Handling non-convexity in low-rank approaches for semidefinite programming

Nicolas Boumal

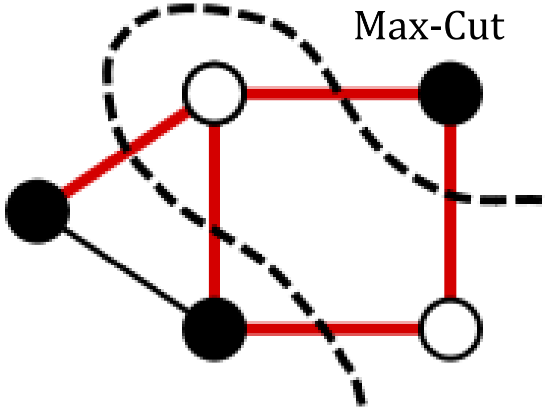
Department of Mathematics, Princeton University

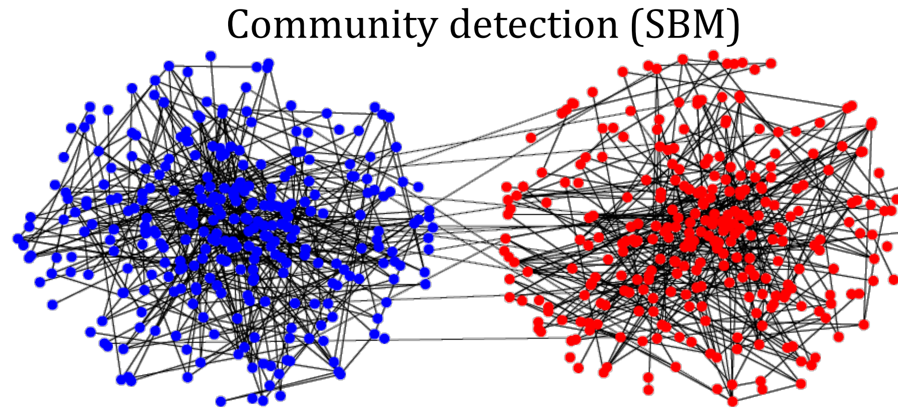
Collaborations with Afonso Bandeira and Vladislav Voroninski; Pierre-Antoine Absil, Naman Agarwal, Brian Bullins and Coralia Cartis; Srinadh Bhojanapalli, Prateek Jain and Praneeth Netrapalli; Samy Jelassi and Thomas Pumir.

This is about semidefinite programs

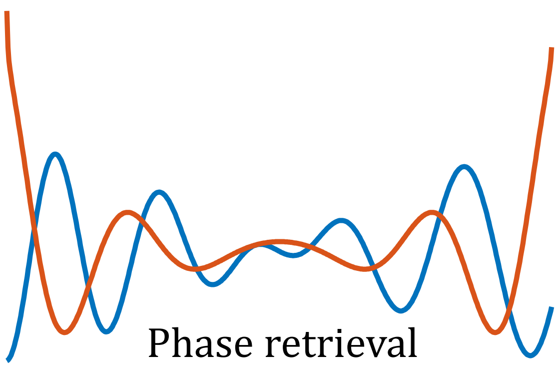
$$\min_{X: n\times n} \operatorname{Tr}(AX) \text{ s.t. } \operatorname{Lin}(X) = b, X \succcurlyeq 0$$

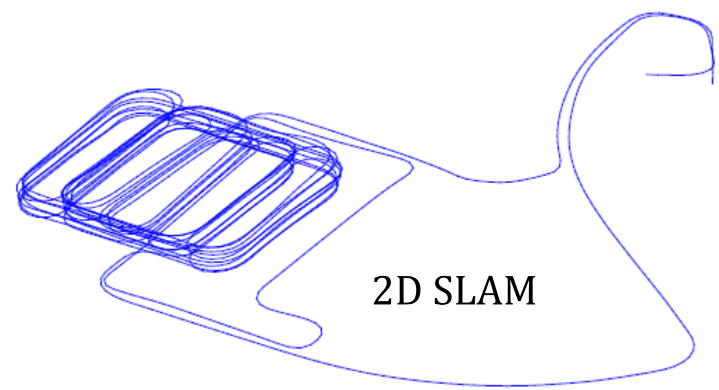
But not just “any” SDP.





$$\min_{X:n\times n} \mathrm{Tr}(AX) \text{ s.t. } \mathrm{diag}(X) = \mathbf{1}, X \succcurlyeq 0$$





Picture credits—Max-Cut: wikipedia; SBM: Abbe et al.?; SLAM: David Rosen's thesis

Convex, but in high dimension

Standard algorithms to solve SDPs iterate on full-rank, dense matrices X,

And need to maintain positive definiteness.

This requires memory and time.

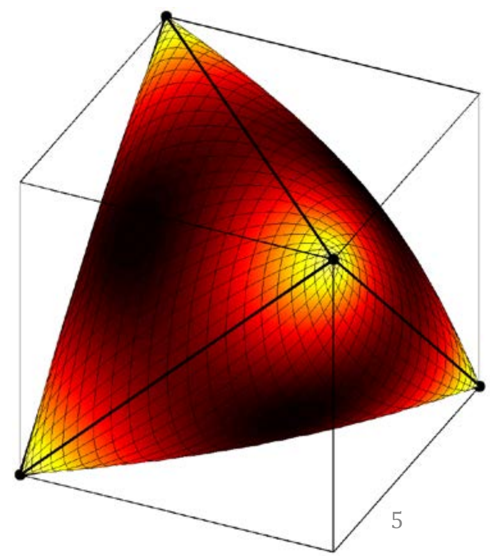
Max-Cut SDP has a low-rank solution

$$\min_X \mathrm{Tr}(AX) \ \text{ s.t. } \ \mathrm{diag}(X) = \mathbf{1}, X \succcurlyeq 0$$

Shapiro '82, Grone et al. '90, Pataki '94, Barvinok '95

There is an optimal X whose rank r satisfies

$$\frac{r(r+1)}{2} \leq n$$



A fortiori, $r \leq \sqrt{2n}$.

This justifies restricting the rank

$$\min_X \operatorname{Tr}(AX) \text{ s.t. } \operatorname{diag}(X) = \mathbf{1}, X \succcurlyeq 0, \operatorname{rank}(X) \leq p$$

Parameterize as $X = YY^T$ with Y of size $n \times p$:

$$\min_{Y:n\times p} \operatorname{Tr}\left(AYY^T\right) \text{ s.t. } \operatorname{diag}\left(YY^T\right) = \mathbf{1}$$

Lower dimension and no conic constraint!

