Lecture 11

- **Introduction to convex optimization**
- ○ Convex optimization modeling
- ○ A model predictive control example
- ○ Duality

Optimization

- Consider the following optimization problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in X \end{array}$$

- Most probably this problem cannot be solved
- However, if f and X are convex, we can, even very large problems

Convex optimization

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in X \end{array}$$

- *objective* f is convex:

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$$

  if $0 \leq \alpha \leq 1$

- *constraint set* X is convex:

$$\alpha x + (1 - \alpha)y \in X$$

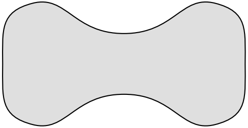
  for every $x, y \in X$ and $\alpha \in [0, 1]$

Convex sets

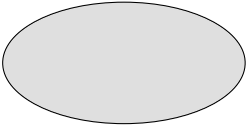
- a set X is convex if for every $x, y \in X$ and $\theta \in [0, 1]$:

$$\theta x + (1 - \theta) y \in X$$

- "every line segment that connect any two points in X is in X"

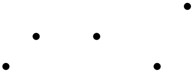


A nonconvex set



A convex set

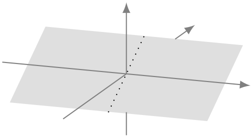
A nonconvex set



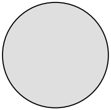
A nonconvex set

Examples of convex sets

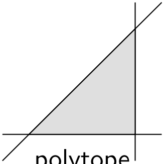
- affine subspaces $\{x \mid Ax = b\}$
- norm balls $\{x \mid \|x\| \leq r\}$ for some $r > 0$
- polytopic sets $\{x \mid Cx \leq d\}$
- level sets $\{x \mid g(x) \leq 0\}$ of convex functions g



affine subspace



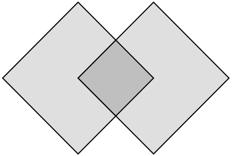
norm ball



polytope

Intersection and union

- the intersection $X_1 \cap X_2$ of two convex sets X_1, X_2 is convex
- the union $X_1 \cup X_2$ of two convex sets X_1, X_2 need not be convex



(intersection: darker gray, union: lighter gray)

- Example: Want $x \in X_1$ and $x \in X_2 \Rightarrow$ want $x \in X_1 \cap X_2$

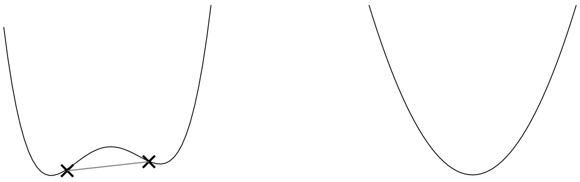
Convex functions

- Convex functions satisfy Jensen's inequality, i.e.,:

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y)$$

  for all x, y and $0 \leq \alpha \leq 1$

- "the whole line between any two points on the graph of f is on or above the graph"



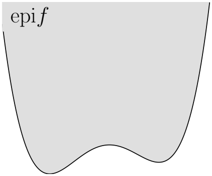
a nonconvex function — a convex function

Epigraphs and convexity

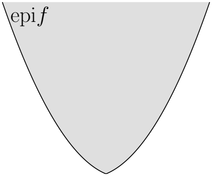
- the epi-graph of f is the set above f, i.e., the set

$$\text{epi } f = \{(x, r) \mid f(x) \leq r\}$$

- result: the function f is convex if and only epi f is a convex set



nonconvex



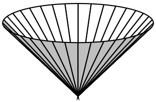
convex

Examples of convex functions

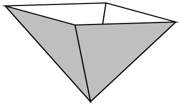
- indicator functions of convex sets X

$$\iota_{\mathcal{S}}(x) = \begin{cases} 0 & \text{if } x \in X \\ \infty & \text{else} \end{cases}$$

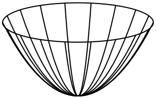
- norms: $\|x\|$ (e.g., 1-norm $\|x\|_1$ or 2-norm $\|x\|_2$)
- norm-squared: $\|x\|^2$
- (shortest) distance to convex set: $\text{dist}_X(y) = \inf_{x \in X}\{\|x - y\|\}$
- linear functions: $f(x) = q^T x$
- quadratic forms: $f(x) = \frac{1}{2}x^T Q x$ with Q positive semi-definite
- compositions of convex f with affine operator: $f(Lx - b)$



$\|x\|_2$



$\|x\|_1$



$\|x\|_2^2$

Linear programming

- A special case of convex optimization is linear programming
- This is obtained if
  - the objective f is linear
  - the constraint set $X = X_1 \cap X_2$ with X_1 affine and X_2 polytope
- The linear program is:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & Cx \leq d \end{array}$$

Convex optimization

- Many problems can be modeled using convex optimization
- In many cases we can achieve state-of-the-art performance
- We will look at a couple of examples
  - Signal reconstruction
  - Image reconstruction
  - Model predictive control of process industry site

Lecture 11

- ◦ Introduction to convex optimization
- • **Convex optimization modeling**
- ◦ A model predictive control example
- ◦ Duality

Optimization modeling

- will show how to model some problems using optimization
- in most examples, we use the following functions

$$\|\cdot\|_1, \quad \tfrac{1}{2}\|\cdot\|_2^2$$

  and often compose them with an affine operator $Lx - b$
- we will also use the following constraint sets

$$\{x \mid Ax = b\}, \quad \{x \mid Cx \leq d\}$$

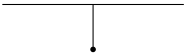
- constructing convex optimization problem using these functions only is very powerful and can model a wide variety of problems!

