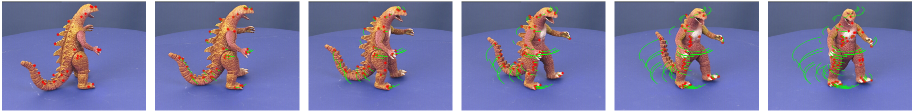
Non-Convex Relaxations for Rank Regularization

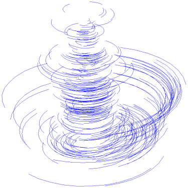
Carl Olsson

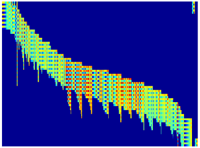
2019-05-01

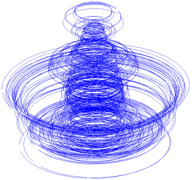
Structure from Motion and Factorization

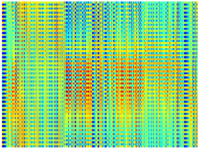


$W \odot X$ | X



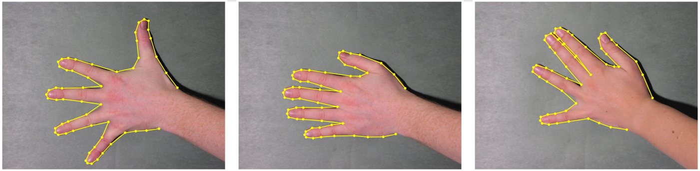






Affine camera model:

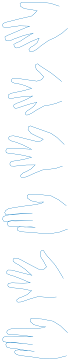
$$X = \underbrace{\begin{bmatrix} P_1 \\ P_2 \\ \vdots \end{bmatrix}}_{\text{camera matrices}} \underbrace{\begin{bmatrix} X_1 & X_2 & \dots \end{bmatrix}}_{\text{3D points}}$$

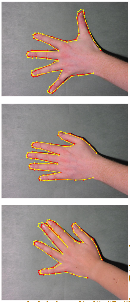


...

Linear shape basis assumption:

$$\begin{pmatrix} 0.1581 & 0.4714 & -0.9782 & 2.0509 & 1.8610 & -2.4750 \\ -0.0366 & -0.0468 & 0.2511 & 0.0532 & 0.2687 & 0.5076 \\ 0.5402 & -1.9804 & 0.4749 & -0.4343 & 2.0293 & 0.3569 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} \text{[image 2]} \end{pmatrix} = \begin{pmatrix} \text{[image 3]} \end{pmatrix}$$





Rank and Factorization

$$X = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_n \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_r \end{bmatrix}}_{B \quad (m \times r)} \underbrace{\begin{pmatrix} c_{11} & c_{21} & \dots \\ c_{12} & c_{21} & \dots \\ \vdots & \vdots & \ddots \\ c_{1r} & c_{21} & \dots \end{pmatrix}}_{C^T \quad (r \times n)}.$$

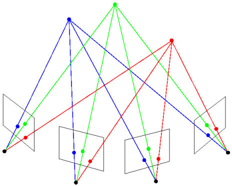
- $\text{rank}(X) = r$
- Factorization not unique: $X = BC^T = \underbrace{BG}_{\tilde{B}} \underbrace{G^{-1}C^T}_{\tilde{C}^T}$.
- DOF: $(m+n)r - r^2 << mn$
- Can reconstruct at most $mn - ((m+n)r - r^2)$ missing elements.

Small DOF desirable!

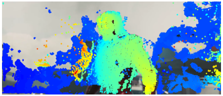
Incorporate as many constraints as possible.

Structure from Motion

Rigid reconstruction:



Non-rigid version:



Low Rank Approximation

Find the best rank r_0 approximation of X_0:

$$\min_{\text{rank}(X)=r_0} \|X - X_0\|_F^2$$

Eckart, Young (1936): Closed form solution via SVD:

If $X_0 = \sum_{i=1}^{n} \sigma_i(X_0) u_i v_i^T$ then $X = \sum_{i=1}^{r_0} \sigma_i(X_0) u_i v_i^T$.

Alternative formulation:

$$\min_{X} \mu \text{rank}(X) + \|X - X_0\|_F^2$$

Eckart, Young:

$$\sigma_i(X) = \begin{cases} \sigma_i(X_0) & \text{if } \sigma_i(X_0) \geq \sqrt{\mu} \\ 0 & \text{otherwise} \end{cases}$$

Low Rank Approximation

Generalizations:

$$\min g(\text{rank}(X)) + \|\mathcal{A}X - b\|^2 + C(X)$$

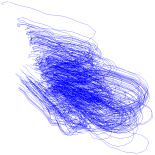
- No closed form solution.
- Non-convex.
- Discontinuous.
- Even local optimization can be difficult.

Goal: Find "flexible, easy to optimize" relaxations.

The Nuclear Norm Approach

Recht, Fazel, Parillo 2008. Replace rank(X) with $\|X\|_* = \sum_{i=1}^{n} \sigma_i(X)$.

$$\min_X \mu \|X\|_* + \|\mathcal{A}X - b\|^2$$



- Convex.
- Can be solved optimally.
- Shrinking bias. Not good for SfM!

Closed form solution to $\min_X \mu\|X\|_* + \|X - X_0\|_F^2$:

$$\text{If } X_0 = \sum_{i=1}^{n} \sigma_i(X_0) u_i v_i^T \text{ then } X = \sum_{i=1}^{n} \underbrace{\max\left(\sigma_i(X_0) - \frac{\mu}{2}, 0\right)}_{\text{soft thresholding}} u_i v_i^T.$$

Just a few prior works

Low rank recovery via Nuclear Norm:
Fazel, Hindi, Boyd. A rank minimization heuristic with application to minimum order system approximation. 2001.
Candès, Recht. Exact matrix completion via convex optimization. 2009.
Candès, Li, Ma, Wright. Robust principal component analysis? 2011.
Non-convex approaches:
Mohan, Fazel. Iterative reweighted least squares for matrix rank minimization. 2010.
Pinghua, Zhang, Lu, Huang, Ye. A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems. 2013.
Sparse signal recovery using the ℓ_1 norm:
Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. 2006.
Candès, Romberg, Tao. Stable signal recovery from incomplete and inaccurate measurements. 2006.
Candès, Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? 2006.
Non-Convex approaches:
Candès, Wakin, Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. 2008.

