School of Mathematics

Interior Point Methods for Large Scale Optimization

Jacek Gondzio
J.Gondzio@ed.ac.uk
http://www.maths.ed.ac.uk/~gondzio

Outline

- **Interior Point Methods for Optimization**
  - log barrier, first-order conditions, Newton method
  - linear, quadratic, semidefinite programming, etc.
  - IPMs are well suited to large scale optimization
- **Sparse Approximations: Signal/Image Processing**
  - Inverse problems
    $\rightarrow$ ℓ_1-regularized least squares
  - Machine Learning (and Big Data)
- **Plastic Truss Layout Optimization**
  - Ground structures
    $\rightarrow$ linear programming formulation
  - Stability constraints
    $\rightarrow$ semidefinite programming formulation
  - Geometry optimization
- **Final Remarks**

Observation

Numerous practical (engineering) problems can be cast as the following optimization problems

LP:

$$\begin{array}{rl} \min & c^T x \\ \text{s.t.} & Ax = b, \\ & x \geq 0. \end{array}$$

QP:

$$\begin{array}{rl} \min & c^T x + \frac{1}{2} x^T Q x \\ \text{s.t.} & Ax = b, \\ & x \geq 0. \end{array}$$

SDP:

$$\begin{array}{rl} \min & C \bullet X \\ \text{s.t.} & \mathcal{A}X = b, \\ & X \succeq 0, \end{array}$$

where $X \in \mathcal{SR}^{n \times n}$.

Observation

All these problems can be solved efficiently using

Interior Point Methods for Optimization

J. Gondzio,
Interior Point Methods 25 Years Later,
European Journal of Operational Research 218 (2012) 587–601.

J. Gondzio,
Convergence Analysis of an Inexact Feasible IPM for Convex QP,
SIAM J. on Optimization, 23 (2013) No 3, 1510–1527.

Interior Point Methods

Shocking mathematical concept: A step against common sense and many centuries of mathematical practice:

"nonlinearize" the linear problem

Take **linear** optimization problem
and add **nonlinear** function to the objective.

Mathematical "elements" of the IPM

What do we need to derive the **IPM**?

- duality theory:
  Lagrangian function;
  first order optimality conditions.
- logarithmic barriers.
- Newton method.

Primal-Dual Pair of Linear Programs

Primal Dual

$$\begin{array}{llll} \min & c^T x & & \\ \text{s.t.} & Ax & = & b, \\ & x \ge 0; & & \end{array} \qquad\qquad \begin{array}{llll} \max & b^T y & & \\ \text{s.t.} & A^T y + s & = & c, \\ & s \ge 0. & & \end{array}$$

Lagrangian

$$L(x,y) = c^T x - y^T (Ax - b) - s^T x.$$

Optimality Conditions

$$\begin{aligned} Ax &= b, \\ A^T y + s &= c, \\ XSe &= 0, \quad (\text{ i.e., } x_j \cdot s_j = 0 \quad \forall j), \\ (x,s) &\ge 0, \end{aligned}$$

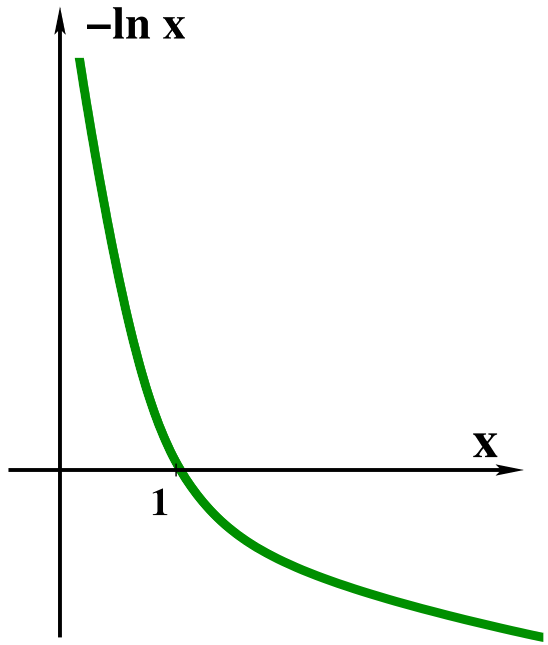
$X = diag\{x_1, \cdots, x_n\}$, $S = diag\{s_1, \cdots, s_n\}$, $e = (1, \cdots, 1) \in \mathcal{R}^n$.

Logarithmic barrier

$$-\ln x_j$$

"replaces" the inequality

$$x_j \geq 0 \ .$$



Observe that

$$\min \ \mathrm{e}^{-\sum_{j=1}^n \ln x_j} \quad \Longleftrightarrow \quad \max \prod_{j=1}^n x_j$$

The minimization of $-\sum_{j=1}^n \ln x_j$ is equivalent to the maximization of the product of distances from all hyperplanes defining the positive orthant: it prevents all x_j from approaching zero.

Logarithmic barrier

Replace the **primal** LP

$$\begin{array}{lrl} \min & c^T x & \\ \text{s.t.} & Ax & = b, \\ & x \geq 0, & \end{array}$$

with the **primal barrier program**

$$\begin{array}{ll} \min & c^T x - \mu \sum_{j=1}^{n} \ln x_j \\ \text{s.t.} & Ax = b. \end{array}$$

Lagrangian: $$L(x, y, \mu) = c^T x - y^T (Ax - b) - \mu \sum_{j=1}^{n} \ln x_j.$$

Conditions for a stationary point of the Lagrangian

$$\begin{aligned} \nabla_x L(x,y,\mu) &= c - A^T y - \mu X^{-1} e = 0 \\ \nabla_y L(x,y,\mu) &= \qquad\qquad\quad Ax - b = 0, \end{aligned}$$

where $X^{-1} = diag\{x_1^{-1}, x_2^{-1}, \cdots, x_n^{-1}\}$.