0-norm

- in many applications we would ideally like to use the 0-norm
- the 0-norm $\|x\|_0$ counts the number of nonzero elements in x
- that is $\|x\|_0 = \sum_i h_i(x_i)$ where

$$h_i(x_i) = \begin{cases} 0 & \text{if } x_i = 0 \\ 1 & \text{else} \end{cases}$$

- graphical representation



- it is obviously nonconvex
- often the 1-norm is used as a convex proxy for this
- why? 1-norm is convex envelope of $\|x\|_0 + \iota_{\|x\| \leq 1}(x)$ for $\|x\| \leq 1$

Signal reconstruction

- in signal reconstruction, we have a noisy signal y
- assume that measurement from process with slow changes
- approximate with signal x that captures process behavior
- therefore: want neighboring time-steps to be close to each other
- we have two competing objectives, want $x \approx y$ and x vary slowly

Signal reconstruction

- introduce difference operator D

$$D = \begin{bmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{bmatrix}$$

- then

$$Dx = \begin{bmatrix} x_1 - x_2 \\ \vdots \\ x_{n-1} - x_n \end{bmatrix}$$

- want Dx small and $x \approx y$
- can you model this as an optimization problem?

Signal reconstruction

- introduce difference operator D

$$D = \begin{bmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{bmatrix}$$

- then

$$Dx = \begin{bmatrix} x_1 - x_2 \\ \vdots \\ x_{n-1} - x_n \end{bmatrix}$$

- want Dx small and $x \approx y$
- can you model this as an optimization problem?
- consider the optimization problem

$$\text{minimize } \tfrac{1}{2}\|x - y\|_2^2 + \lambda\|Dx\|_2^2$$

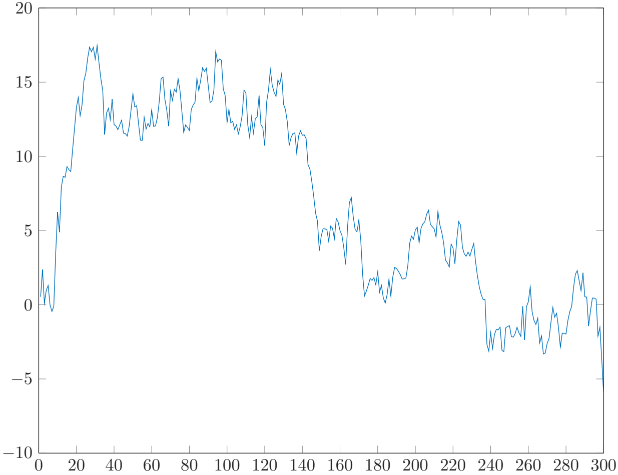
where y contains measurements and $\lambda > 0$ trades off objectives

Numerical example

- we have $y \in \mathbb{R}^{300}$
- y constructed by random walk in $\mathbb{R}$