Our Approach

Replace $\mu\text{rank}(X)$ with $\mathcal{R}_\mu(\boldsymbol{\sigma}(X)) = \sum_i \mu - max(\sqrt{\mu} - \sigma_i(X), 0)^2$.

$$\min_X \mathcal{R}_\mu(X) + \|\mathcal{A}X - b\|^2$$

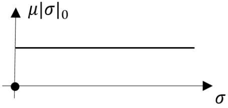
- $\mathcal{R}_\mu$ continuous, but non-convex.
- The global minimizer does not change if $\|\mathcal{A}\| < 1$.
- $\mathcal{R}_\mu(\boldsymbol{\sigma}(X)) + \|\mathcal{A}X - b\|^2$ lower bound on $\mu\text{rank}(X) + \|\mathcal{A}X - b\|^2$.

$f_\mu^{**}(X) = \mathcal{R}_\mu(\boldsymbol{\sigma}(X)) + \|X - X_0\|_F^2$ is the convex envelope of $f_\mu(X) = \mu\text{rank}(X) + \|X - X_0\|_F^2$.

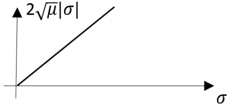
Larsson, Olsson. Convex Low Rank Regularization. IJCV 2016.

Shrinking Bias

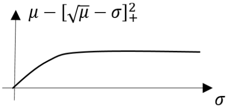
1D versions:



$\text{rank}(X) = \sum_i |\sigma_i(X)|_0$

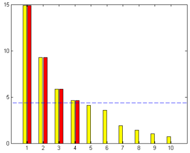


$\|X\|_* = \sum_i \sigma_i(X)$

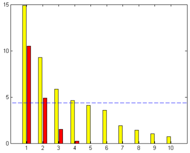


$\mathcal{R}_\mu(\boldsymbol{\sigma}(X)) = \sum_i \mu - [\sqrt{\mu} - \sigma_i(X)]_+^2$

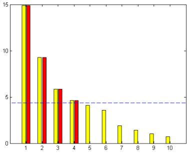
Singular value thresholding:



$\mu\text{rank}(X) + \|X - X_0\|_F^2$



$2\sqrt{\mu}\|X\|_* + \|X - X_0\|_F^2$



$\mathcal{R}_\mu(\boldsymbol{\sigma}(X)) + \|X - X_0\|_F^2$

More General Framework

Computation of the **convex envelopes**

$$f_g^{**}(X) = \mathcal{R}_g(\boldsymbol{\sigma}(X)) + \|X - X_0\|_F^2$$

of

$$f_g(X) = g(\text{rank}(X)) + \|X - X_0\|_F^2$$

where $g(k) = \sum_{i=1}^{k} g_i$ and $0 \leq g_1 \leq g_2 \leq ....$ And **proximal operators**.

Another special case:

$$f_{r_0}(X) = \mathbb{I}(\text{rank}(X) \leq r_0) + \|X - X_0\|_F^2$$

Larsson, Olsson. Convex Low Rank Regularization. IJCV 2016.

Results

General Case

If

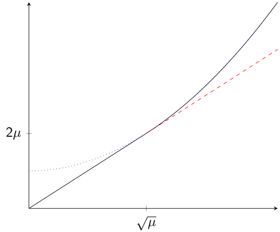
$$f_g(X) = g(\text{rank}(X)) + \|X - X_0\|_F^2$$

then

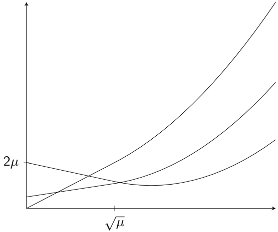
$$f_g^{**}(X) = \max_{\boldsymbol{\sigma}(Z)} \left(\sum_{i=1}^{n} \min\left(g_i,\ \sigma_i^2(Z)\right) - \|\boldsymbol{\sigma}(Z) - \boldsymbol{\sigma}(X)\|^2 \right) + \|X - X_0\|_F^2.$$

The maximization over Z reduces to a 1D-search.
(piece-wise quadratic concave objective function)
Can be done in $O(n)$ time ($n =$number of singular values).

Convexity of f_μ^{**}



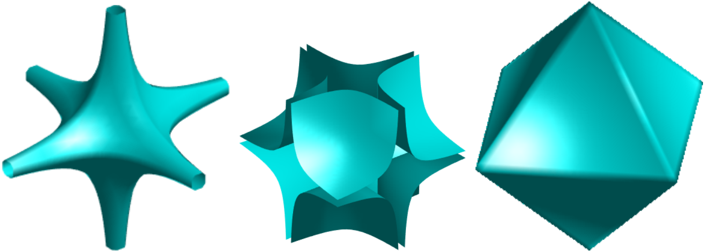
$\mu - \left[\sqrt{\mu} - \sigma\right]_+^2 + \sigma^2$



$\mu - \left[\sqrt{\mu} - \sigma\right]_+^2 + (\sigma - \sigma_0)^2$ for $\sigma_0 = 0, 1, 2$

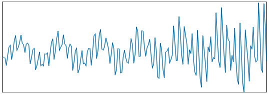
If $\sigma_0 = 2$ the function will not try to make $\sigma = 0$!

$$f_{r_0}(X) = \mathbb{I}(\text{rank}(X) \leq r_0) + \|X - X_0\|_F^2$$

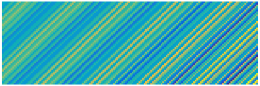


Level set surfaces $\{X \mid \mathcal{R}_{r_0}(X) = \alpha\}$ for $X = \text{diag}(x_1, x_2, x_3)$ with $r_0 = 1$ (*Left*) and $r_0 = 2$ (*Middle*). Note that when $r_0 = 1$ the regularizer promotes solutions where only one of x_k is non-zero. For $r_0 = 2$ the regularizer instead favors solutions with two non-zero x_k. For comparison we also include the level set of the nuclear norm.