Let us denote

$$s = \mu X^{-1} e, \qquad \text{i.e.} \qquad XSe = \mu e.$$

The **First Order Optimality Conditions** are:

$$\begin{aligned} Ax &= b, \\ A^T y + s &= c, \\ XSe &= \boldsymbol{\mu} e, \\ (x,s) &> 0. \end{aligned}$$

Apply Newton Method to the FOC

The first order optimality conditions for the barrier problem form a large system of nonlinear equations

$$f(x, y, s) = 0,$$

where $f : \mathcal{R}^{2n+m} \mapsto \mathcal{R}^{2n+m}$ is a mapping defined as follows:

$$f(x, y, s) = \begin{bmatrix} Ax - b \\ A^T y + s - c \\ XSe - \mu e \end{bmatrix}.$$

Actually, the first two terms of it are **linear**; only the last one, corresponding to the complementarity condition, is **nonlinear**.

Interior-Point Framework

The **logarithmic barrier**

$$-\ln x_j$$

"replaces" the inequality $\quad x_j \geq 0.$

We derive the **first order optimality conditions** for the primal barrier problem:

$$\begin{aligned} Ax &= b, \\ A^T y + s &= c, \\ XSe &= \mu e, \end{aligned}$$

and apply **Newton method** to solve this system of (nonlinear) equations.

Actually, we fix the barrier parameter μ and make only **one** (damped) Newton step towards the solution of FOC. We do not solve the FOC exactly. Instead, we immediately reduce the barrier parameter μ (to ensure progress towards optimality) and repeat the process.

Self-concordant Barrier

Def: Let $C \in \mathcal{R}^n$ be an open nonempty convex set.

Let $f : C \mapsto \mathcal{R}$ be a 3 times continuously diff'able convex function.

A function f is called **self-concordant** if there exists a constant $p > 0$ such that

$$|\nabla^3 f(x)[h,h,h]| \leq 2p^{-1/2}(\nabla^2 f(x)[h,h])^{3/2},$$

$\forall x \in C$, $\forall h: \ x+h \in C$. (We then say that f is p-self-concordant).
Note that a self-concordant function is always well approximated by the quadratic model because the error of such an approximation can be bounded by the 3/2 power of $\nabla^2 f(x)[h,h]$.

Lemma The barrier function $-\log x$ is self-concordant on $\mathcal{R}_+$.

Nesterov and Nemirovskii,
Interior Point Polynomial Algorithms in Convex Programming: Theory and Applications, SIAM, 1994.

From LP via QP to NLP, SOCP and SDP

For the quadratic cone

$$K_q = \{(x,t) : x \in \mathcal{R}^{n-1}, t \in \mathcal{R},\ t^2 \geq \|x\|^2,\ t \geq 0\},$$

define the *logarithmic barrier function*, $f : \mathcal{R}^n \mapsto \mathcal{R}$

$$f(x,t) = \begin{cases} -\ln(t^2 - \|x\|^2) & \text{if } \|x\| < t \\ +\infty & \text{otherwise.} \end{cases}$$

For the cone $\mathcal{SR}_+^{n \times n}$ of positive definite matrices,
define the *logarithmic barrier function*, $f : \mathcal{SR}_+^{n \times n} \mapsto \mathcal{R}$

$$f(X) = \begin{cases} -\ln \det X & \text{if } X \succ 0 \\ +\infty & \text{otherwise.} \end{cases}$$

LP: Replace $x \geq 0$ with $-\mu \sum_{j=1}^{n} \ln x_j$.

SDP: Replace $X \succeq 0$ with $-\mu \sum_{j=1}^{n} \ln \lambda_j = -\mu \ln(\prod_{j=1}^{n} \lambda_j)$.

Interior Point Methods:

- Unified view of optimization
  $\rightarrow$ from LP via QP to NLP, SOCP and SDP
- Predictable behaviour
  $\rightarrow$ small number of iterations
- Unequalled efficiency
  - competitive for small problems ($n \leq 10^6$)
  - beyond competition for large problems ($n \geq 10^6$)

Problem of size $\mathbf{10^9}$ solved in 2005.

Object-Oriented Parallel IPM Solver (OOPS):
`http://www.maths.ed.ac.uk/~gondzio/parallel/solver.html`

Gondzio and Grothey, Parallel IPM solver for structured QPs: application to financial planning problems, *Annals of Operations Research* 152 (2007) 319-339.

Overarching Feature of IPMs

They possess an unequalled ability to identify
the **"essential subspace"**
in which the optimal solution is hidden.

