** (2015), no. 2, 1160–1178. MR 3357643

D. Drusvyatskiy and H. Wolkowicz, *The many faces of degeneracy in conic optimization*, Tech. report, University of Waterloo, Waterloo, Ontario, 2016, survey in progress.

S. Huang and H. Wolkowicz, *Low-rank matrix completion using nuclear norm with facial reduction*, Tech. report, University of Waterloo, Waterloo, Ontario, 2016, in progress.

N. Krislock and H. Wolkowicz, *Explicit sensor network localization using semidefinite representations and facial reductions*, SIAM Journal on Optimization **20** (2010), no. 5, 2679–2708.

G. Reid, F. Wang, H. Wolkowicz, and W. Wu, *Facial reduction and SDP methods for systems of polynomial equations*, Tech. report, University of Western Ontario, London, Ontario, 2014, submitted Dec. 2014, 38 pages.

Low-Rank Matrix Completion (LRMC) using Nuclear Norm (NN) with Facial Reduction (FR)

Henry Wolkowicz

Dept. Combinatorics and Optimization, University of Waterloo, Canada

Sat. Nov. 5, 2016, 10:30-11:00

4th Alpen-Adria-Workshop on Optimization, Klagenfurt