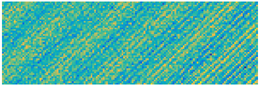
Signal



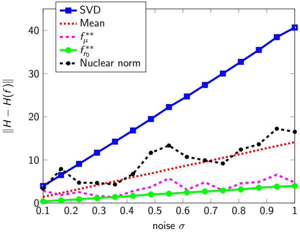
Hankel Matrix

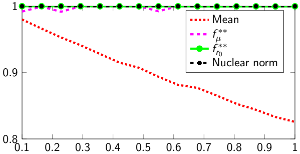


Matrix+Noise



$$\min_{H \in \mathcal{H}} \mathbb{I}(\mathrm{rank}(H) \leq r_0) + \|H - X_0\|_F^2$$

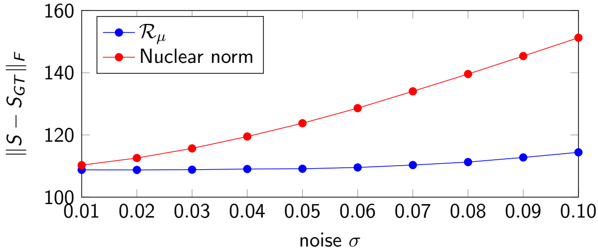


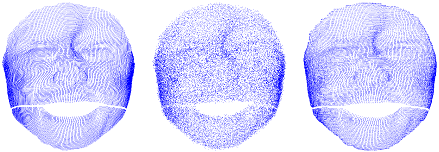


Smooth Linear Shape Basis

$$f_N(S) = \|S - S_0\|_F^2 + \mu\|P(S)\|_* + \tau \mathrm{TV}(S)$$

$$f_{\mathcal{R}}(S) = \|S - S_0\|_F^2 + \mathcal{R}_\mu(P(S)) + \tau \mathrm{TV}(S)$$





RIP problems

Linear observations: $b = \mathcal{A}X_0 + \epsilon$, $\mathcal{A} : \mathbb{R}^{m \times n} \to \mathbb{R}^p$, X_0 low rank, ϵ noise.

Recover X_0 using rank-penalties/constraints

$$\mathcal{R}_\mu(\boldsymbol{\sigma}(X)) + \|\mathcal{A}X - b\|^2 \quad \text{or} \quad \mathcal{R}_{r_0}(\boldsymbol{\sigma}(X)) + \|\mathcal{A}X - b\|^2$$

Restricted Isometry Property (RIP):

$$(1 - \delta_q)\|X\|_F^2 \leq \|\mathcal{A}X\|^2 \leq (1 + \delta_q)\|X\|_F^2, \quad \text{rank}(X) \leq q$$

Olsson, Carlsson, Andersson, Larsson. Non-Convex Rank/Sparsity Regularization and Local Minima. ICCV 2017.

Olsson, Carlsson, Bylow. A Non-Convex Relaxation for Fixed-Rank Approximation.

RSL-CV 2017.

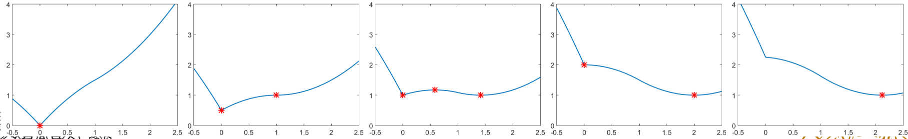
Near Convex Rank/Sparsity Estimation

Intuition:

- If RIP holds then $\|\mathcal{A}X\|^2$ behaves like $\|X\|_F^2$.
- $\mathcal{R}_\mu(\boldsymbol{\sigma}(X)) + \|X - X_0\|_F^2 \approx \mathcal{R}_\mu(\boldsymbol{\sigma}(X)) + \|X\|_F^2 - 2\langle X, X_0\rangle$ is convex.
- What about
  $\mathcal{R}_\mu(\boldsymbol{\sigma}(X)) + \|\mathcal{A}X - b\|^2 \approx \mathcal{R}_\mu(\boldsymbol{\sigma}(X)) + \|\mathcal{A}X\|^2 - 2\langle X, \mathcal{A}^* b\rangle$?
  Near convex?

1D-example: $\mathcal{R}_\mu(x) + (\frac{1}{\sqrt{2}}x - b)^2 \quad (\mu = 1)$

$b = 0$ $\qquad b = \frac{1}{\sqrt{2}}$ $\qquad b = 1$ $\qquad b = \sqrt{2}$ $\qquad b = 1.5$



Main Result (Rank Penalty)

Def. $F_\mu(X) := \mathcal{R}_\mu(\sigma(X)) + \|\mathcal{A}X - b\|^2$
$Z := (I - \mathcal{A}^*\mathcal{A})X_s + \mathcal{A}^* b$

X_s stationary point of $F_\mu(X) \Leftrightarrow$

$$X_s \in \arg\min_X \mathcal{R}_\mu(\sigma(X)) + \|X - Z\|_F^2.$$

$\|X - Z\|_F^2$ local approximation of $\|\mathcal{A}X - b\|^2$ around X_s.
X_s obtained by thresholding SVD of Z.