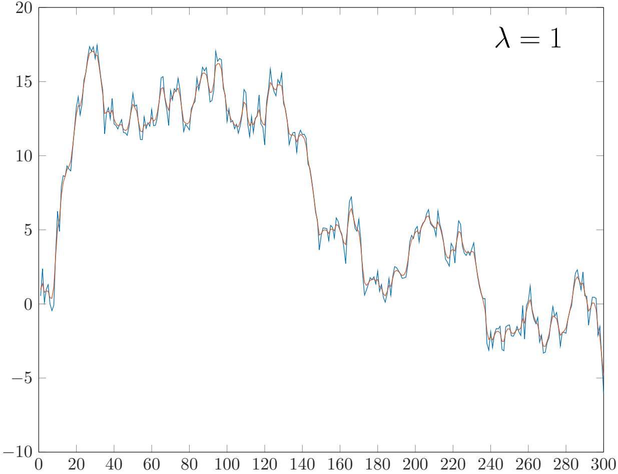
Result

$$\text{minimize } \frac{1}{2}\|x - y\|_2^2 + \lambda\|Dx\|_2^2$$



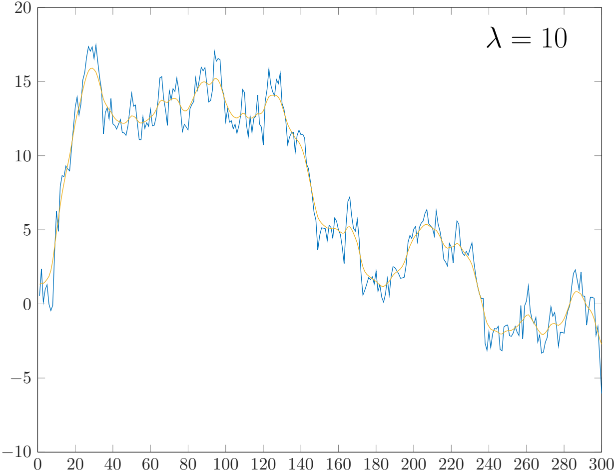
Result

$$\text{minimize } \frac{1}{2}\|x - y\|_2^2 + \lambda\|Dx\|_2^2$$



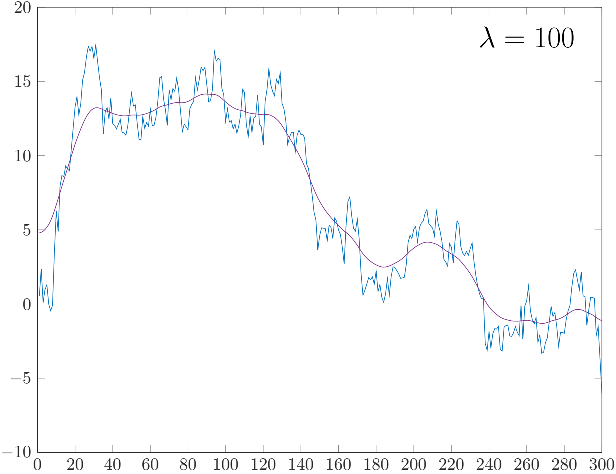
Result

$$\text{minimize } \frac{1}{2}\|x - y\|_2^2 + \lambda\|Dx\|_2^2$$



Result

$$\text{minimize } \frac{1}{2}\|x - y\|_2^2 + \lambda\|Dx\|_2^2$$



Example

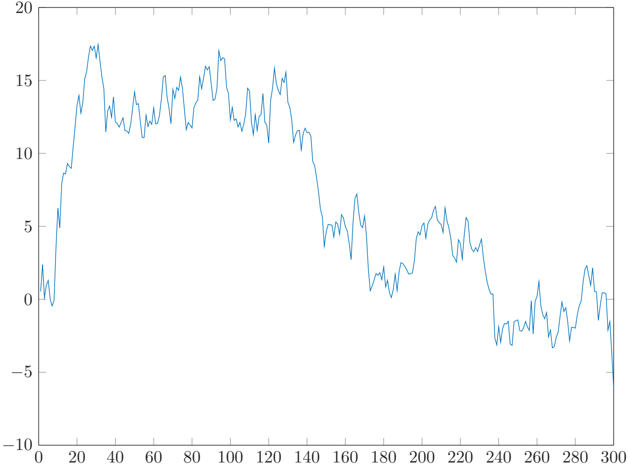
- what if we instead want piece-wise constant approximation?
- then we want Dx to be sparse
- how to model this?

Example

- what if we instead want piece-wise constant approximation?
- then we want Dx to be sparse
- how to model this?
- typically we want to minimize $\|Dx\|_0$
- nonconvex, use our convex proxy $\|Dx\|_1$

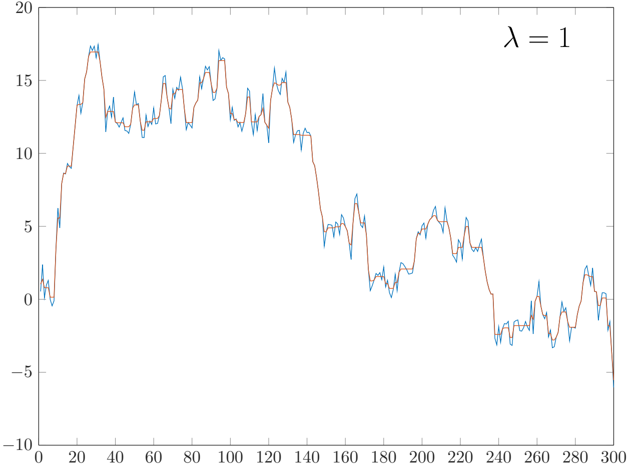
Result

$$\text{minimize } \frac{1}{2}\|x-y\|_2^2 + \lambda\|Dx\|_1$$



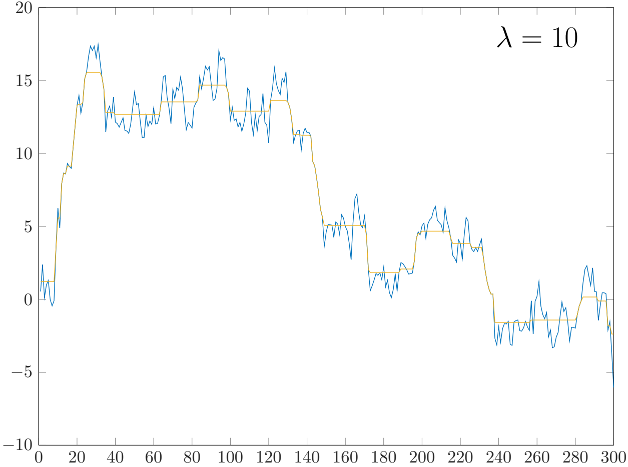
Result

$$\text{minimize } \frac{1}{2}\|x - y\|_2^2 + \lambda\|Dx\|_1$$



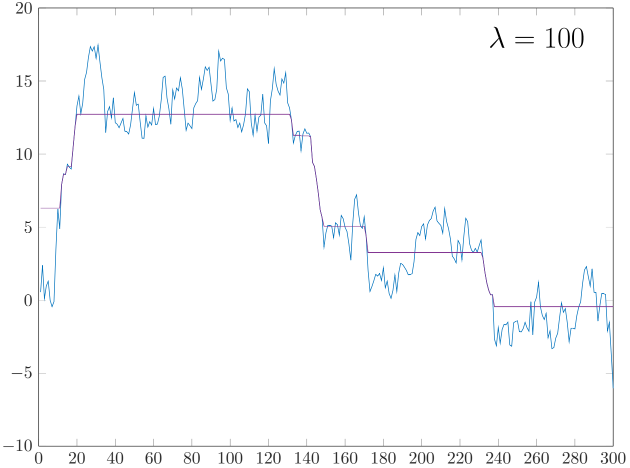
Result

$$\text{minimize } \frac{1}{2}\|x-y\|_2^2 + \lambda\|Dx\|_1$$



Result

$$\text{minimize } \frac{1}{2}\|x - y\|_2^2 + \lambda\|Dx\|_1$$



Piece-wise linear approximation

- maybe we want a piece-wise linear approximation instead
- introduce the second order discrete difference

$$D_2 = \begin{bmatrix} 1 & -2 & 1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & 1 & -2 & 1 & \end{bmatrix}$$

- this is zero on any line
- how to model piece-wise linear approximation?