Machine Learning/Big Data

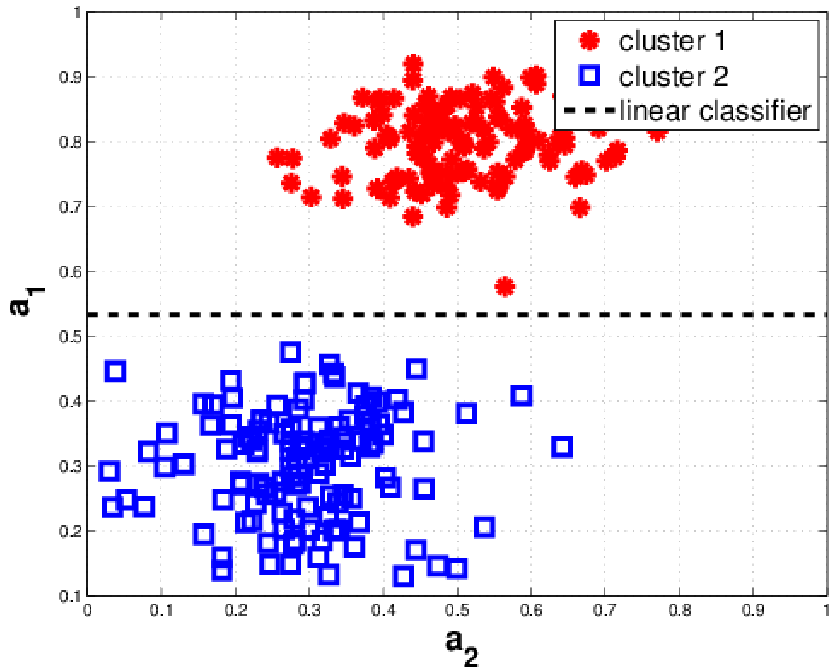
Sparse Approximation

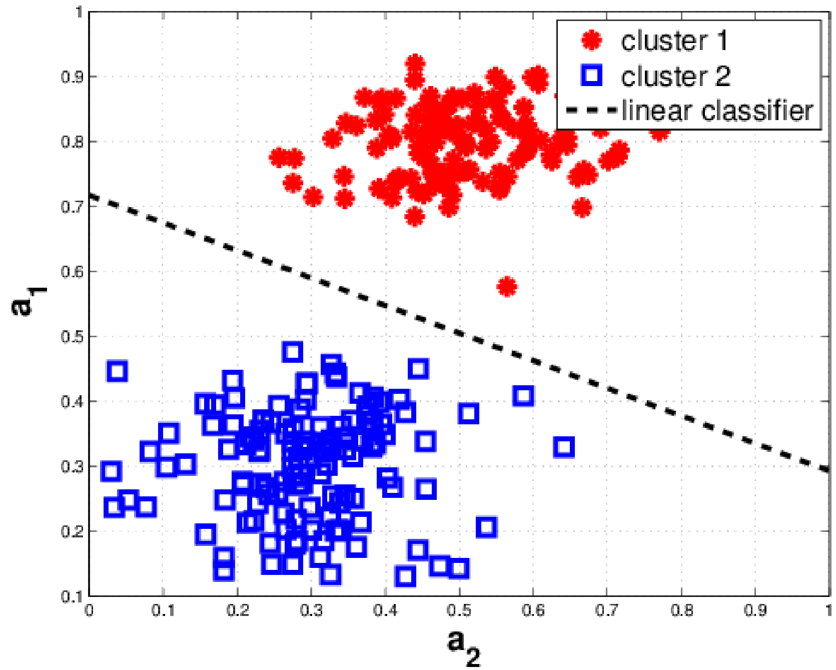
- Machine Learning: Classification with SVMs
- Statistics: Estimate x from observations
- Wavelet-based signal/image reconst. & restoration
- Compressed Sensing (Signal Processing)

All such problems lead to the same dense, possibly very large QP.

Binary Classification

$$\min \tau \|x\|_1 + \sum_{i=1}^{m} \log(1 + e^{-b_i x^T a_i}) \qquad \min \tau \|x\|_2^2 + \sum_{i=1}^{m} \log(1 + e^{-b_i x^T a_i})$$





ℓ_1-regularization

$$\min_x \quad \tau \|x\|_1 + \phi(x).$$

think of LASSO:

$$\min_x \; f(x) = \tau \|x\|_1 + \|Ax - b\|_2^2$$

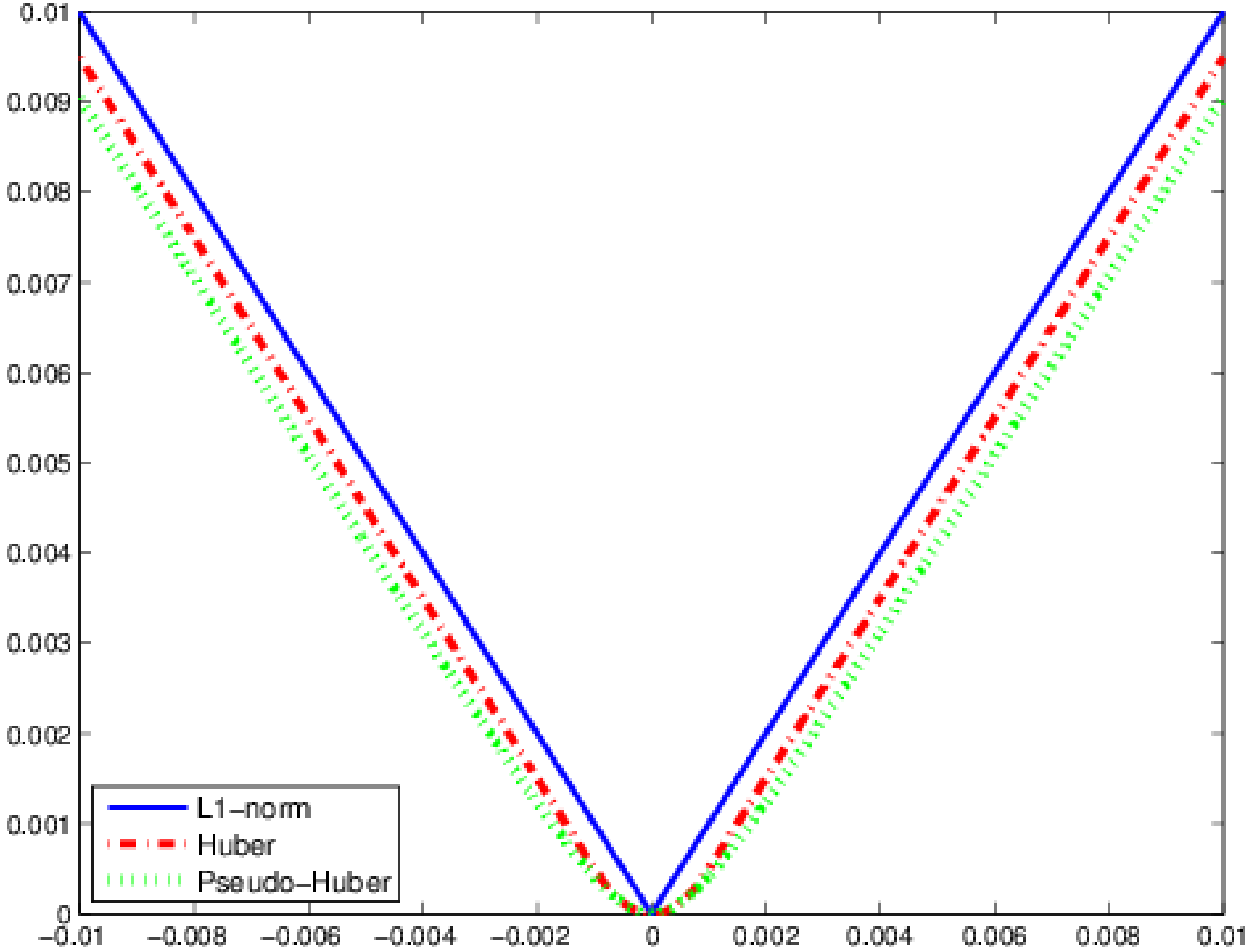
Unconstrained optimization $\Rightarrow$ easy
Serious Issue: nondifferentiability of $\|.\|_1$

Two possible tricks:

- Splitting $x = u - v$ with $u, v \geq 0$
- Smoothing with pseudo-Huber approximation

  replaces $\|x\|_1$ with $\psi_\mu(x) = \sum_{i=1}^n (\sqrt{\mu^2 + x_i^2} - \mu)$

Huber:


L1–norm
Huber
Pseudo–Huber
0.01
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0
−0.01
−0.008
−0.006
−0.004
−0.002
0
0.002
0.004
0.006
0.008
0.01

Continuation

Embed inexact Newton Meth into a *homotopy* approach:

- Inequalities $u \geq 0,\ v \geq 0 \quad \longrightarrow$ use **IPM**

  replace $z \geq 0$ with $-\mu\, log z$ and drive μ to zero.