Theorem

If X_s is a stationary point of F_μ, and the singular values of Z fulfill $\sigma_i(Z) \notin [(1-\delta_r)\sqrt{\mu}, \frac{\sqrt{\mu}}{1-\delta_r}]$. then for any another stationary point X_s' we have $\text{rank}(X_s' - X_s) > r$.

Main Result (Rank Constraint)

Def. $F_{r_0}(X) := \mathcal{R}_{r_0}(\sigma(X)) + \|\mathcal{A}X - b\|^2$
$Z := (I - \mathcal{A}^*\mathcal{A})X_s + \mathcal{A}^* b$

X_s stationary point of $F_{r_0}(X) \Leftrightarrow$

$$X_s \in \arg\min_X \mathcal{R}_{r_0}(\sigma(X)) + \|X - Z\|_F^2.$$

$\|X - Z\|_F^2$ local approximation of $\|\mathcal{A}X - b\|^2$ around X_s.
X_s obtained by thresholding SVD of Z.

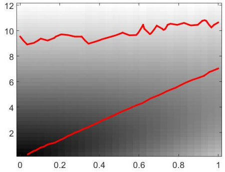
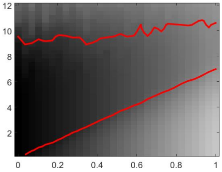
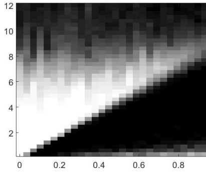
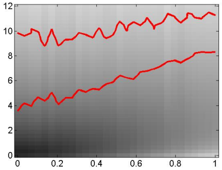
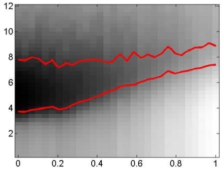
Theorem

If X_s is a stationary point of F_{r_0} with $\text{rank}(X_s) = r_0$, and the singular values of Z fulfill $\sigma_{r_0+1}(Z) < (1 - 2\delta_{2r_0})\sigma_{r_0}(Z)$ then any other stationary point X_s' has $\text{rank}(X_s') > r_0$.

Experiments (Rank Penalty)

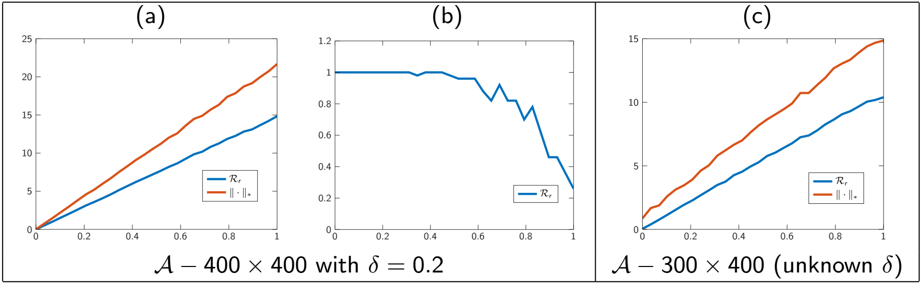
Low rank recovery for varying noise level (x-axis) and regularization strength (y-axis).

$$b = \mathcal{A}X_0 + \epsilon, \quad \epsilon_i \in \mathcal{N}(0,\sigma), \text{ where } \mathrm{rank}(X_0) = 10.$$

$\mathcal{A}$:	$2\sqrt{\mu}\|X\|_* + \|\mathcal{A}X - b\|_F^2$	$r_\mu(\boldsymbol{\sigma}(X)) + \|\mathcal{A}X - b\|_F^2$	Verified optima
$20^2 \times 400$ ($\delta = 0.2$)			
300×400 (δ unknown)			

Experiments (Fixed Rank)

$$\mu\|X\|_* + \|\mathcal{A}X - b\|_F^2 \text{ vs. } \mathcal{R}_r(\sigma(X)) + \|\mathcal{A}X - b\|_F^2$$



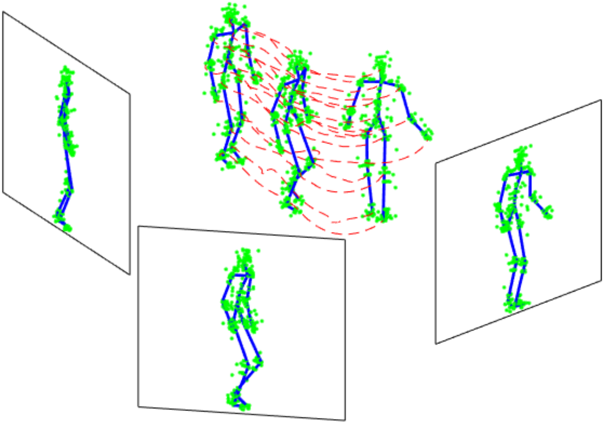
(a) - Noise level (x-axis) vs. data fit $\|\mathcal{A}X - \mathbf{b}\|$ (y-axis).
(b) - Fraction of instances verified to be globally optimal.
(c) - Same as (a).

$$X \in \mathbb{R}^{20\times 20} \Rightarrow vec(X) \in \mathbb{R}^{400}$$

(a) and (b) use 400 × 400 $\mathcal{A}$ with $\delta = 0.2$ while (c) uses 300 × 400 $\mathcal{A}$.

NRSfM

Reconstruct moving and deforming object from image projections.



CMU Motion capture sequences.