Piece-wise linear approximation

- maybe we want a piece-wise linear approximation instead
- introduce the second order discrete difference

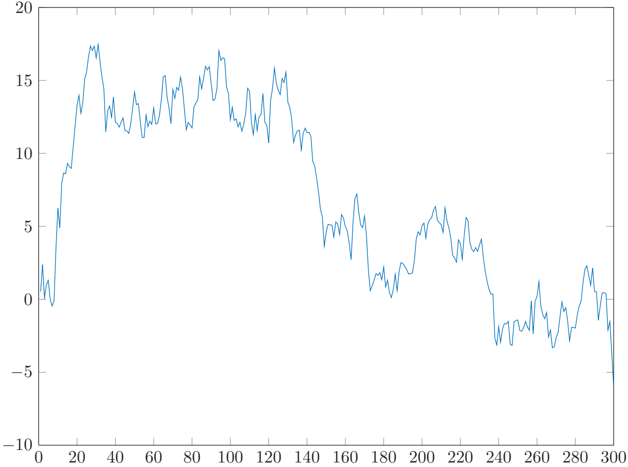
$$D_2 = \begin{bmatrix} 1 & -2 & 1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & 1 & -2 & 1 & \end{bmatrix}$$

- this is zero on any line
- how to model piece-wise linear approximation?

$$\text{minimize } \tfrac{1}{2}\|x - y\|_2^2 + \lambda \|D_2 x\|_1$$

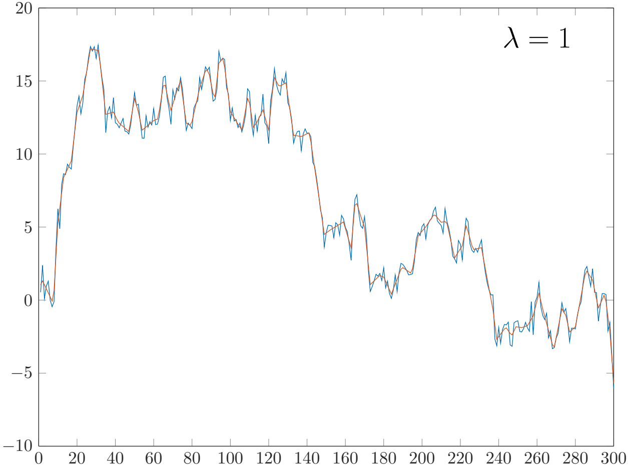
Result

minimize $\frac{1}{2}\|x - y\|_2^2 + \lambda \|D_2 x\|_1$



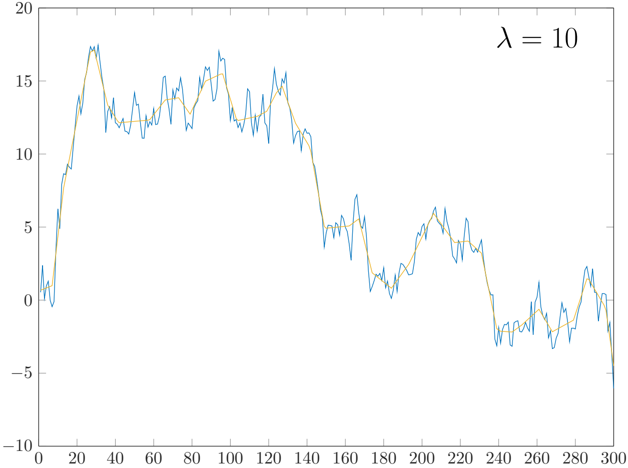
Result

minimize $\frac{1}{2}\|x-y\|_2^2+\lambda\|D_2x\|_1$



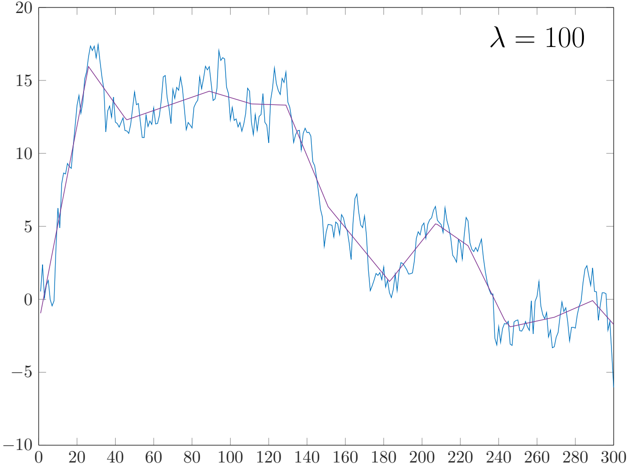
Result

$$\text{minimize } \frac{1}{2}\|x - y\|_2^2 + \lambda\|D_2 x\|_1$$



Result

$$\text{minimize } \tfrac{1}{2}\|x - y\|_2^2 + \lambda \|D_2 x\|_1$$



Smooth second derivative

- we might instead want a smooth second derivative
- how to model this?

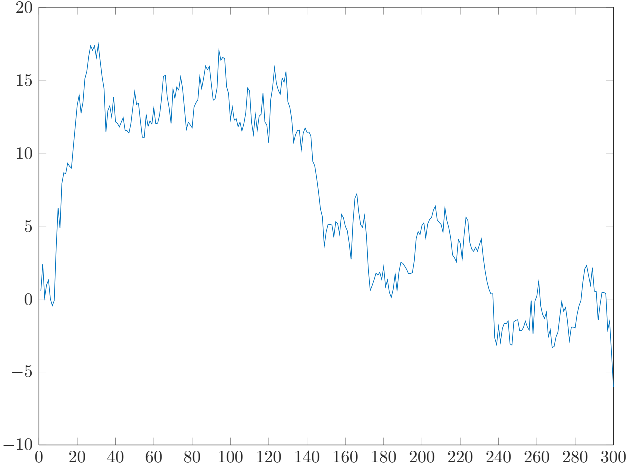
Smooth second derivative

- we might instead want a smooth second derivative
- how to model this?