- pseudo-Huber regression $\quad \longrightarrow$ use **continuation**

  replace $|x_i|$ with $\mu(\sqrt{1+\frac{x_i^2}{\mu^2}}-1)$ and drive μ to zero.

Questions:

- Theory?
- Practice?

Compressed Sensing and Continuation

Replace $$\min_x \quad f(x) = \tau \|W^* x\|_1 + \frac{1}{2}\|Ax - b\|_2^2, \qquad \longrightarrow \boldsymbol{x_\tau}$$

with $$\min_x \quad f_\mu(x) = \tau \psi_\mu(W^* x) + \frac{1}{2}\|Ax - b\|_2^2, \qquad \longrightarrow \boldsymbol{x_{\tau,\mu}}$$

Solve approximately a family of problems for a (short) decreasing sequence of μ's: $\boldsymbol{\mu_0 > \mu_1 > \mu_2 \cdots}$

Theorem (Brief description)

There exists a $\tilde{\mu}$ such that $\forall \mu \leq \tilde{\mu}$ the difference of the two solutions satisfies

$$\|x_{\tau,\mu} - x_\tau\|_2 = \mathcal{O}(\mu^{1/2}) \quad \forall\, \tau, \mu$$

Primal-Dual Newton Conjugate Gradient Method:

Fountoulakis and Gondzio, A Second-order Method for Strongly Convex ℓ_1-regularization Problems, *Mathematical Programming*, 156 (2016) 189–219.

Dassios, Fountoulakis and Gondzio, A Preconditioner for a Primal-Dual Newton Conjugate Gradient Method for Compressed Sensing Problems, *SIAM J on Scientific Computing*, 37 (2015) A2783–A2812.

Simple test example for ℓ_1-regularization

$$\min_x \ \tau \|x\|_1 + \|Ax - b\|_2^2$$

Special matrix given in SVD form $A = Q \Sigma G^T$, where Q and G are products of Givens rotations. The user controls:

- the condition number $\kappa(A)$,
- the sparsity of matrix A.

Matlab generator:
http://www.maths.ed.ac.uk/ERGO/trillion/

K. Fountoulakis and J. Gondzio,

Performance of First- and Second-Order Methods for ℓ_1-regularized Least Squares Problems,

Computational Optimization and Applications 65 (2016) 605–635.

Let us go big: a trillion (2^{40}) variables

n (billions)	Processors	Memory (TB)	time (s)
1	64	0.192	1923
4	256	0.768	1968
16	1024	3.072	1986
64	4096	12.288	1970
256	16384	49.152	1990
1,024	65536	196.608	2006

ARCHER (ranked 25 on `top500.com`, 11 March 2015)

Linpack Performance (Rmax) 1,642.54 TFlop/s
Theoretical Peak (Rpeak) 2,550.53 TFlop/s

Optimization of truss structures

Potential applications:

the design of

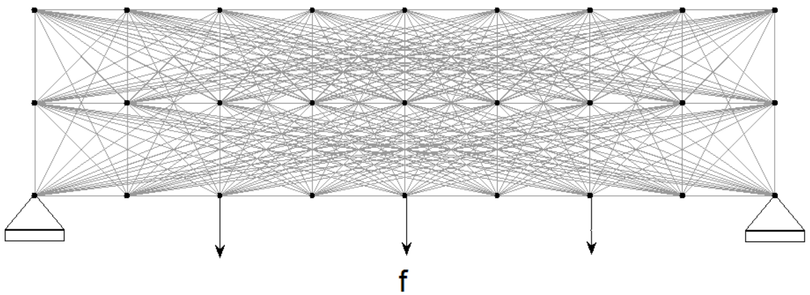
- bridges
- exoskeleton of tall buildings
- large span roof structures, etc

Optimization of truss structures

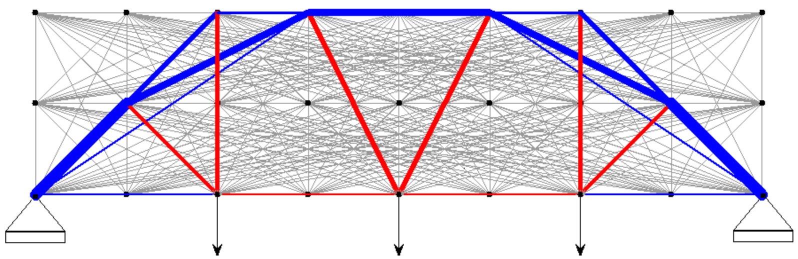
Given the following:

- d nodes,
- n bars and their lengths,
- external forces f,
- boundary conditions (some fixed nodes),

find the **lightest** truss structure that can support the applied loads.



(a) Design domain, boundary conditions, and loads

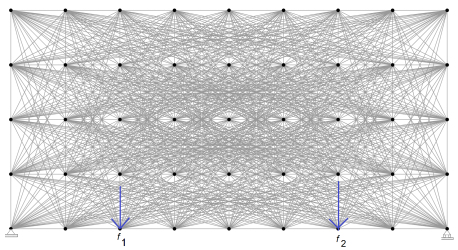


(b) Optimal design

The truss problem: plastic design formulation

$$
\begin{array}{lll}
\underset{a, q_\ell}{\text{minimize}} & l^T a & \\
\text{subject to} & B q_\ell = f_\ell, & \ell = 1, \cdots, n_L \\
 & -\sigma^- a \le q_\ell \le \sigma^+ a, & \ell = 1, \cdots, n_L \\
 & a \ge 0 &
\end{array}
$$

- n_L number of load cases,
- $l \in \mathcal{R}^n$ is a vector of bar lengths,
- $a \in \mathcal{R}^n$ is a vector of bar cross-sectional areas,
- $f_\ell \in \mathcal{R}^m$ is a vector of applied load forces,
- $q_\ell \in \mathcal{R}^n$ are axial forces in members,
- $\sigma^- > 0$ and $\sigma^+ > 0$ are the the material's yield stresses in compression and tension,
- $B \in \mathcal{R}^{m \times n}$ nodal equilibrium matrix.