NRSfM (Dai et al. 2012)

X_i, Y_i, Z_i: x-,y- and z-coordinats in image i.
R_i: 2×3 matrix encoding orientation of camera i (affine model).

$$R = \begin{bmatrix} R_1 & 0 & \dots & 0 \\ 0 & R_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & R_F \end{bmatrix}, X = \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \\ \vdots \\ X_F \\ Y_F \\ Z_F \end{bmatrix} \text{ and } X^{\#} = \begin{bmatrix} X_1 & Y_1 & Z_1 \\ \vdots & \vdots & \vdots \\ X_F & Y_F & Z_F \end{bmatrix}.$$

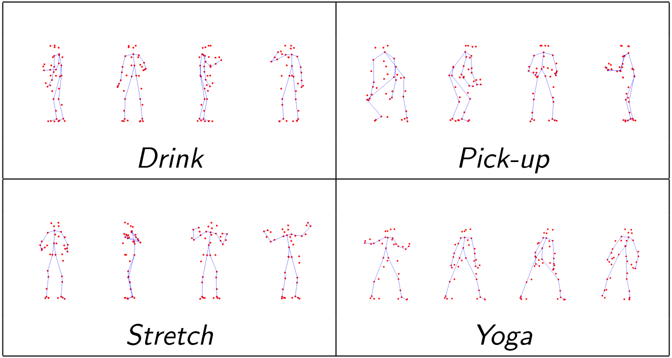
Projections: $M \approx RX$
Linear shape basis assumption: $\text{rank}(X^{\#}) \leq r$.

$$\min_{X^{\#}} \mathcal{I}(\text{rank}(X^{\#}) \leq r_0) + \|\mathcal{A}X^{\#} - M\|_F^2,$$

$$\mathcal{A} : X^{\#} \mapsto RX.$$

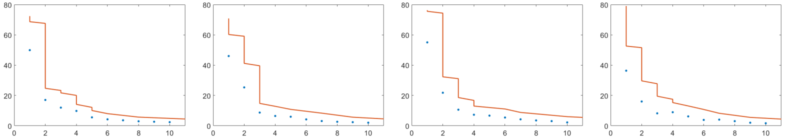
MOCAP Experiments

CMU Motion capture sequences:



MOCAP Experiments

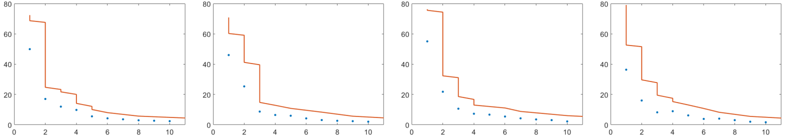
$$\mathcal{R}_{r_0}(X^{\#}) + \|RX - M\|_F^2 \text{ (blue) vs.}$$
$$\mu\|X^{\#}\|_* + \|RX - M\|_F^2 \text{ (orange)}$$



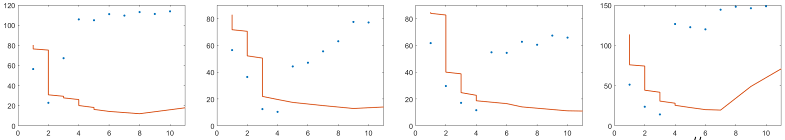
Data fit $\|RX - M\|_F$ (y-axis) versus rank$(X^{\#})$ (x-axis).

MOCAP Experiments

$$\mathcal{R}_{r_0}(X^{\#}) + \|RX - M\|_F^2 \text{ (blue) vs.}$$
$$\mu\|X^{\#}\|_* + \|RX - M\|_F^2 \text{ (orange)}$$



Data fit $\|RX - M\|_F$ (y-axis) versus rank$(X^{\#})$ (x-axis).



Distance to ground truth $\|X - X_{gt}\|_F$ (y-axis) versus rank$(X^{\#})$ (x-axis).

MOCAP Experiments

RIP does not hold for $\mathcal{A}(X^{\#}) = RX$!
If $R_i N_i = 0$, $N_i \in \mathbb{R}^{3\times 1}$ then $R_i N_i C_i = 0$, $\forall C_i \in \mathbb{R}^{1\times m}$.
Therefore $\mathcal{A}(N(C)) = 0$ for any matrix of the form

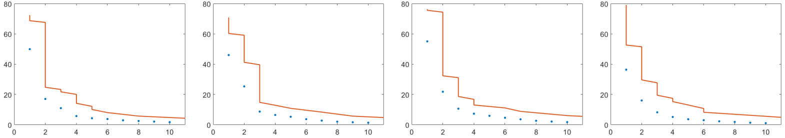
$$N(C) = \begin{bmatrix} n_{11}C_1 & n_{21}C_1 & n_{31}C_1 \\ n_{12}C_2 & n_{22}C_2 & n_{32}C_2 \\ \vdots & \vdots & \vdots \\ n_{1F}C_F & n_{2F}C_F & n_{3F}C_F \end{bmatrix},$$

where n_{ij} are the elements of N_i.

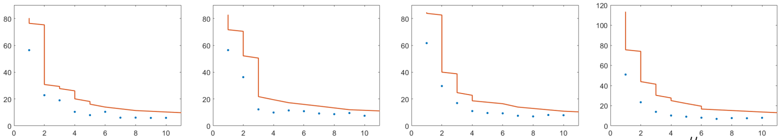
- $N(C)$ does not affect the projections.
- If the row space of an optimal solution contains $N(C)$ (for some C) the solution is not unique.

MOCAP Experiments

$$\mathcal{R}_{r_0}(X^\#) + \|RX - M\|_F^2 + \|DX^\#\|_F^2 \text{ (blue) vs.}$$
$$\mu\|X^\#\|_* + \|RX - M\|_F^2 + \|DX^\#\|_F^2 \text{ (orange)}$$



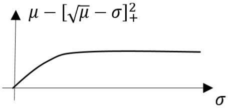
Data fit $\|RX - M\|_F$ (y-axis) versus rank($X^\#$) (x-axis).



Distance to ground truth $\|X - X_{gt}\|_F$ (y-axis) versus rank($X^\#$) (x-axis).

Optimization via ADMM/Splitting

- $\mathcal{R}_\mu(X)$ not differentiable X.
- $\mathcal{R}_\mu(X) + \|\mathcal{A}X - b\|^2$ (near) convex in X.
- Proximal operator $\arg\min_X \mathcal{R}_\mu(X) + \rho\|X - X_0\|_F^2$ computable.