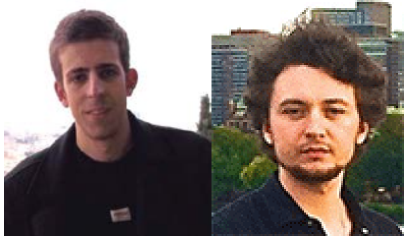
Burer & Monteiro ’03, ’05

But nonconvex…

Burer-Monteiro's early work ('03, '05)

> *"How large must we take p so that the local minima map to solutions of the SDP?*
>
> *Our theorem asserts we need only* $\frac{p(p+1)}{2} > n$
>
> *(with the important caveat that some faces of the SDP [...] can harbor non-global local minima)."*

With Bandeira & Voroninski, arXiv 1804.02008, to appear in CPAM



Our main result for this SDP

$$\min_{Y:n\times p} \mathrm{Tr}\left(AYY^T\right) \text{ s.t. } \mathrm{diag}\left(YY^T\right) = \mathbf{1}$$

If $\frac{p(p+1)}{2} > n$, for almost all A, all 2° points are optimal.

(If $p > n/2$, for all A, all 2° points are optimal.)

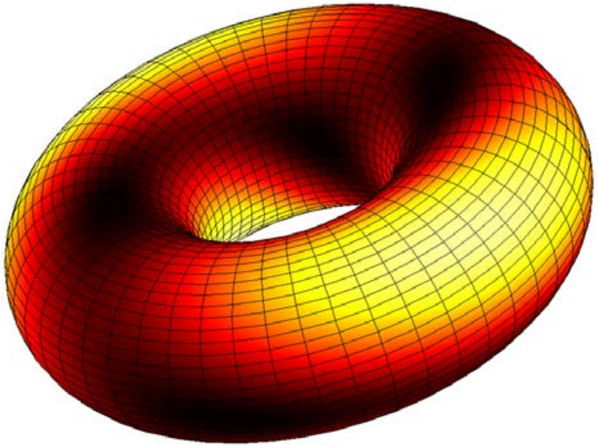
2° point: second-order critical point

The search space is smooth

$$\min_{Y:n\times p} \operatorname{Tr}\left(AYY^T\right) \text{ s.t. } \operatorname{diag}\left(YY^T\right) = \mathbf{1}$$

Constraints → rows of Y have unit norm.

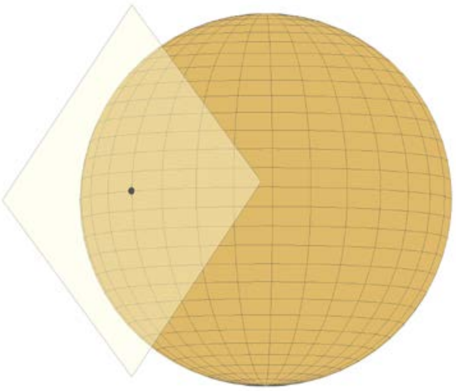
The search space is a product of spheres: smooth cost function on a smooth manifold.



Journée, Bach, Absil, Sepulchre '10

Necessary optimality conditions are straightforward on smooth spaces

$$\min_{Y:n\times p} \mathrm{Tr}\left(AYY^T\right) \text{ s.t. } \mathrm{diag}\left(YY^T\right) = \mathbf{1}$$



1° $\mathrm{Proj}_Y AY = SY = 0$, with

$S = A - \mathrm{ddiag}\left(AYY^T\right)$

2° $\langle \dot{Y}, S\dot{Y} \rangle \geq 0$ for all $\dot{Y}$ tangent:

$\mathrm{T}_Y M = \{\dot{Y} : \mathrm{diag}\left(\dot{Y}Y^T + Y\dot{Y}^T\right) = 0\}$

Main proof ingredients

1. $X = YY^T$ is optimal iff

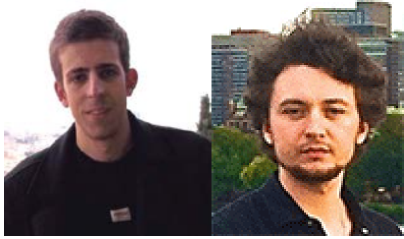
   For all feasible $\hat{X}$,
   $$0 \le \mathrm{Tr}\big(S\hat{X}\big) = \mathrm{Tr}\big(A\hat{X}\big) - \mathrm{Tr}\big(\mathrm{ddiag}(AX)\hat{X}\big) = \mathrm{Tr}\big(A\hat{X}\big) - \mathrm{Tr}(AX).$$

   $$S = A - \mathrm{ddiag}(AX) \succcurlyeq 0$$

2. If Y is a 2° point and rank deficient, $S \succcurlyeq 0$

3. If $\frac{p(p+1)}{2} > n$, for almost all A,
   all critical points Y are rank deficient.

With Bandeira & Voroninski, arXiv 1804.02008, to appear in CPAM



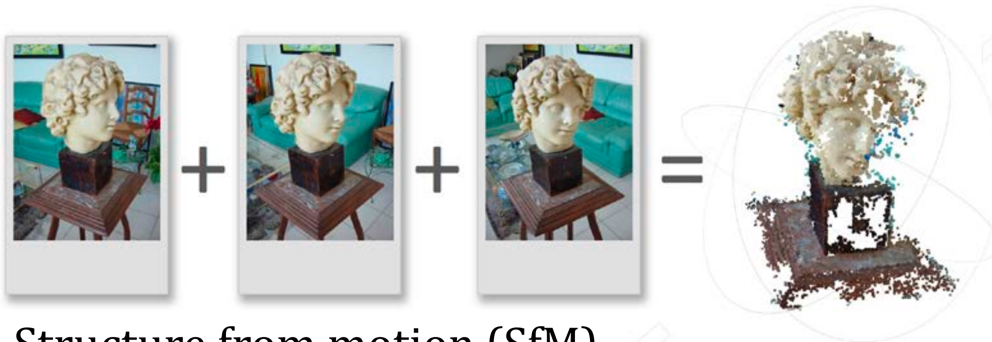
Main result for smooth SDPs

$$\min_{X:n\times n} \operatorname{Tr}(AX) \text{ s.t. } \operatorname{Lin}(X) = b, X \succcurlyeq 0$$

$$\min_{Y:n\times p} \operatorname{Tr}\left(AYY^T\right) \text{ s.t. } \operatorname{Lin}\left(YY^T\right) = b$$