This is a large-scale LP.
For d nodes, N-dim problem:
$m = Nd, \quad n = \frac{d(d-1)}{2}$
hence $n \gg m$.

A challenge for optimization methods!

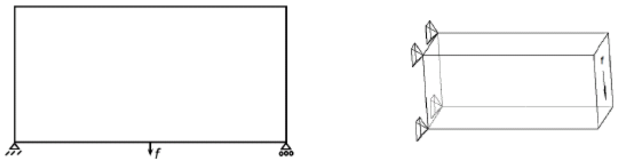
Difficult optimization problem

For fine grid, the resulting linear programming problem may be very large (m is in millions and n easily goes to billions).

Design a specialized IPM for it

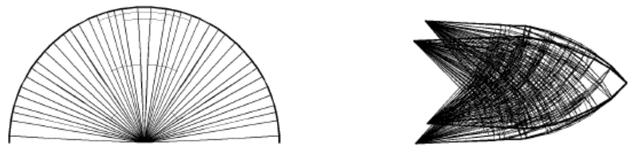
- $n \gg m$
  $\rightarrow$ use column generation technique
- a sequence of 'similar' problems to be solved
  $\rightarrow$ use warm-starting ability of IPM
- difficult linear systems to solve
  $\rightarrow$ exploit special structure of the problem, i.e.,
  use appropriately preconditioned Krylov-subspace method

Numerical results



(a) 2-D (b) 3-D

Design domain, boundary conditions, and load

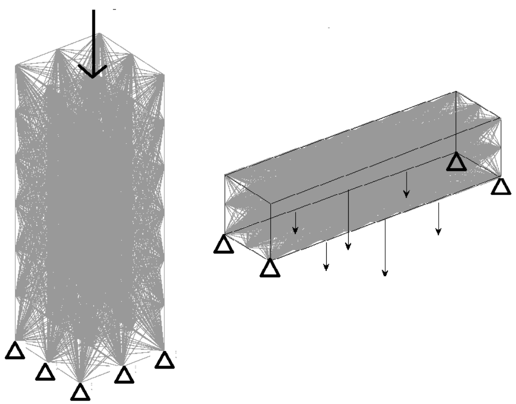


(a) 2-D (b) 3-D

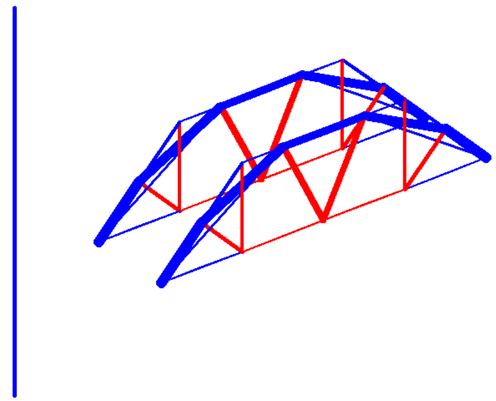
Optimal designs for the problems

	2D	3D
# of potential bars in the original problem	85,027,320	163,452,240
# of primal vars in the original problem	340,109,280	653,808,960
# of constraints in the original problem	85,053,399	16,350,647
# of member adding iterations	9	8
# of bars in the largest LP solved	201,796	498,058
# of primal vars in the largest LP solved	807,184	1,992,232
# of constraints in the largest LP solved	227,875	552,289
Total CPU[s] (applying) column generation	311	1220

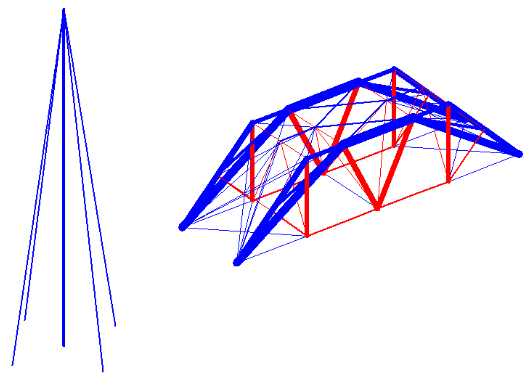
Stability constraints



Design domains, bc, and loads.



Without stability considerations.



With stability considerations.

- **Without stability considerations:**
  - A slender bar lacks any kind of support or bracing (?)
  - A bridge includes only independent planar trusses (?)
- **With stability considerations:**
  - The bar has bracing.
  - The planar trusses in the bridge are connected.

M. Stingl, On the solution of nonlinear semidefinite programs by augmented Lagrangian method, *PhD thesis* 2006, IAM II, Friedrich-Alexander U. of Erlangen–Nuremberg.

Stability constraints

The stiffness matrix $K(a)$ is given by

$$K(a) = \sum_{j=1}^{n} a_j K_j, \quad \text{with} \quad K_j = \frac{E}{l_j}\gamma_j\gamma_j^T \quad (E = \text{Young's modulus})$$

and the geometry stiffness matrix $G(q)$ is given by

$$G(q) = \sum_{j=1}^{n} q_j G_j, \quad \text{with} \quad G_j = \frac{1}{l_j}(\delta_j\delta_j^T + \eta_j\eta_j^T),$$

such that $(\delta_j, \gamma_j, \eta_j)$ are mutually orthogonal ($\eta = 0$ for 2D probs).

Global stability constraint:

$$K(a) + \tau G(q) \succeq 0.$$

M. Kočvara, On the modelling and solving of the truss design problem with global stability constraints, *Structural and Multidisciplinary Optimization* 23(2002), 189–203.

The truss problem with stability constraints

We consider the so-called global stability which is based on linear buckling. This leads to the following SDP formulation

$$
\begin{array}{lll}
\displaystyle\min_{a,q_\ell,u_\ell} & l^T a & \\
\text{s.t.} & \dfrac{a_j E}{l_j}\gamma_j^T u_\ell = q_{\ell,j}, & \forall \ell, \forall j \\
& B q_\ell = f_\ell, & \forall \ell \\
& -\sigma^- a \le q_\ell \le \sigma^+ a, & \forall \ell \\
& K(a) + \tau_\ell G(q_\ell) \succeq 0, & \forall \ell \\
& a \ge 0. &
\end{array}
$$

Ignore the kinematic compatibility constraint $\frac{a_j E}{l_j}\gamma_j^T u_\ell = q_{\ell,j}$.

