Applications

Goran Banjac

Large-Scale Convex Optimization
ETH Zurich

May 26, 2020

Optimal control

- consider the state-space representation of a discrete-time dynamical system

$$x_{k+1} = f(x_k, u_k)$$

- we want to drive the system to some desired state x_{ref} while respecting constraints on inputs u_k and states x_k
- we can compute the sequence of inputs u_k that solve the following optimal control problem

$$\begin{array}{ll}
\text{minimize} & l_N(x_N - x_{\text{ref}}) + \displaystyle\sum_{k=0}^{N-1} l_k(x_k - x_{\text{ref}}, u_k) \\
\text{subject to} & x_0 = x_{\text{init}} \\
& x_{k+1} = f(x_k, u_k), \quad k = 0, \ldots, N-1 \\
& x_k \in \mathcal{X},\ \ u_k \in \mathcal{U}, \quad\ k = 0, \ldots, N-1 \\
& x_N \in \mathcal{X}_N
\end{array}$$

Optimal spacecraft landing

https://youtu.be/2t15vP1PyoA

Optimal spacecraft landing

- we want to optimize the thrust profile of a spacecraft to carry out landing at a target position
- the spacecraft dynamics are

$$m\ddot{p} = f - mge_3$$

  where m is its mass, $p(t) \in \mathbb{R}^3$ the position, with 0 the target landing position, $f(t) \in \mathbb{R}^3$ the thrust force
- the thrust force is constrained by $\|f(t)\|_2 \leq F^{\mathsf{max}}$
- the spacecraft must remain in a region given by the glide slope constraint

$$p_3(t) \geq \alpha\|(p_1(t), p_2(t))\|_2$$

  second order conic constraint
- the fuel use rate is proportional to the thrust force magnitude, so the total fuel use is

$$\int_0^{T^{\mathsf{td}}} \gamma\,\|f(t)\|_2\,dt$$

Time discretization

- we discretize the thrust force in time, *i.e.*, it is constant over time period of length $h > 0$, where $T^{\text{td}} = Kh$
- the spacecraft dynamics then take the following form

$$\begin{aligned} v_{k+1} &= v_k + (h/m) f_k - hge_3 \\ p_{k+1} &= p_k + (h/2)(v_k + v_{k+1}) \end{aligned}$$

- for simplicity, we will impose the glide slope constraint only at the times $t = 0, h, 2h, \ldots, Kh$
- the total fuel use is then

$$\sum_{k=1}^{K} \gamma\, h \|f_k\|_2$$

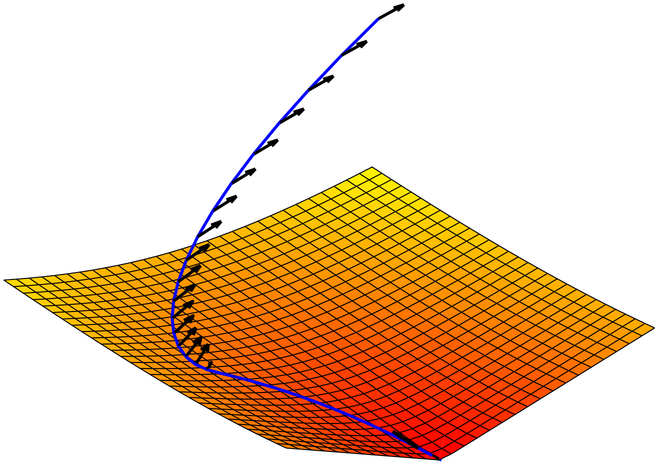
Minimum fuel descent

- we minimize fuel consumption, given the touchdown time $T^{\text{td}} = Kh$

$$
\begin{array}{ll}
\text{minimize} & \sum_{k=0}^{K-1} \|f_k\|_2 \\
\text{subject to} & p_0 = p(0),\ v_0 = \dot{p}(0) \\
& v_{k+1} = v_k + (h/m) f_k - hge_3 \\
& p_{k+1} = p_k + (h/2)(v_k + v_{k+1}) \\
& \|f_k\|_2 \leq F^{\text{max}}, \quad (p_k)_3 \geq \|((p_k)_1, (p_k)_2)\|_2 \\
& p_K = 0,\ v_K = 0
\end{array}
$$