Can apply splitting schemes:

$$L(X, Y, \Lambda) = \mathcal{R}_\mu(X) + \rho\|X - Y + \Lambda\|_F^2 + \|\mathcal{A}Y - b\|^2 - \rho\|\Lambda_i\|_F^2.$$

Alternate:

$$\begin{aligned} X_{t+1} &= \arg\min_X \mathcal{R}_\mu(X) + \rho\|X - Y_t + \Lambda_t\|_F^2, \\ Y_{t+1} &= \arg\min_Y \rho\|X_{t+1} - Y + \Lambda_t\|_F^2 + \|\mathcal{A}Y - b\|^2, \\ \Lambda_{t+1} &= \Lambda_t + X_{t+1} - Y_{t+1}. \end{aligned}$$

First order method.

Quadratic Approximation

Reformulate into differentiable objective. Bilinear parameterization:

$$\min_{B,C} \tilde{R}_\mu(B, C) + \|\mathcal{A}(BC^T) - b\|^2,$$

- $\tilde{R}_\mu(B, C)$ two times differentiable a.e.
- Optimize with 2nd order methods.
- Introduces non-optimal stationary points.

Characterization of local minima under RIP and optimization with VarPro/Wiberg.

Valtonen-Örnhag, Olsson, Heyden. Bilinear Parameterization for Differentiable Rank Regularization, arXiv 2018.

Bilinear Parameterization

Assumption: $\mathcal{R}(X) = \sum_{i=1}^{r} f(\sigma_i(X))$, f concave, non-decreasing on $[0, \infty)$.

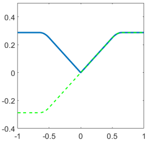
Theorem

$\mathcal{R}(X) = \min_{BC^T = X} \tilde{\mathcal{R}}(B, C)$, where

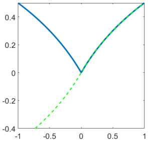
$$\tilde{\mathcal{R}}(B, C) = \sum_{i=1}^{k} f\left(\frac{\|B_i\|^2 + \|C_i\|^2}{2}\right),$$

and B_i,C_i columns of B,C.
$\tilde{R}(B, C)$ differentiable if $f(\sigma) = h(|\sigma|)$ where h is differentiable.

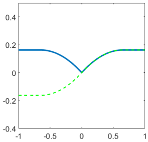
SCAD:



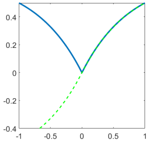
Log:



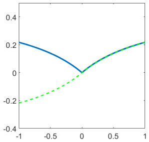
MCP:



ETP:



Geman:



Bilinear Parameterization

If $X = U\Sigma V^T$, $B = U\sqrt{\Sigma}$, $C = V\sqrt{\Sigma}$ then

$$B_i = \sqrt{\sigma_i} U_i$$

and

$$C_i = \sqrt{\sigma_i} V_i.$$

Therefore

$$\frac{\|B_i\|^2 + \|C_i\|^2}{2} = \frac{\sigma_i \|U_i\|^2 + \sigma_i \|V_i\|^2}{2} = \sigma_i.$$

Uniqueness of Low-Rank-Minima

Over parameterization with our relaxation:

Theorem

Let $(\bar{B}, \bar{C}) \in \mathbb{R}^{m\times 2k} \times \mathbb{R}^{n\times 2k}$ be a **local minimizer** of

$$\tilde{\mathcal{R}}_\mu(B, C) + \|\mathcal{A}(BC^T) - b\|^2,$$

with $\text{rank}(\bar{B}\bar{C}^T) < k$ and $\tilde{\mathcal{R}}_\mu(\bar{B}, \bar{C}) = \mathcal{R}_\mu(\bar{B}\bar{C}^T)$. If the singular values of $Z = (I - \mathcal{A}^*\mathcal{A})\bar{B}\bar{C}^T + \mathcal{A}^* b$ fulfill $\sigma_i(Z) \notin [(1-\delta_{2k})\sqrt{\mu}, \frac{\sqrt{\mu}}{(1-\delta_{2k})}]$ then

$$\bar{B}\bar{C}^T \in \underset{\text{rank}(X)\leq k}{\arg\min}\ \mathcal{R}_\mu(X) + \|\mathcal{A}X - b\|^2.$$

VarPro/Wiberg

$$\min_{B,C} \|\mathcal{A}(BC^T) - b\|^2$$

- Least Squares problem in C for fixed B. $\Rightarrow$ compute (closed form)

$$C^*(B) = \arg\min_{C} \|\mathcal{A}(BC^T) - b\|^2.$$

- Linearize $\mathcal{A}(BC^*(B)^T) - b \approx \mathcal{L}\delta B + \ell$ at B^k.
- Solve

$$\delta B^k = \arg\min \|\mathcal{L}\delta B + \ell\|^2 + \lambda\|\delta B\|^2$$
$$B^{k+1} = B^k + \delta B^k.$$

Quadratic approximation. Rapid convergence.
Similar to GN/LM.

Regweighted VarPro

$$\min_{B,C} \sum_i f\left(\frac{\|B_i\|^2 + \|C_i\|^2}{2}\right) + \|\mathcal{A}(BC^T) - b\|^2$$