But control its violation:

$$\min_{u_\ell} \quad \max_{\ell} \sum_j \left(\frac{a_j E}{l_j}\gamma_j^T u_\ell - q_{\ell,j}\right)^2$$

Constraints on eigenfrequency

Minimum compliance problem with a constraint on eigenfrequency

$$\begin{array}{ll} \min\limits_{a,u_\ell} & l^T a \\ \text{s.t.} & K(a)u_\ell = f_\ell, \quad \forall \ell \\ & f_\ell^T u_\ell \le c, \qquad \forall \ell \\ & K(a) - \lambda M(a) \succeq 0, \\ & a \ge 0. \end{array} \quad \rightarrow \quad \begin{array}{ll} \min\limits_{a} & l^T a \\ \text{s.t.} & \begin{bmatrix} c & f_\ell^T \\ f_\ell & K(a) \end{bmatrix} \succeq 0, \quad \forall \ell \\ & K(a) - \lambda M(a) \succeq 0, \\ & a \ge 0. \end{array}$$

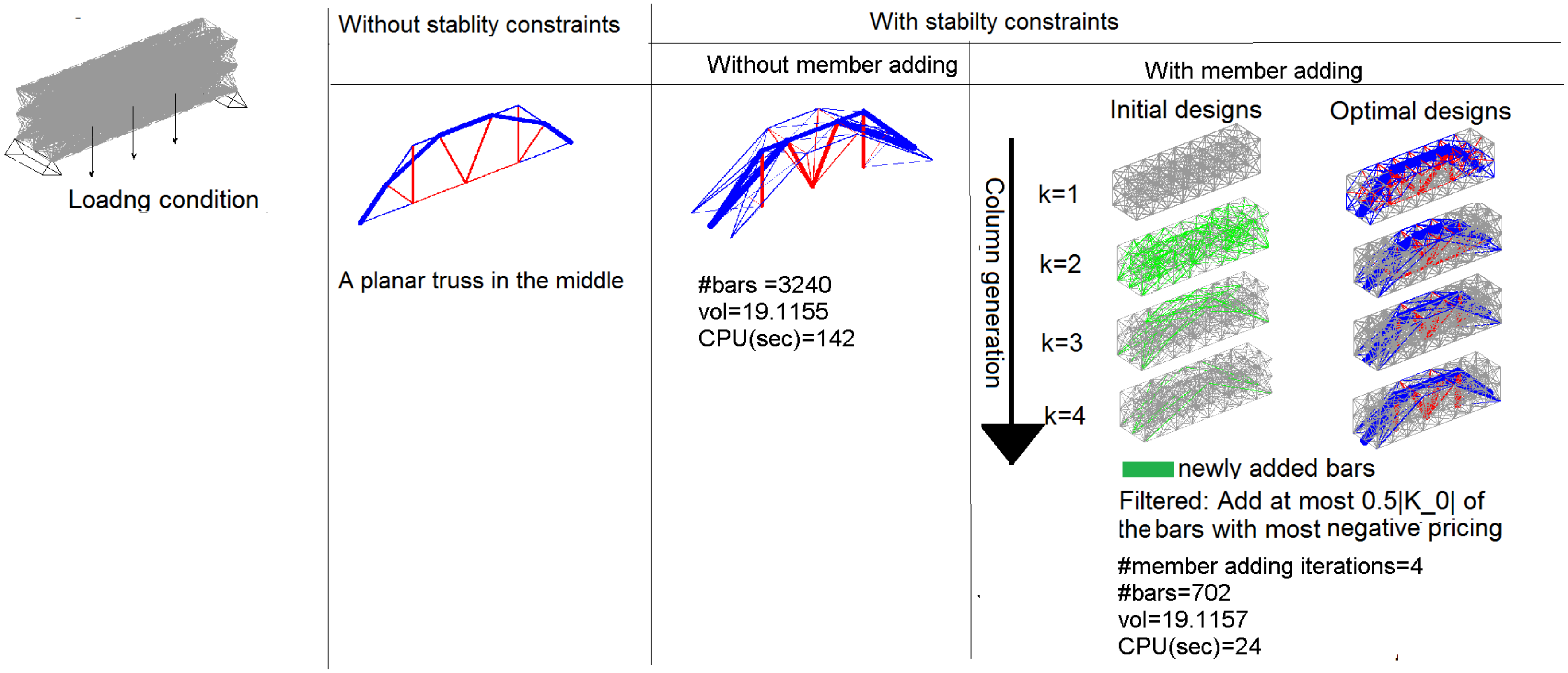
The three constraints $K(a)u = f$, $K(a) \succeq 0$ and $f^T u \le c$ are replaced with a (linear) SDP constraint

$$\begin{bmatrix} c & f^T \\ f & K(a) \end{bmatrix} \succeq 0.$$

M. Kočvara, On the modelling and solving of the truss design problem with global stability constraints, *Structural and Multidisciplinary Optimization* 23(2002), 189–203.