- this is a convex optimization problem

Minimum fuel descent



Minimum time descent

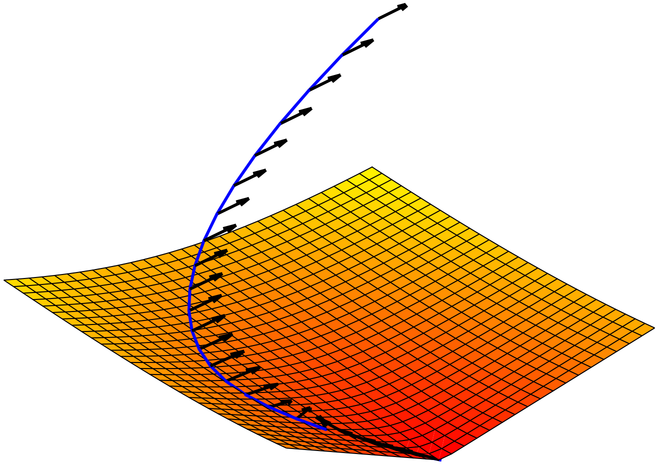
- to find the thrust profile that minimizes the touchdown time, we can solve a sequence of feasibility problems, *i.e.*, for each K we solve

$$
\begin{array}{ll}
\text{minimize} & 0 \\
\text{subject to} & p_0 = p(0),\ v_0 = \dot{p}(0) \\
& v_{k+1} = v_k + (h/m) f_k - hge_3 \\
& p_{k+1} = p_k + (h/2)(v_k + v_{k+1}) \\
& \|f_k\|_2 \leq F^{\mathsf{max}}, \quad (p_k)_3 \geq \|((p_k)_1, (p_k)_2)\|_2 \\
& p_K = 0,\ v_K = 0
\end{array}
$$

- if the problem is feasible, we reduce K, otherwise we increase K

can use bisection method, choose a large K, a smaller K, and then take the mid.

Minimum time descent



Portfolio allocation

- invest fraction w_i in asset i, $i = 1, \ldots, n$
- *portfolio allocation vector* $w \in \mathbb{R}^n$ satisfies $\mathbf{1}^T w = 1$
- initial prices $p_i > 0$; end of period prices $p_i^+ > 0$
- asset (fractional) returns $r_i = (p_i^+ - p_i)/p_i$
- portfolio (fractional) return $R = r^T w$
- common model: r is a random variable with mean $\mathbb{E}r = \mu$ and covariance $\mathbb{E}\left[(r-\mu)(r-\mu)^T\right] = \Sigma$
- therefore, R is a random variable with $\mathbb{E}R = \mu^T w$ and $\operatorname{var} R = w^T \Sigma w$
- $\mathbb{E}R$ is (mean) *return* of portfolio
- $\operatorname{var} R$ is *risk* of portfolio

- two objectives: high return, low risk

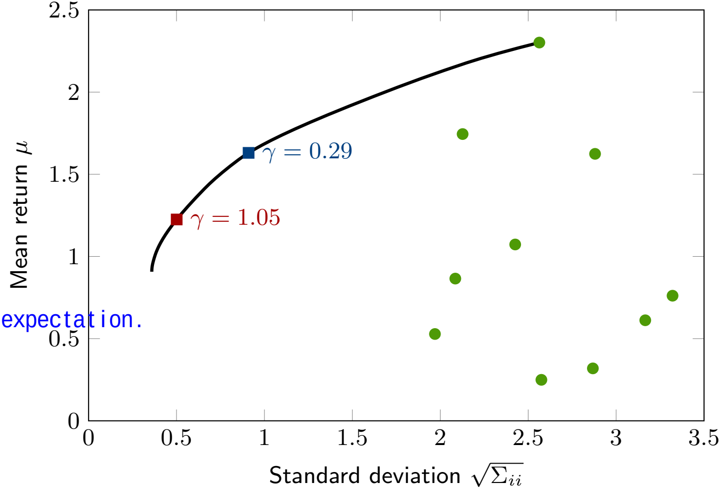
Markowitz portfolio optimization

$$
\begin{array}{ll}
\text{maximize} & \mu^T w - \gamma w^T \Sigma w \\
\text{subject to} & \mathbf{1}^T w = 1 \\
& w \in \mathcal{W}
\end{array}
$$

- $\mathcal{W}$ is set of allowed portfolios; common case: $\mathcal{W} = \mathbb{R}^n_+$
- $\gamma > 0$ is the *risk aversion parameter*
- $\mu^T w - \gamma w^T \Sigma w$ is *risk-adjusted return*
- varying γ gives optimal *risk-return trade-off*
- can also fix return and minimize risk, etc.

Numerical example

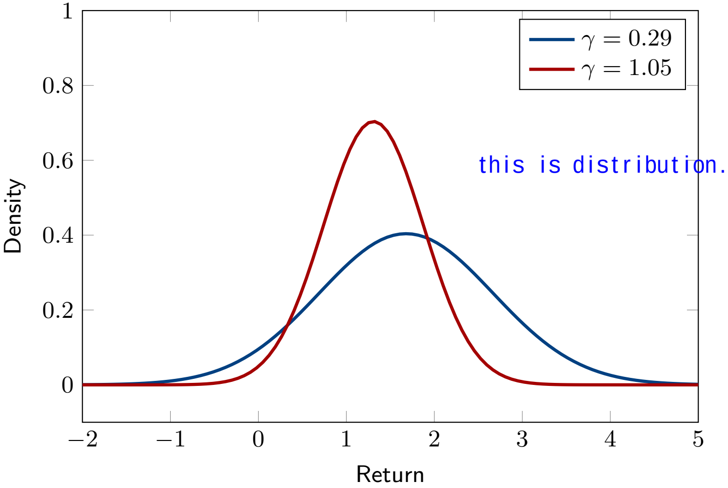
- optimal risk-return trade-off for 10 assets



this is expectation.