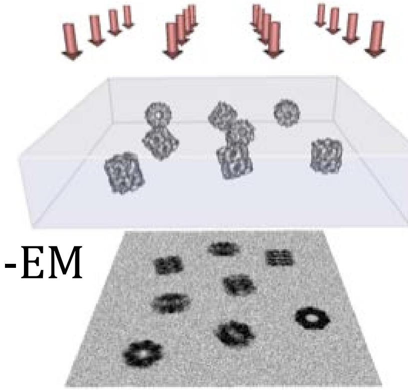
If the search space in X is compact
and the search space in Y is a manifold,
and if $\frac{p(p+1)}{2} > \#\text{constraints}$, then,
for almost all A, all 2° points are optimal.

$$\min_{X:n\times n} \mathrm{Tr}(AX) \text{ s.t. } X_{ii} = I_d, X \succcurlyeq 0$$



Structure from motion (SfM)

Cryo-EM



3D registration

3D SLAM

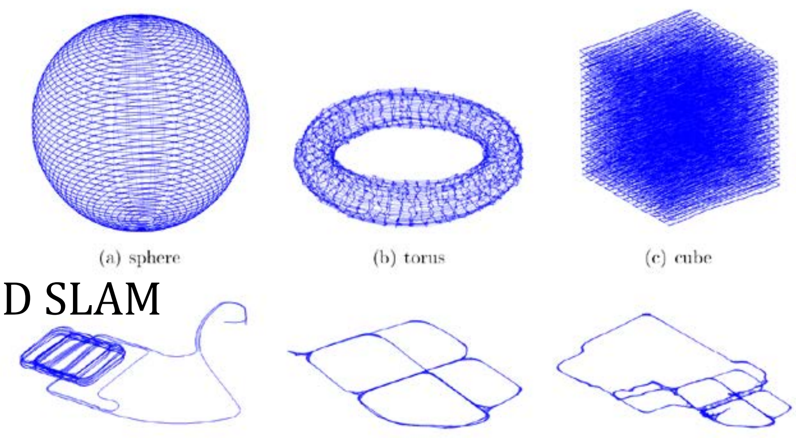


Figure 5-2: Visualizations of the optimal solutions recovered by SE-Sync for the large-scale 3D pose-graph SLAM datasets listed in Table 5.1. Note that (a)–(c) are synthetic, while (d)–(f) are real.

Picture credits—SfM: Princeton Vision & Robotics group; Cryo: ?; SLAM: David Rosen; Registration: Stanford 3D scanning repository

Approximate second-order points?

Our main result states exact 2° points are optimal for almost all A, for $\frac{p(p+1)}{2} > m$ (conditions apply).

In pratice, can only compute approximate 2° points.

Are they approximately optimal?
Well, there's the issue of "bad As"…

If bad As exist, need to be careful

A is bad if there exists a suboptimal 2° point Y.

Minuscule random perturbations of A kill all such points, almost surely…

But Y is still approximately 2° point,
while its suboptimality didn't change much…

If bad As exist,

Then, there exists a *non-zero measure* set of As

For which approximate 2° points are *not* all approximately optimal with $\frac{p(p+1)}{2} > m$.

How thick is this set?

Might it force us to take $p \gg \sqrt{m}$?

With Bhojanapalli, Jain & Netrapalli,
arXiv 1803.00186, COLT2018



Smoothed analysis

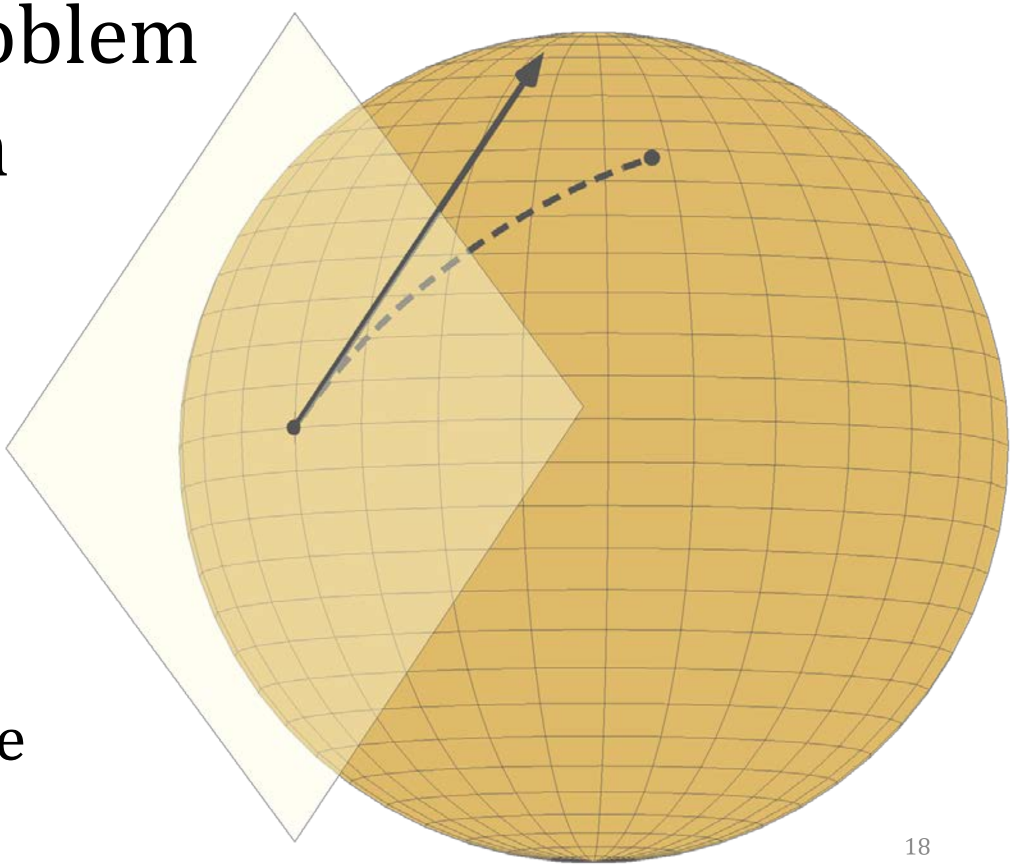
Deterministically, an approximate 2° point which is approximately rank deficient is approximately optimal.