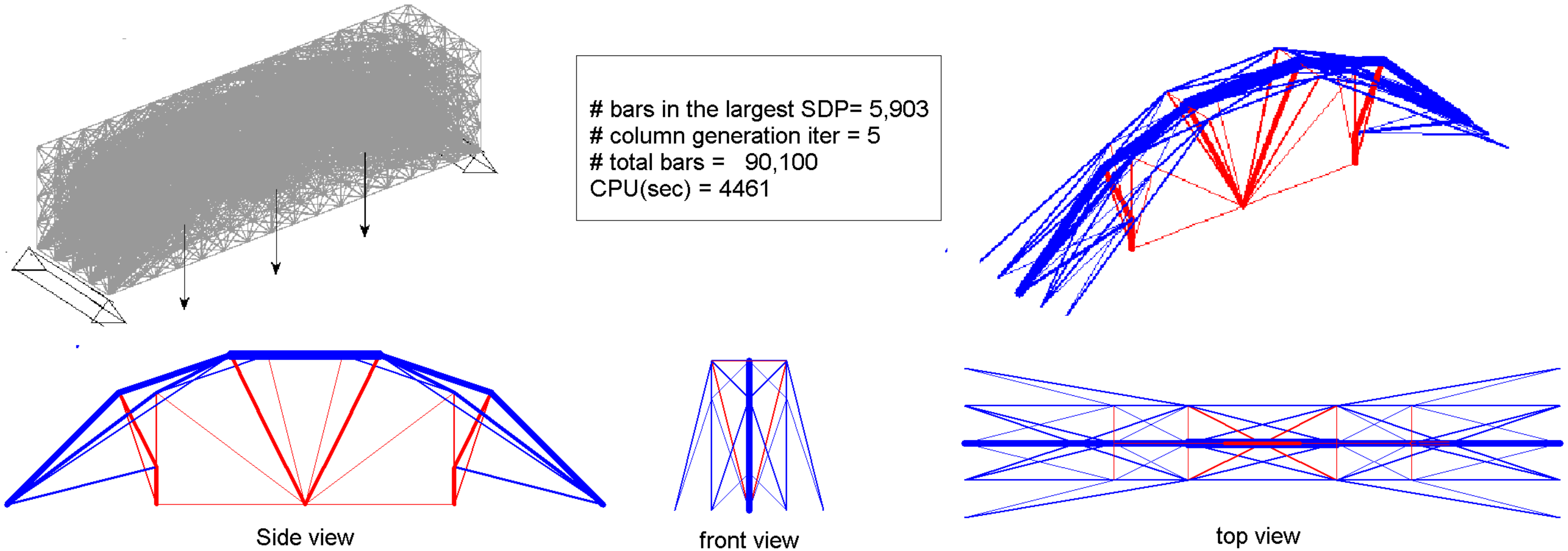
Example: The bridge problem

Small-scale problem: 3,240 bars.



Example: The bridge problem

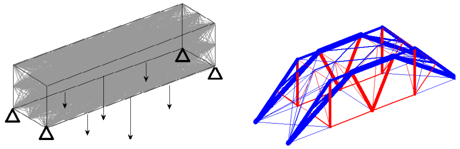
Large-scale problem: 90,100 bars.



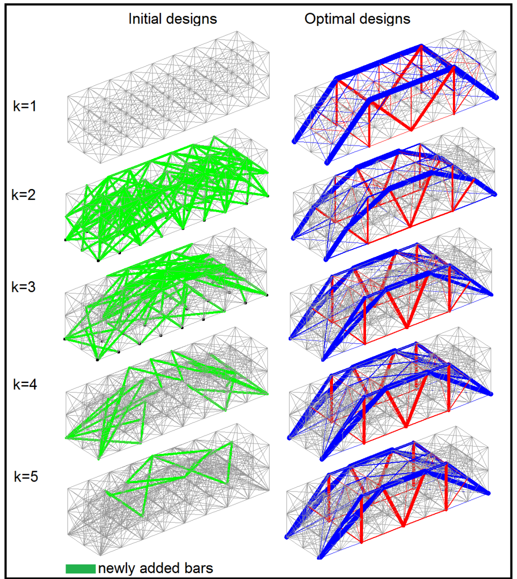
A.G. Weldeyesus and J. Gondzio,
A specialized primal-dual interior point method for the plastic truss layout optimization,
Computational Optimization and Applications, 71(2018) 613–640.

A.G. Weldeyesus, J. Gondzio, L. He, M. Gilbert, P. Shepherd, A. Tyas,
Adaptive solution of truss layout optimization problems with global stability constraints,
Structural and Multidisciplinary Optimization. https://doi.org/10.1007/s00158-019-02312-9

Example (cont'd)

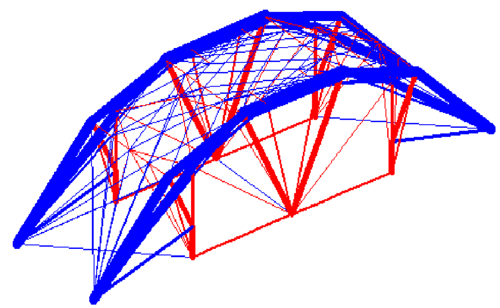


	With member adding	Without member adding
Objective function value	38.235	38.235
Final number of bars	666(21%)	3240
# col. generation iter	5	1
Total CPU(sec)	45	310



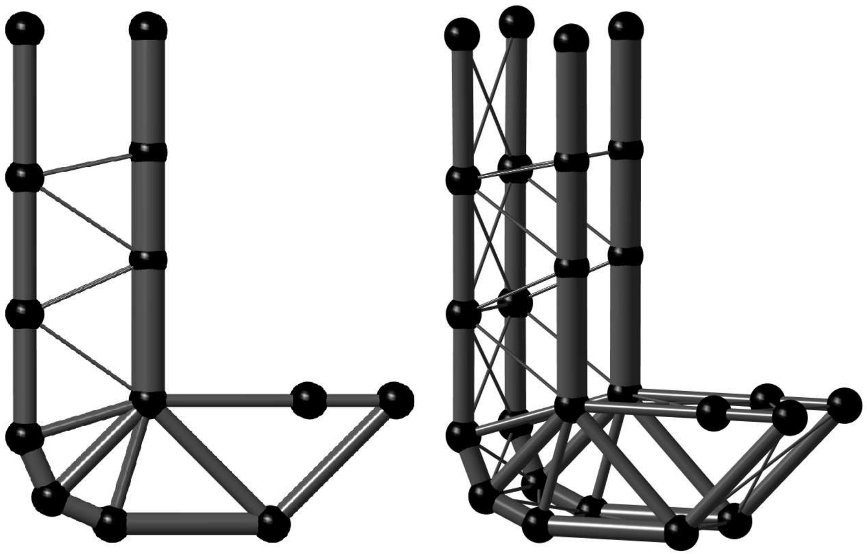
#bars in the largest SDP = 90,100 (m = 1,263, n = 180,200 in standard SDP)
#bars in original SDP = 5,644(6.4%) (m = 1,263, n = 11,288 in standard SDP)