$$\text{minimize } \frac{1}{2}\|x - y\|_2^2 + \lambda\|D_2 x\|_2^2$$

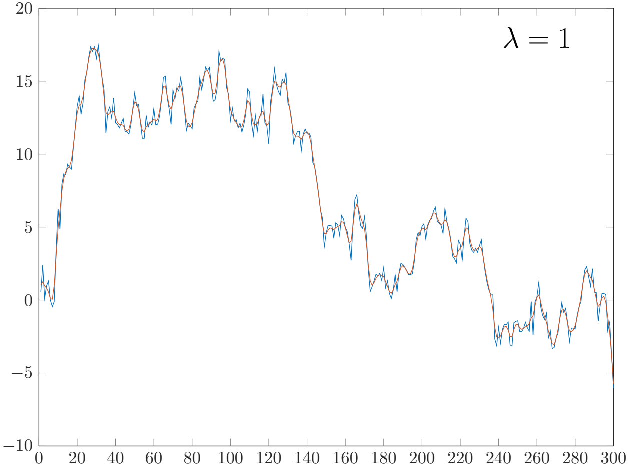
Result

$$\text{minimize } \frac{1}{2}\|x - y\|_2^2 + \lambda\|D_2 x\|_2^2$$



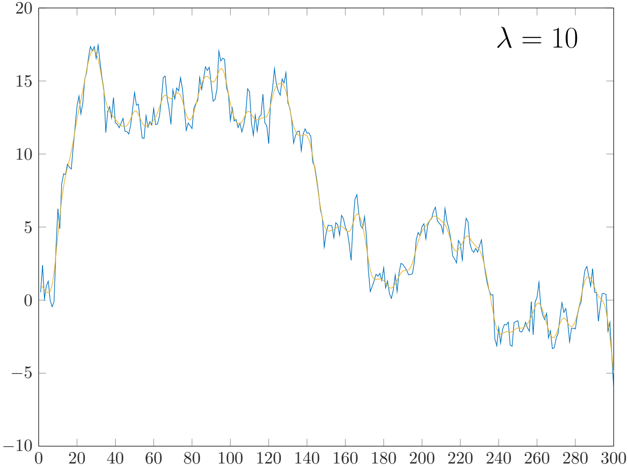
Result

minimize $\frac{1}{2}\|x - y\|_2^2 + \lambda \|D_2 x\|_2^2$



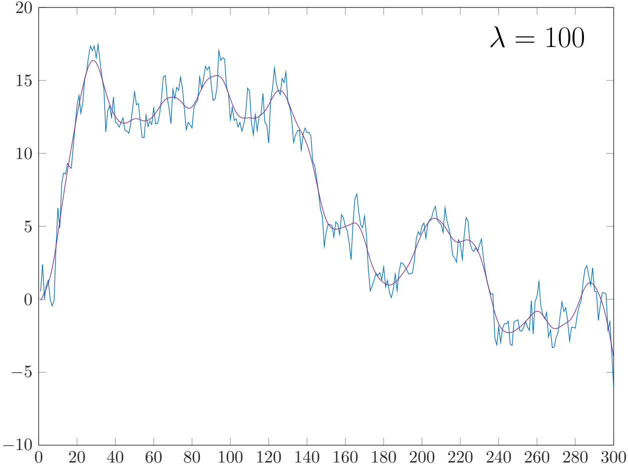
Result

minimize $\frac{1}{2}\|x - y\|_2^2 + \lambda \|D_2 x\|_2^2$



Result

$$\text{minimize } \frac{1}{2}\|x - y\|_2^2 + \lambda\|D_2 x\|_2^2$$



Periodic disturbances

- assume that our signal is disturbed by a periodic signal $p_d \in \mathbb{R}^n$
- p_d could model yearly/weekly/daily variations
- our measurement is still y
- we are interested in, say, a piece-wise linear estimation of $y - p$
- how to model this?

Periodic disturbances

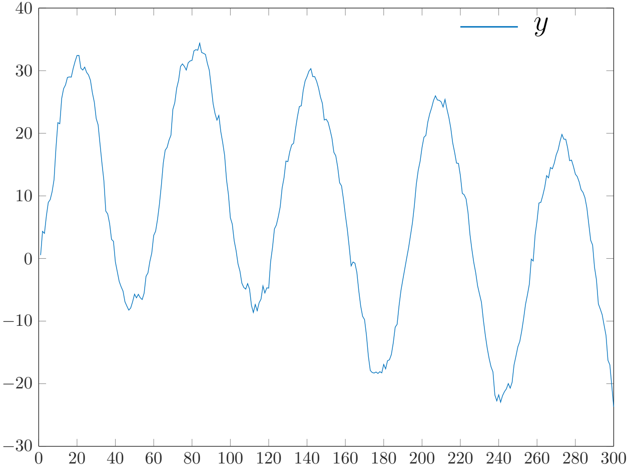
- assume that our signal is disturbed by a periodic signal $p_d \in \mathbb{R}^n$
- p_d could model yearly/weekly/daily variations
- our measurement is still y
- we are interested in, say, a piece-wise linear estimation of $y - p$
- how to model this? assume period is T

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|x - (y - p)\|_2^2 + \lambda \|D_2 x\|_1 \\ \text{subject to} & p_i = p_{i+k_i T} \text{ for } i = 1, \ldots, T \text{ as long as } i + k_i T \leq n \end{array}$$

- x and p optimization variables! (p should estimate p_d)

Result

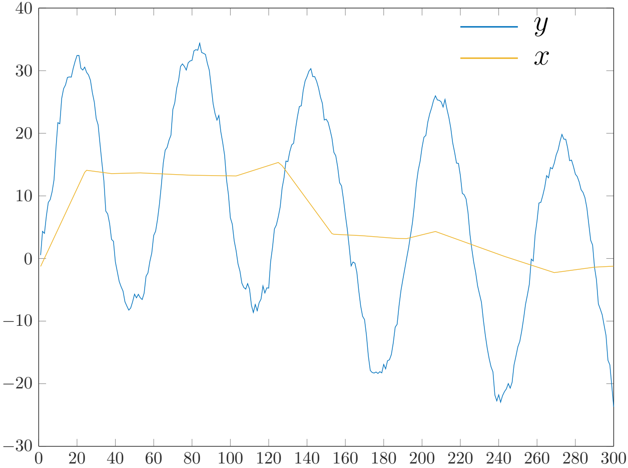
$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|x-(y-p)\|_2^2 + \lambda \|D_2 x\|_1 \\ \text{subject to} & p_i = p_{i+k_i T} \text{ for } i = 1, \ldots, T \text{ as long as } i + k_i T \leq n \end{array}$$