With $p = \tilde{\Omega}(\sqrt{m})$, with high probability upon random perturbation of A, all approximate 1° points are approximately rank deficient.


With Jelassi & Pumir, arXiv 1806.03763, NIPS2018

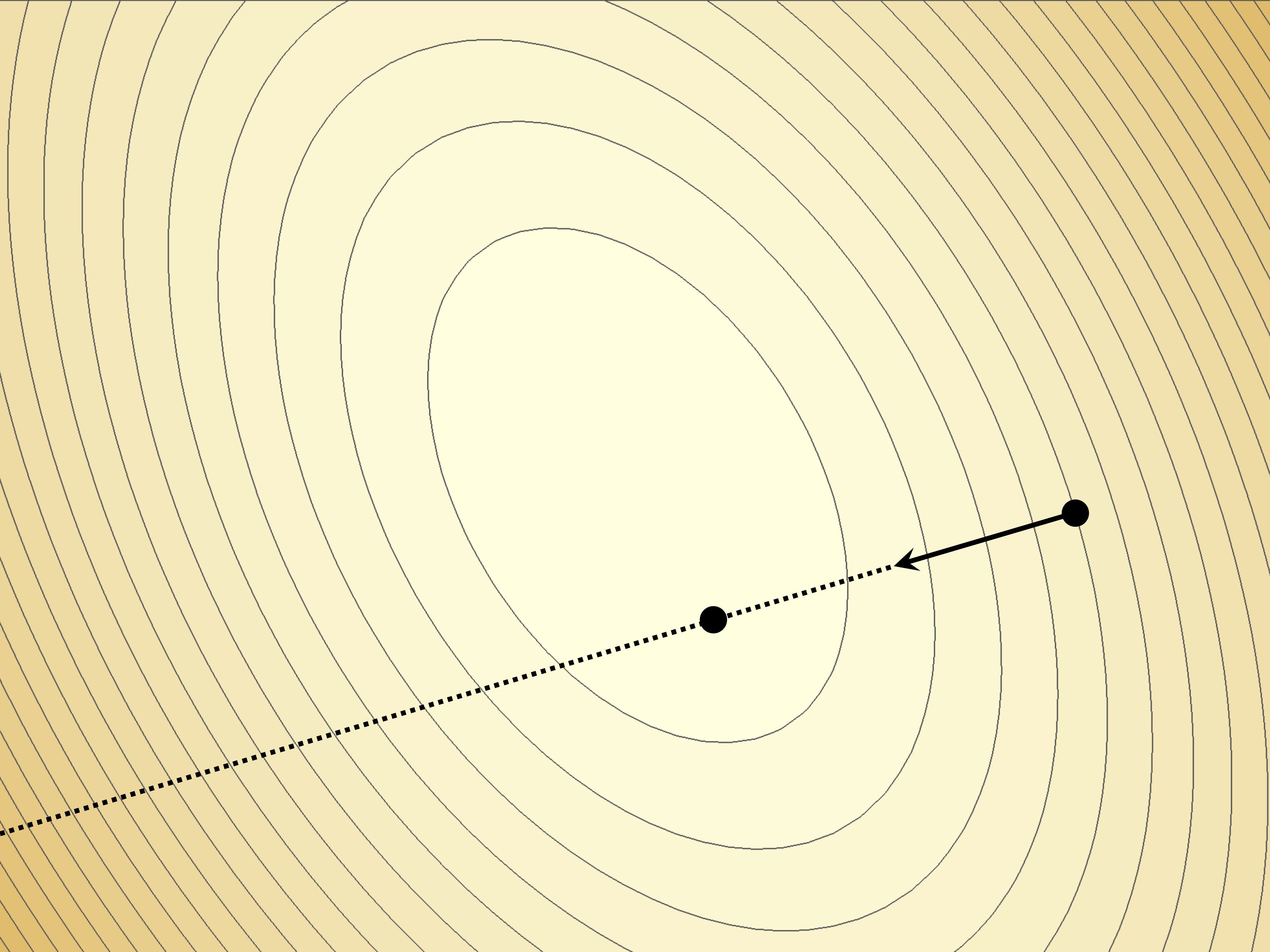
How to solve it?

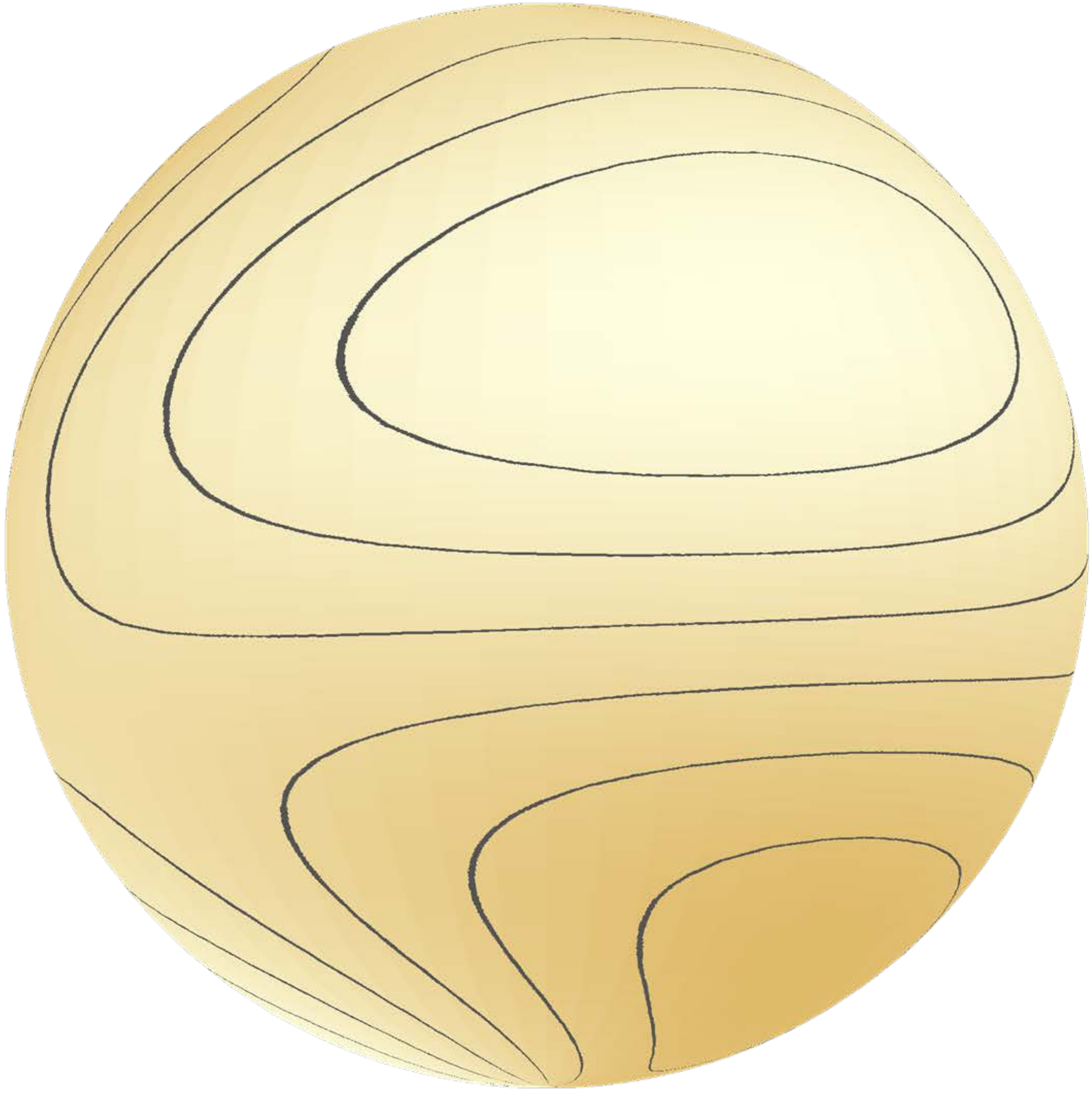
Under the smoothness assumption,
the non-convex problem
is optimization of a
smooth cost
over a
smooth manifold.