	With WS		Without WS	
memAddIt	IPMitr	CPU(sec)	IPMitr	CPU(sec)
1	23	184	23	184
2	25	360	25	360
3	26	560	26	560
4	23	592	30	799
5	25	783	35	991
6	24	1035	35	1170
Total	146	3514	174	4064

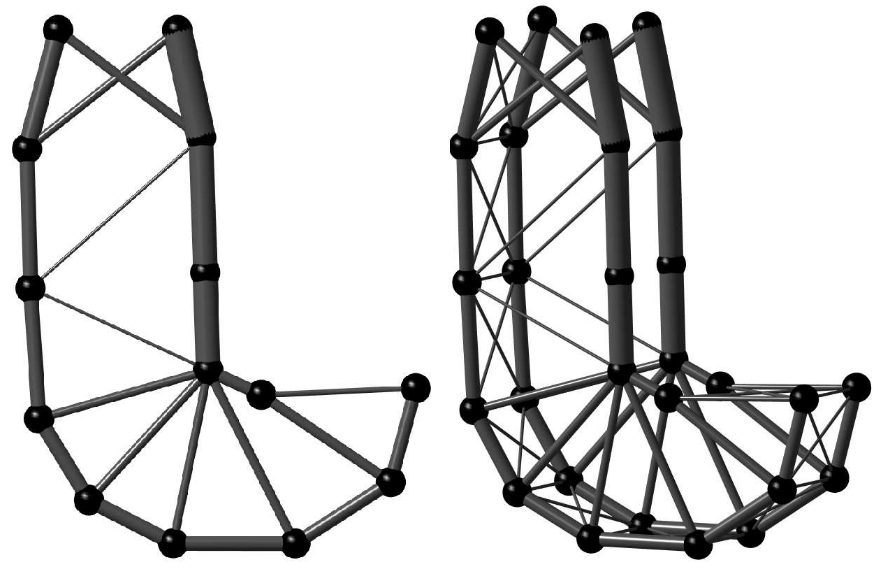


Geometry Optimization

Allow to move nodes

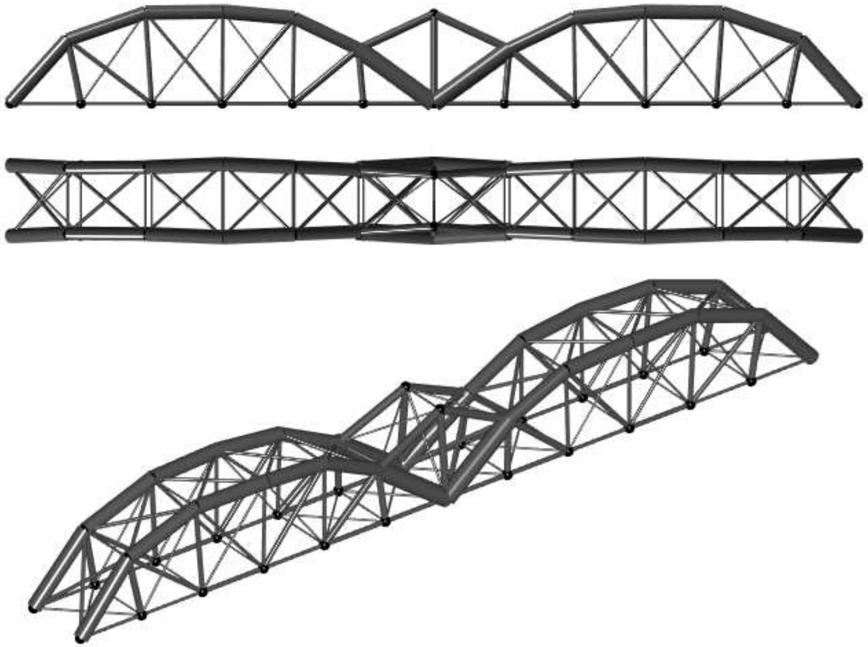


vol = 7.1191

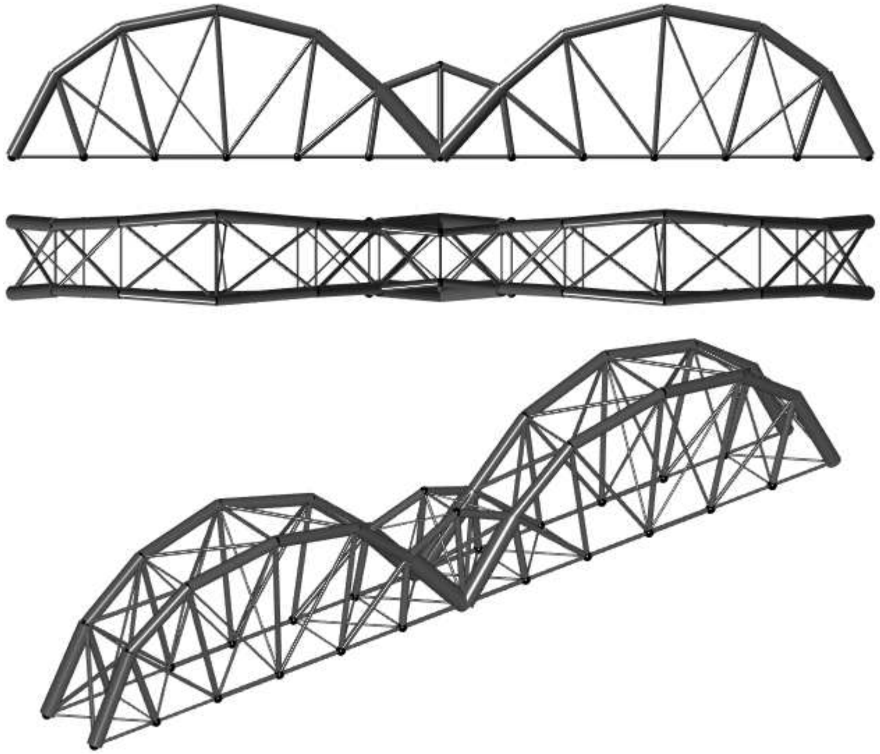


vol = 6.3128

Geometry Optimization: Bridge design



vol = 3.6980



vol = 3.1796

Conclusions

- **IPMs are well-suited to solving large scale optimization problems**
  - predictable behaviour
  - high accuracy
- **IPMs can be applied in various contexts**

Use IPMs in your research!