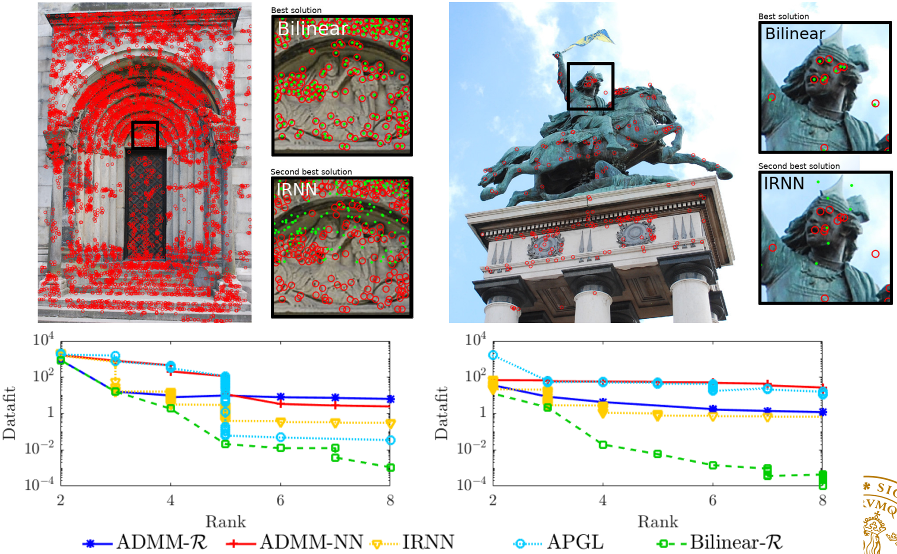
- Taylor: $f(x) \approx f(x^k) + f'(x^k)(x - x^k) = f'(x^k)x + \text{const}$

$$\min_{B,C} \sum_i w_i^k \left(\frac{\|B_i\|^2 + \|C_i\|^2}{2}\right) + \|\mathcal{A}(BC^T) - b\|^2,$$

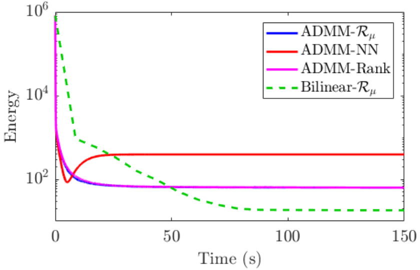
$w_i^k = f'\left(\frac{\|B_i^k\|^2 + \|C_i^k\|^2}{2}\right).$

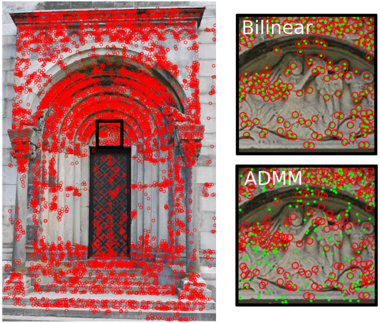
- One iteration of regular VarPro, recompute weights.
- Refactorize into B^{k+1}, C^{k+1} using SVD.

SfM-results


Best solution
Bilinear
Second best solution
IRNN
Best solution
Bilinear
Second best solution
IRNN
Datafit
Rank
ADMM-R
ADMM-NN
IRNN
APGL
Bilinear-R

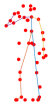
VarPro vs. ADMM

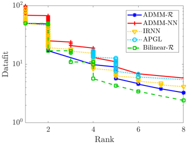




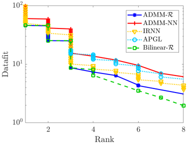
MOCAP Results

Drink

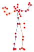


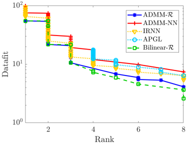


Pickup

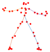


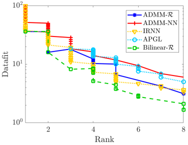
Stretch





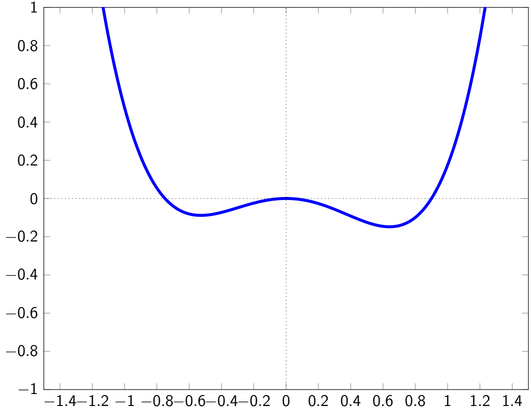
Yoga



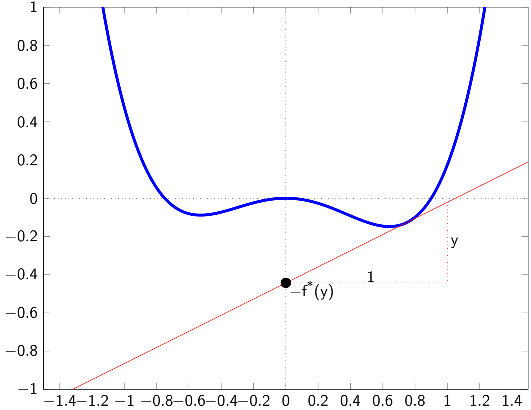


The End

Convex Envelopes and Conjugate Functions

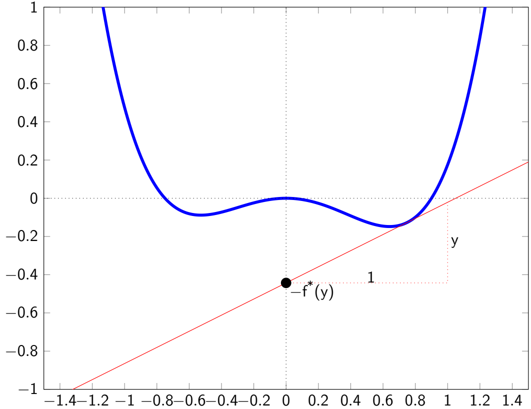


Convex Envelopes and Conjugate Functions



$$f^*(y) = \max_x yx - f(x)$$

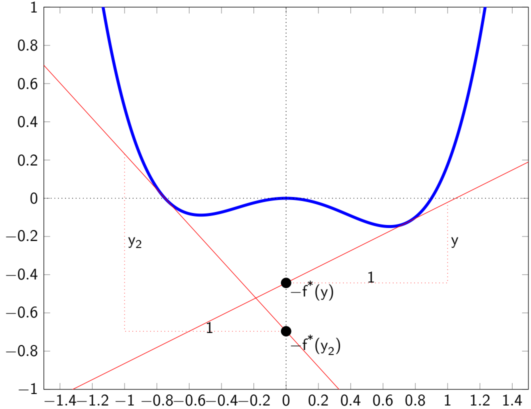
Convex Envelopes and Conjugate Functions