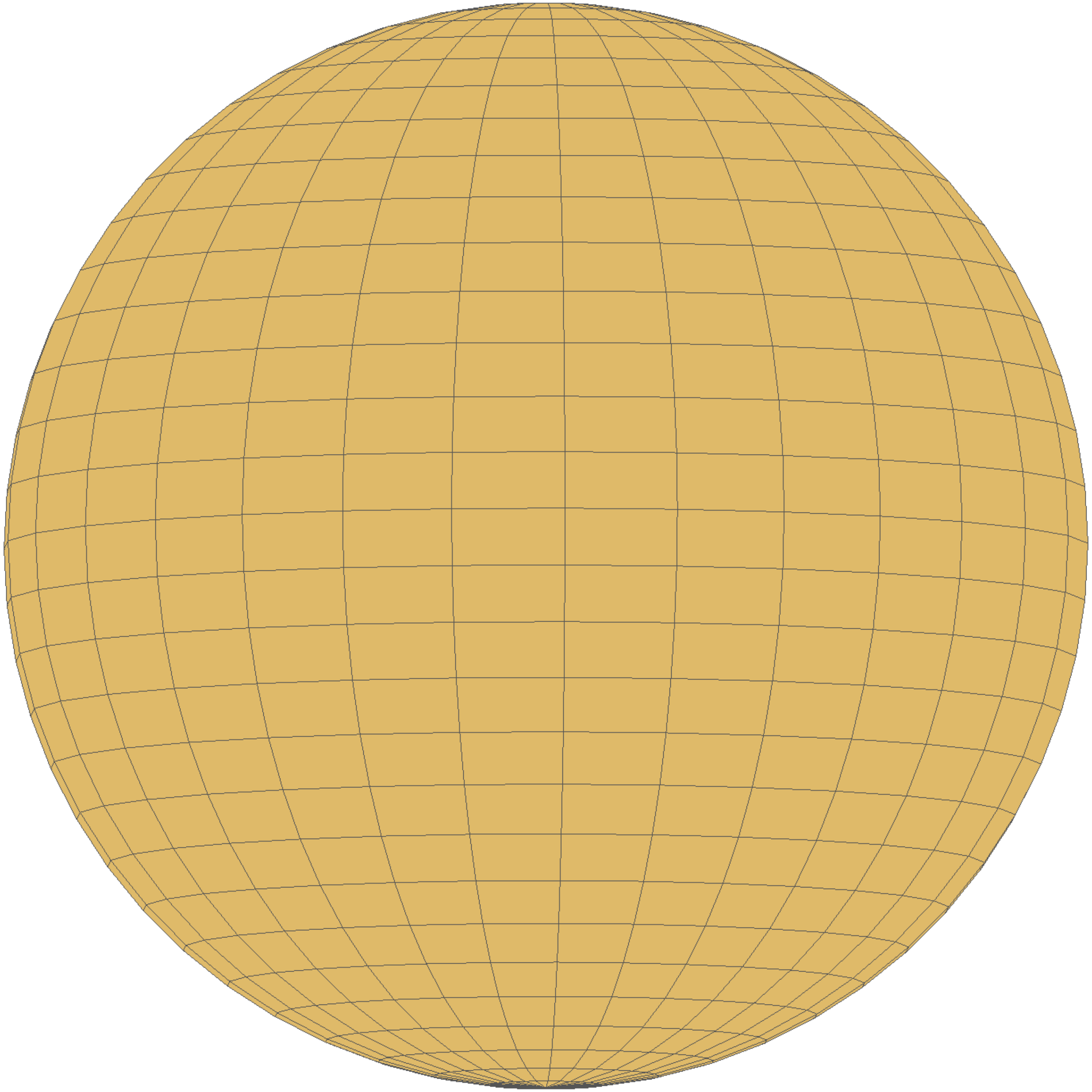


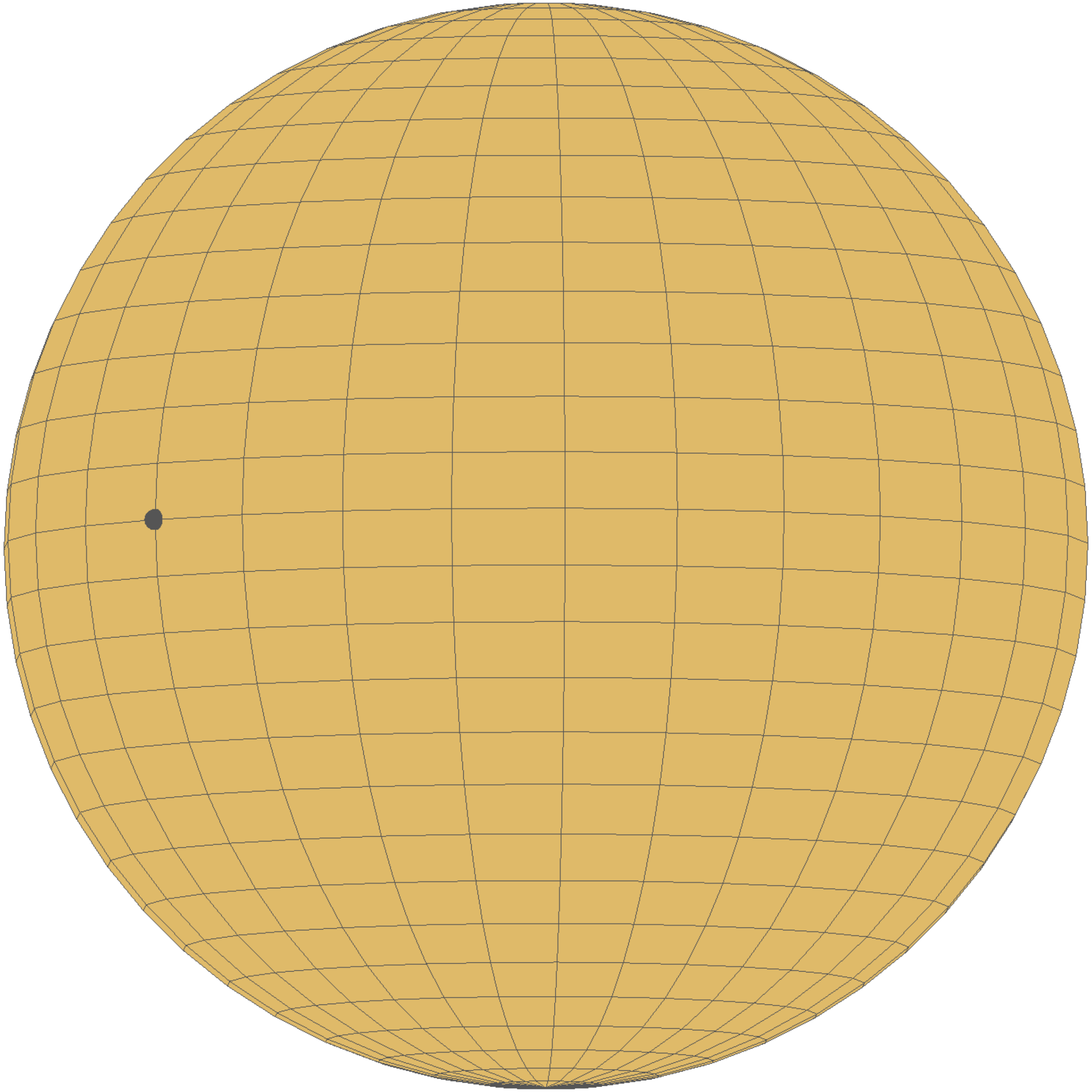
Ad: upcoming grad course

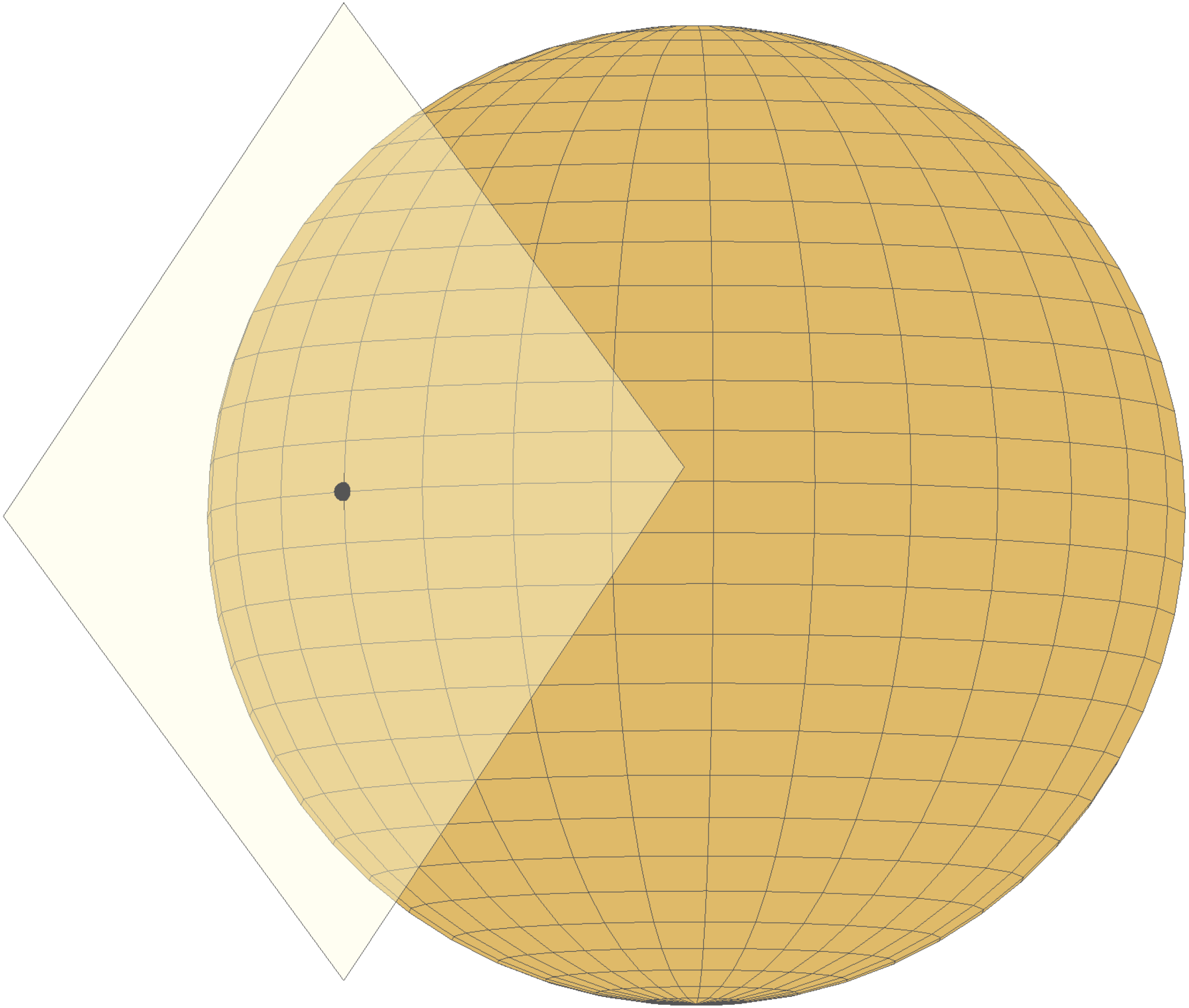
Taking a close look at gradient descent

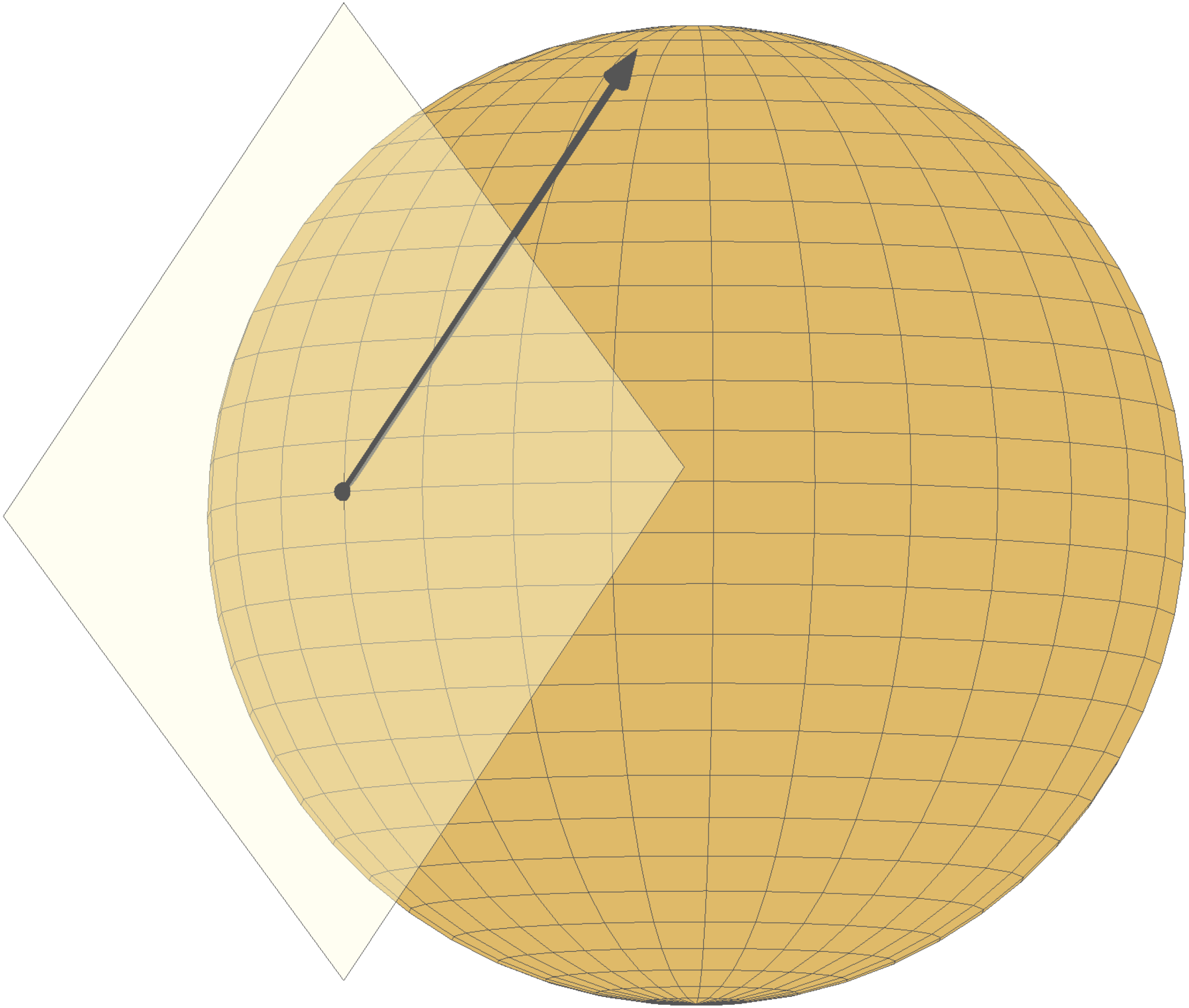


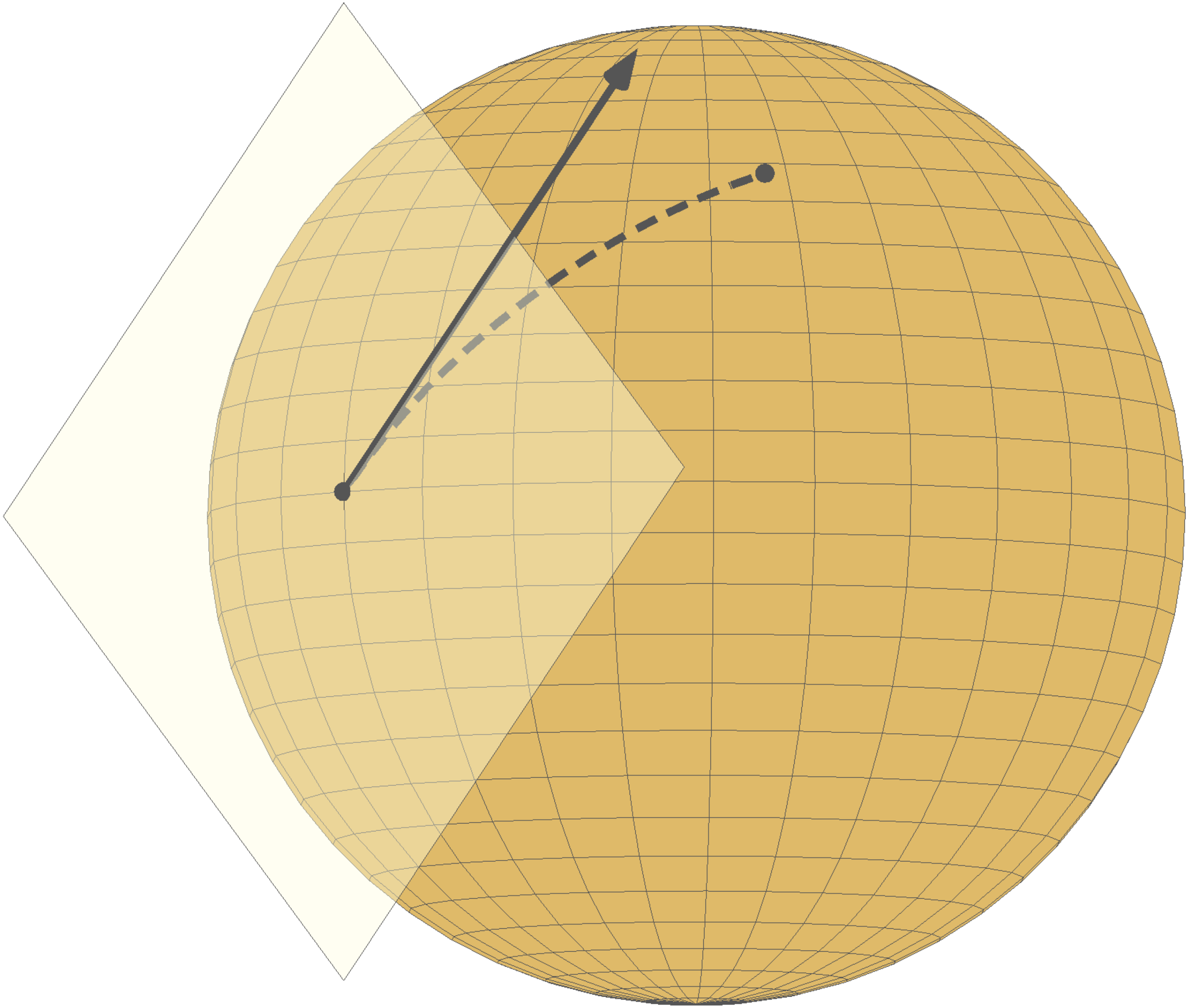