Numerical example

- return distributions for two risk aversion values



Matrix completion

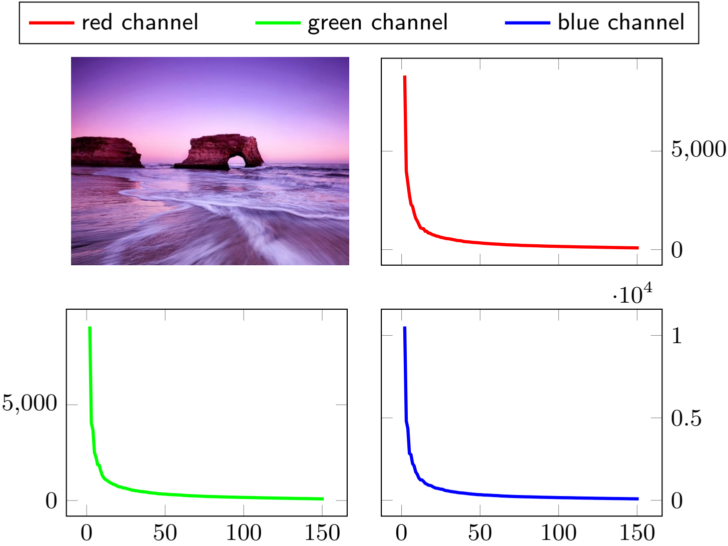
- consider the problem of estimating missing values of an unknown matrix

$$\begin{bmatrix} ? & \mathbf{4} & ? & \mathbf{2} & \mathbf{4} & ? & ? & ? & ? \\ \mathbf{3} & ? & \mathbf{3} & \mathbf{1} & ? & ? & \mathbf{3} & ? & \mathbf{3} \\ ? & \mathbf{3} & \mathbf{2} & \mathbf{3} & ? & ? & ? & ? & \mathbf{4} \\ ? & ? & \mathbf{2} & ? & ? & \mathbf{4} & ? & \mathbf{1} & \mathbf{2} \\ \mathbf{3} & ? & ? & \mathbf{3} & ? & ? & \mathbf{1} & ? & ? \end{bmatrix}$$

- it arises in many real applications such as image inpainting, video denoising, and recommender systems
- obviously, the problem is ill-posed
- to make the problem well-defined, a common assumption is that the matrix comes from a restricted class

assume image is low rank, because singular values are sometimes dominating , you can actually use a part of values to do computation, which implies low rank.

Singular values of an image (400×300)



Recovering a low-rank matrix

- in many applications, it is natural to assume that the unknown matrix has a low rank
- the matrix completion problem can thus be formulated as

$$\begin{array}{ll} \text{minimize} & \operatorname{rank}(X) \\ \text{subject to} & X_{ij} = M_{ij}, \quad (i,j) \in \Omega \end{array}$$

  where Ω is the set of locations corresponding to the observed entries
- unfortunately, the rank function is nonconvex
- in fact, the rank of a diagonal matrix corresponds to the cardinality of its diagonal

Nuclear norm regularization

- as the ℓ_1 norm is a convex approximation of the cardinality of a vector, the nuclear norm of a matrix approximates its rank
- therefore, we approximate the matrix completion problem by

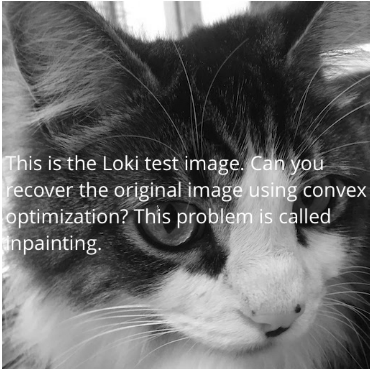
$$\begin{array}{ll} \text{minimize} & \|X\|_* \\ \text{subject to} & X_{ij} = M_{ij}, \quad (i,j) \in \Omega \end{array}$$

  where $\|X\|_* = \sum_{i=1}^{\min(m,n)} \sigma_i(X)$ is the nuclear norm of $X \in \mathbb{R}^{m \times n}$
- the proximal operator of $\|\cdot\|_*$ can be computed efficiently
- can also penalize the distance to observed entries instead of enforcing them by constraints

for diagonal matrix, singular value are absolute of eigen values.just like one-norm in Lasso we used.

Image reconstruction

- example: text over image



References

- these lecture notes are based to a large extent on those for the Stanford EE364a class developed by Stephen Boyd
- the original slides can be downloaded from

  `https://web.stanford.edu/class/ee364a/lectures.html`

  `https://web.stanford.edu/~boyd/papers/cvx_short_course.html`