Result

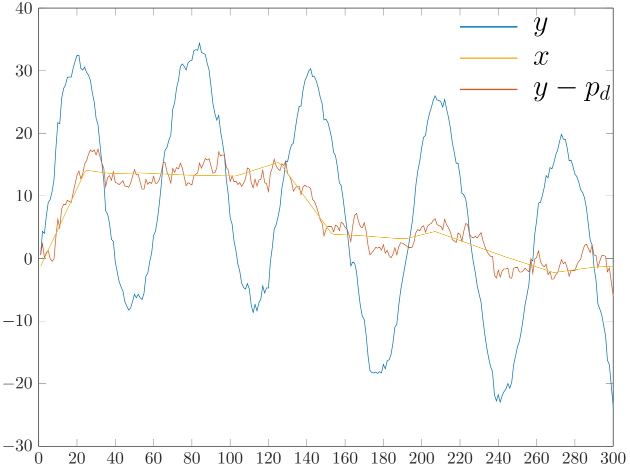
minimize $\quad \frac{1}{2}\|x - (y - p)\|_2^2 + \lambda\|D_2 x\|_1$

subject to $\quad p_i = p_{i+k_i T}$ for $i = 1, \ldots, T$ as long as $i + k_i T \leq n$



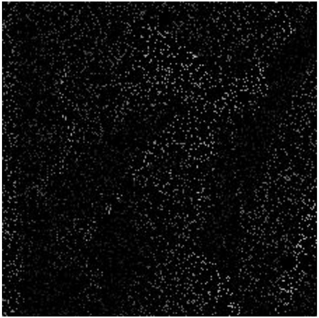
Result

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|x-(y-p)\|_2^2+\lambda\|D_2x\|_1 \\ \text{subject to} & p_i = p_{i+k_iT} \text{ for } i=1,\ldots,T \text{ as long as } i+k_iT \leq n \end{array}$$



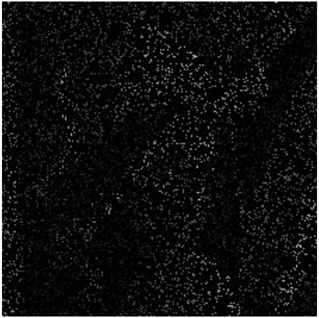
Two-dimensional reconstruction

- can also reconstruct images (2D-signals)
- example: 90% of pixels in image lost



Two-dimensional reconstruction

- can also reconstruct images (2D-signals)
- example: 90% of pixels in image lost



- reconstruct using difference in 2D (TV-norm)

$$\text{minimize} \quad \sum_{i=1}^{n-1}\sum_{j=1}^{m} |x_{i,j} - x_{i+1,j}| + \sum_{i=1}^{n}\sum_{j=1}^{m-1} |x_{i,j} - x_{i,j+1}|$$

- known pixels are set to correct value

Two-dimensional reconstruction

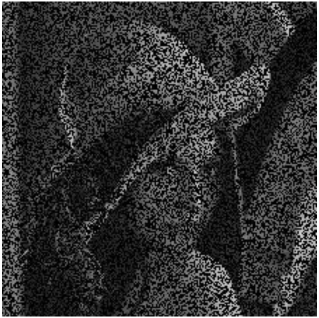
- example: 70% of pixels in image lost

Two-dimensional reconstruction

- example: 70% of pixels in image lost

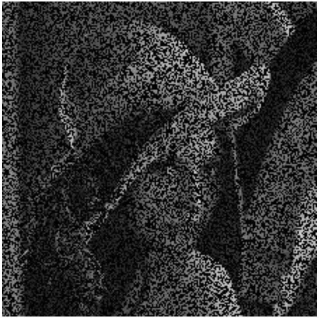
Two-dimensional reconstruction

- example: 50% of pixels in image lost



Two-dimensional reconstruction

- example: 50% of pixels in image lost



Two-dimensional reconstruction

- example: 30% of pixels in image lost

Two-dimensional reconstruction

- example: 30% of pixels in image lost

Comparison to ground truth

Modeling idea

If we want something to

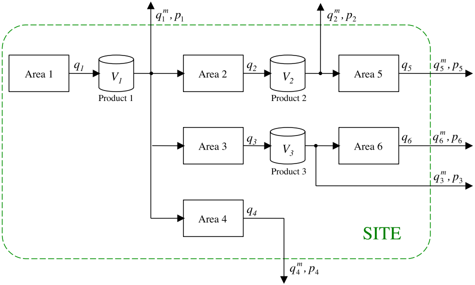
- hold approximately, use $\|\cdot\|_2^2$
- be sparse, use $\|\cdot\|_1$
- enforce something, use constraints

Lecture 11

- ○ Introduction to convex optimization
- ○ Convex optimization modeling
- • **A model predictive control example**
- ○ Duality

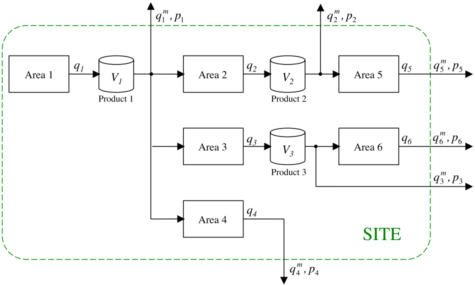
A model predictive control example

- Schematic view of production site:



- Maximize profit dispite disturbances in utilities
- Use buffer tanks and routing to achieve this
- Satisfy constraints for buffer levels and product rates
- Base routing and rate decisions on optimization problem solution!

Model site dynamic behavior



Site model (mass balance for tanks, rates assumed static):

$$V_1(t+1) = V_1(t) + q_1(t) - q_1^m(t) - q_2^{\text{in}}(t) - q_3^{\text{in}}(t) - q_4^{\text{in}}(t)$$
$$V_2(t+1) = V_2(t) + q_2(t) - q_2^m(t) - q_5^{\text{in}}(t)$$
$$V_3(t+1) = V_3(t) + q_3(t) - q_3^m(t) - q_6^{\text{in}}(t)$$

Buffer tank constraints

$$V_1^{\min} \leq V_1(t) \leq V_1^{\max}$$
$$V_2^{\min} \leq V_2(t) \leq V_2^{\max}$$
$$V_3^{\min} \leq V_3(t) \leq V_3^{\max}$$

Production rate constraints

- Want production rates q_i to satisfy

$$q_i(t) = 0 \qquad \text{or} \qquad q_i^{\min} \leq q_i(t) \leq q_i^{\max}$$

  where $q_i^{\min} > 0$ (rather shut off than produce too little)

- But a cost $l_i(t)$ is associated with shutting down:

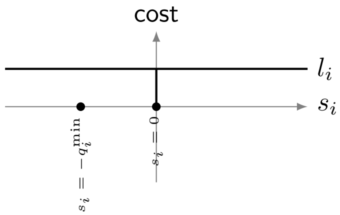
$$l_i(t) = \begin{cases} l_i & \text{if } q_i(t) = 0 \\ 0 & \text{else} \end{cases}$$

Optimization formulation

- Can be modeled with the following constraints:

$$q_i^{\min} + s_i(t) \le q_i(t) \le q_i^{\max} \text{ with } s_i(t) = \{0, -q_i^{\min}\}$$

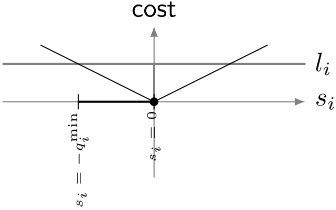
and the additional objective cost that $l_i \|s_i(t)\|_0$



- Both the cost and the constraint are nonconvex!

Convex relaxation

- Relax the inclusion $s_i(t) \in \{0, -q_i^{\min}\}$ to $s_i(t) \in [0, -q_i^{\min}]$
- Approximate the cost $l_i \|s_i(t)\|_0$ with $\hat{l}_i \|s_i(t)\|_1$



- This convex formulation very often gives either

$$s_i = 0 \; (q_i \in [q_i^{\min}, q_i^{\max}]) \qquad \text{or} \qquad s_i = -q_i^{\min} \; (q_i = 0)$$

Utility constraints

Area	1	2	3	4	5	6
Steam HP	x		x			
Steam MP		x		x		x
Cooling water	x	x	x	x	x	x

Constraints due to shared utilities

$$c_{11}q_1(t) + c_{13}q_3(t) \leq U_1(t)$$
$$c_{22}q_2(t) + c_{24}q_4(t) + c_{26}q_6(t) \leq U_2(t)$$
$$\sum_{i=1}^{6} c_{3i}q_i(t) \leq U_3(t)$$

Cost function

- We have the cost from the start-up $l_i(t)$ for all units i and times t
- We also want to maximize profit $p_i q_i^m(t)$ for all units i and times t
- Combined cost function to maximize:

$$\sum_{i=1}^{6} \sum_{t=0}^{N} (p_i q_i^m(t) - l_i \|s_i(t)\|_1)$$

MPC of site

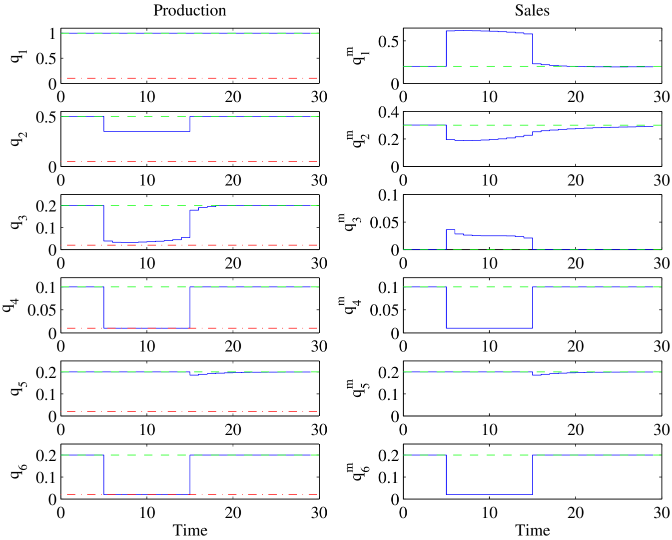
- Consider the full optimization problem:

$$
\begin{array}{ll}
\text{maximize} & \sum_{i=1}^{6} \sum_{t=0}^{N} \left(p_i q_i^m(t) - l_i \|s_i(t)\|_1\right) \\
\text{subject to} & V_1(t+1) = V_1(t) + q_1(t) - q_1^m(t) - q_2^{\text{in}}(t) - q_3^{\text{in}}(t) - q_4^{\text{in}}(t) \\
& V_2(t+1) = V_2(t) + q_2(t) - q_2^m(t) - q_5^{\text{in}}(t) \\
& V_3(t+1) = V_3(t) + q_3(t) - q_3^m(t) - q_6^{\text{in}}(t) \\
& V_j^{\min} \le V_j(t) \le V_j^{\max} \\
& s_i(t) = [0, -q_i^{\min}] \\
& q_i^{\min} + s_i(t) \le q_i(t) \le q_i^{\max} \\
& c_{11} q_1(t) + c_{13} q_3(t) \le U_1(t) \\
& c_{22} q_2(t) + c_{24} q_4(t) + c_{26} q_6(t) \le U_2(t) \\
& \sum_{i=1}^{6} c_{3i} q_i(t) \le U_3(t)
\end{array}
$$

for $j = 1, 2, 3$, $i = 1, \ldots, 6$, $t = 1, \ldots, N$

- Solve this repeaditely with measured $V_i(0)$ and estimated $U_j(t)$

Solution to one problem



Nonconvex problem can be solved using MIQP.
This convex relaxation gives the same solutions. Much faster to solve.

Lecture 11

- ◦ Introduction to convex optimization
- ◦ Convex optimization modeling
- ◦ A model predictive control example
- • **Duality**

Linear Programming Example

Product	# of items	Profit / item
Garden Furniture 1	x_1	c_1
Garden Furniture 2	x_2	c_2
Sled 1	x_3	c_3
Sled 2	x_4	c_4

Constraints for sub-division 1:

$$7x_1 + 10x_2 \leq 100 \qquad \text{(Sawing)}$$
$$16x_1 + 12x_2 \leq 135 \qquad \text{(Assembling)}$$

Constraints for sub-division 2:

$$10x_3 + 9x_4 \leq 70 \qquad \text{(Sawing)}$$
$$6x_3 + 9x_4 \leq 60 \qquad \text{(Assembling)}$$

Painting Constraint:

$$5x_1 + 3x_2 + 3x_3 + 2x_4 \leq 45$$

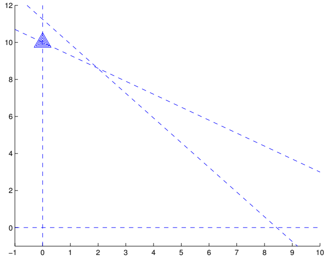
Linear Programming Example

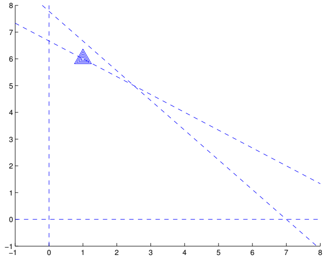
Mathematical formulation:

$$
\begin{array}{ll}
\text{Maximize} & c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4 \\
\\
\text{subject to} & 7x_1 + 10x_2 \le 100 \\
& 16x_1 + 12x_2 \le 135 \\
& 10x_3 + 9x_4 \le 70 \\
& 6x_3 + 9x_4 \le 60 \\
& 5x_1 + 3x_2 + 3x_3 + 2x_4 \le 45 \\
& x \ge 0
\end{array}
$$

Numerical Results

Optimal solution for Division 1 (left) and Division 2 (right). Common constraint active (i.e. equality holds).





Dual Variables

- Dual variables are the marginal prices for resources
- If the capacity for a resource is increased by 1, the total profit is increased by the corresponding dual variable
- This gives insight to which resource to increase to gain most

Numerical Results

Optimal dual variables and their respective constraints:

Constraint	Dual variable
$7x_1 + 10x_2 \leq 100$	1.04
$16x_1 + 12x_2 \leq 135$	0
$10x_3 + 9x_4 \leq 70$	0
$6x_3 + 9x_4 \leq 60$	0.4
$5x_1 + 3x_2 + 3x_3 + 2x_4 \leq 45$	3.2

Optimal value: $p^* = c^T x^* = 272$

If common (painting) constraint capacity increased to 46, optimal value becomes $272 + 3.2 = 275.2$

Company would gain most by increasing painting capacity

Linear Programming Duality

Linear Program:

$$p^* = \begin{cases} \max\limits_{x} & c^T x \\ \text{subject to} & Ax \leq b,\ x \geq 0 \end{cases}$$

where $p^* = c^T x^*$ is the optimal value attained by x^*.

For the constraints $Ax \leq b$, introduce dual variables $\lambda \geq 0$ and construct the corresponding dual function $g(\lambda)$:

$$g(\lambda) = \max_{x \succeq 0} \left[c^T x + \lambda^T (b - Ax) \right]$$

We have $g(\lambda) \geq p^*$. Let $x^\star$ be optimal and $\lambda \geq 0$, then $\lambda^T (b - Ax^\star) \geq 0$. Therefore

$$g(\lambda) = \max_{x \succeq 0} \left[c^T x + \lambda^T (b - Ax) \right] \geq c^T x^\star + \lambda^T (b - Ax^\star) \geq c^T x^\star$$

Linear Programming Duality cont'd

Tightest upper bound to p^* obtained by minimizing $g(\lambda)$:

$$d^* = \min_{\lambda \geq 0} g(\lambda) = \min_{\lambda \geq 0} \max_{x \geq 0} \left[c^T x + \lambda^T (b - Ax) \right]$$

Optimal value d^* for this min-max problem is attained by $x = x^*$ and $\lambda = \lambda^*$.

Further we have that $p^* = c^T x^* = d^*$. This equality is referred to as *strong duality*

Dual optimal values and d^* can be obtained by solving

$$\begin{array}{ll} \min\limits_{\lambda} & b^T \lambda \\ \text{subject to} & A^T \lambda \geq c, \lambda \geq 0 \end{array}$$

Note symmetry to primal problem

Linear Programming Duality

$$\begin{array}{lll} \max\limits_{x} & c^T x \\ \text{with} & Ax \leq b \\ & x \geq 0 \end{array} \quad = \quad \begin{array}{ll} \min\limits_{\lambda} & b^T \lambda \\ \text{with} & A^T \lambda \geq c \\ & \lambda \geq 0 \end{array}$$

Lecture 10 and 11

Lecture 10

- Linear Programming (LP)
- LP in production planning example
- Model Predictive Control
- A portfolio optimization problem

Lecture 11

- Introduction to convex optimization
- Convex optimization modeling
- A model predictive control example
- Duality