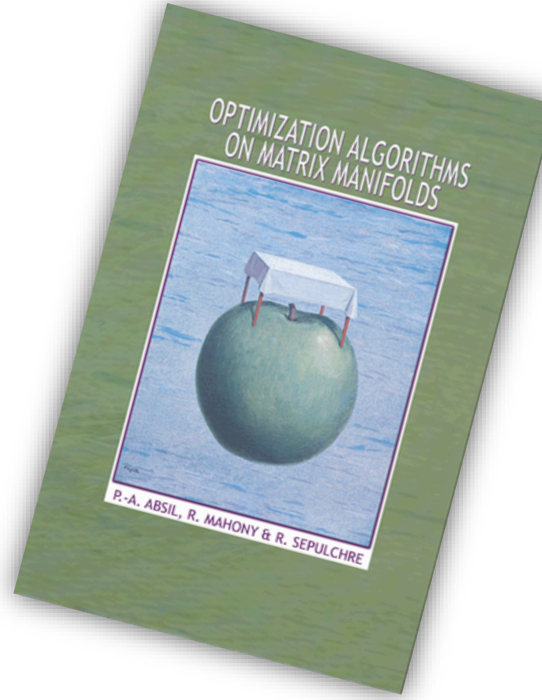






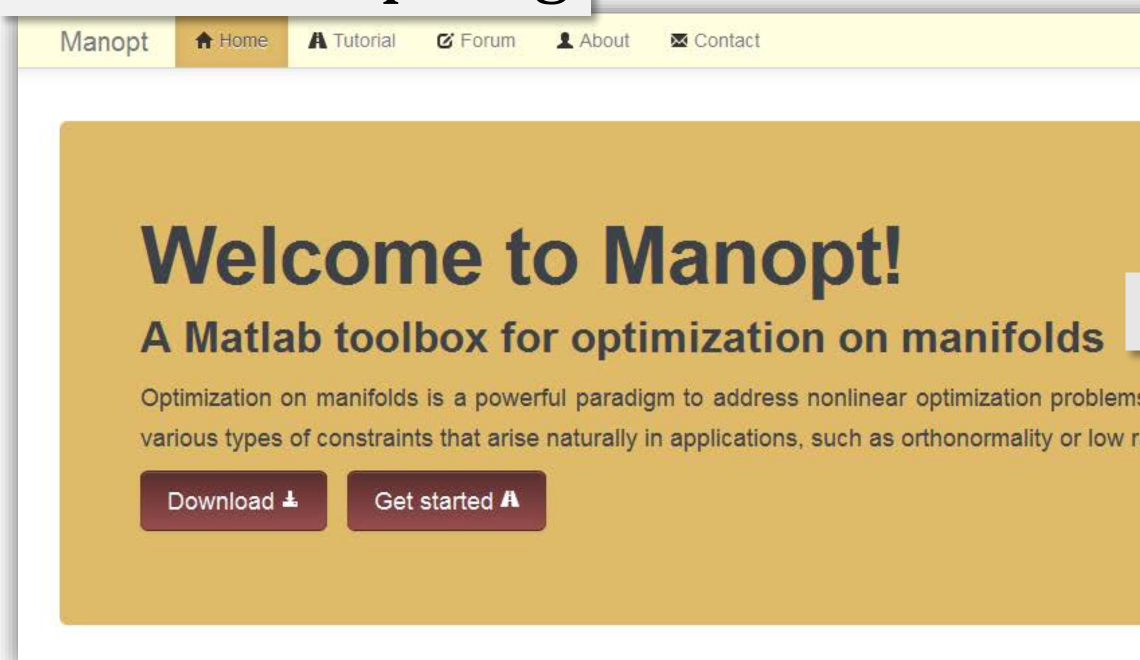
An excellent book

Optimization algorithms on matrix manifolds



A Matlab toolbox

www.manopt.org



With Mishra, Absil & Sepulchre

Iteration complexity for computing approximate 2° points on manifolds


Riemannian trust regions

arXiv:1605.08101, IMAJNA

$O(\varepsilon^{-2})$ for small gradient
$O(\varepsilon^{-3})$ for second-order too


Adaptive regularization with cubics (ARC)

arXiv:1806.00065

$O(\varepsilon^{-1.5})$ for small gradient
$O(\varepsilon^{-3})$ for second-order too

(Under Lipschitz conditions, satisfied on compact manifolds.)

Take home messages

The Burer-Monteiro low-rank approach works, generically, for smooth, compact SDPs.

Solve them with optimization on manifolds.

To ponder: can we relax the assumptions?