$$f^*(y) = \max_x yx - f(x) \Rightarrow f^*(y) \geq yx - f(x)$$

$$\Rightarrow f(x) \geq yx - f^*(y)$$

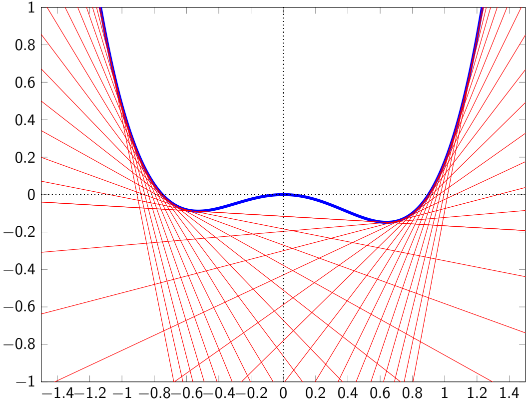
Convex Envelopes and Conjugate Functions



$$f^*(y) = \max_x yx - f(x) \Rightarrow f^*(y) \geq yx - f(x)$$

$$\Rightarrow f(x) \geq yx - f^*(y)$$

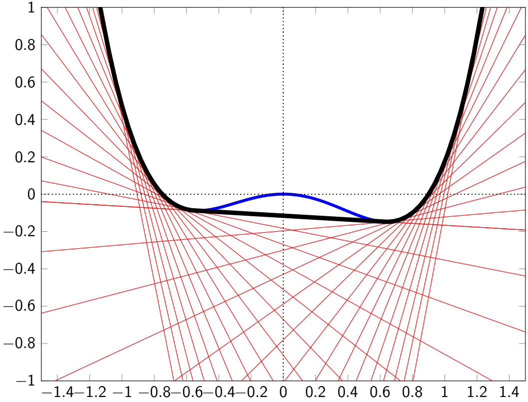
Convex Envelopes and Conjugate Functions



$$f^*(y) = \max_x yx - f(x) \Rightarrow f^*(y) \geq yx - f(x)$$

$$\Rightarrow f(x) \geq yx - f^*(y)$$

Convex Envelopes and Conjugate Functions



$$f^{**}(x) = \max_y xy - f^*(y)$$

Computing the Conjugate

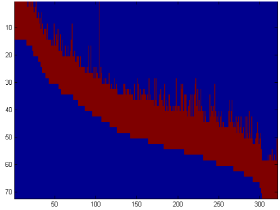
$$f_g^*(X) = \max_Y \langle X, Y \rangle - f_g(X)$$

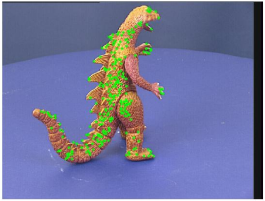
$$= \max_k - \sum_{i=1}^{k} g_k + \max_{\text{rank}(Y)=k} \langle X, Y \rangle - \|X - X_0\|_F^2$$

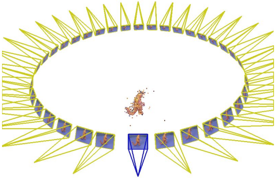
Inner maximization can be solved with SVD (Eckart, Young) and completion of squares.

The Missing Data Problem

$$\|W \odot (X - M)\|_F^2$$



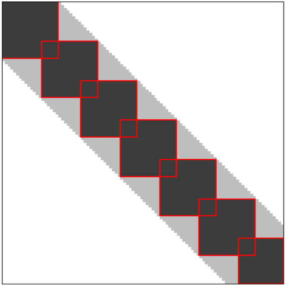




Red - visible element. Blue - missing measurement.
No known closed for solution.

A Block Decomposition Approach

Solve the problem on sub-blocks with no missing data:



$$f(X) = \sum_{i=1}^{K} \mu_i \text{rank}(\mathcal{P}_i(X)) + \|\mathcal{P}_i(X) - \mathcal{P}_i(M)\|_F^2$$

Convex relaxation:

$$\tilde{f}(X) = \sum_{i=1}^{K} \mathcal{R}_{\mu_i}(\boldsymbol{\sigma}(\mathcal{P}_i(X))) + \|\mathcal{P}_i(X) - \mathcal{P}_i(M)\|_F^2$$

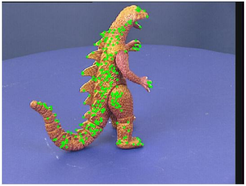
A Block Decomposition Approach

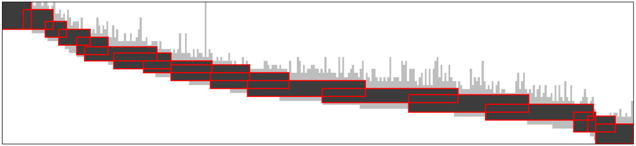
$$X = \begin{bmatrix} X_{11} & X_{12} & ? \\ X_{21} & X_{22} & X_{23} \\ ? & X_{32} & X_{33} \end{bmatrix}$$

Lemma

Let X_1 and X_2 be two given matrices with overlap matrix X_{22}, and let $r_1 = rank(X_1)$ and $r_2 = rank(X_2)$. Suppose that $rank(X_{22}) = \min(r_1, r_2)$, then there exists a matrix X with $rank(X) = \max(r_1, r_2)$. Additionally if $rank(X_{22}) = r_1 = r_2$ then X is unique.

Results





Observed: Our Solution: Nuclear Norm: Ground Truth: