EE/ACM 150 - Applications of Convex Optimization in Signal Processing and Communications
Lecture 13

Andre Tkacenko

Signal Processing Research Group
Jet Propulsion Laboratory

May 15, 2012

Caltech

Outline

Basic Norm Approximation Problem

The most elementary *norm approximation problem* is the following unconstrained one.

Norm Approximation Problem:

$$\text{minimize} \quad \|\mathbf{A}\mathbf{x} - \mathbf{b}\| \quad ,$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$ are the problem data, and $\mathbf{x} \in \mathbb{R}^n$ is the variable.

A solution is often called an approximate solution to $\mathbf{A}\mathbf{x} \approx \mathbf{b}$ in the norm $\|\cdot\|$.

Residual: The vector

$$\mathbf{r} \triangleq \mathbf{A}\mathbf{x} - \mathbf{b} \, ,$$

is called the *residual* for the problem and its components are sometimes called the individual *residuals* associated with $\mathbf{x}$.

Properties of Norm Approximation Problem:

- It is always convex and solvable.
- Its optimal value is zero if and only if $\mathbf{b} \in \mathcal{R}(\mathbf{A})$.
- For the more interesting case where $\mathbf{b} \notin \mathcal{R}(\mathbf{A})$, we assume the following.
  - The columns of $\mathbf{A}$ are linearly independent.
  - The matrix $\mathbf{A}$ is 'tall' in that $m > n$.

Interpretations of the Norm Approximation Problem

There are several interesting interpretations of the solution

$$\mathbf{x}^\star = \underset{\mathbf{x}}{\operatorname{argmin}} \, ||\mathbf{A}\mathbf{x} - \mathbf{b}|| \ ,$$

to the norm approximation problem.

- **Geometric:** The point $\mathbf{A}\mathbf{x}^\star$ is the one in $\mathcal{A} \triangleq \mathcal{R}(\mathbf{A})$ that is closest to $\mathbf{b}$ in the norm $||\cdot||$. In other words, it is the *projection* of $\mathbf{b}$ onto the linear subspace $\mathcal{A}$.
- **Regression:** Expressing $\mathbf{A}\mathbf{x}$ as
$$\mathbf{A}\mathbf{x} = x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n \ ,$$
where $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^n$ are the columns of $\mathbf{A}$, the approximation problem is to find the best fit of the vector $\mathbf{b}$ by a linear combination of the columns of $\mathbf{A}$. Viewing $\mathbf{a}_1, \ldots, \mathbf{a}_n$ as *regressors*, the vector $\mathbf{A}\mathbf{x}^\star$ is called the *regression of* $\mathbf{b}$.
- **Estimation:** Consider a linear measurement model of the form
$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v} \, ,$$
where $\mathbf{y} \in \mathbb{R}^n$ is the vector measurement and $\mathbf{v} \in \mathbb{R}^m$ is an unknown measurement error or noise. The goal is to find the unknown input $\mathbf{x}$. Given $\mathbf{y} = \mathbf{b}$, the most plausible guess of $\mathbf{x}$ is $\mathbf{x}^\star$.
- **Optimal Design:** The components of $\mathbf{x}$ are input design variables to a linear system with output $\mathbf{y} = \mathbf{A}\mathbf{x}$. Here, $\mathbf{x}^\star$ yields the result $\mathbf{y}^\star = \mathbf{A}\mathbf{x}^\star$ that best approximates the desired result $\mathbf{b}$.

Examples of the Norm Approximation Problem

- **Weighted norm approximation:** minimize $\|\mathbf{W}(\mathbf{A}\mathbf{x}-\mathbf{b})\|$, where $\mathbf{W} \in \mathbb{R}^{m \times m}$ is a *weighting matrix* used to emphasize some residuals over others.
  This is identical to the basic problem with $\widetilde{\mathbf{A}} \triangleq \mathbf{W}\mathbf{A}$ and $\widetilde{\mathbf{b}} \triangleq \mathbf{W}\mathbf{b}$.
- **Least-squares approximation:** minimize $\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2 = r_1^2 + \cdots + r_m^2$.
  A solution $\mathbf{x}^\star$ is optimal if and only if it satisfies the *normal equations*

  $$\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}\,.$$

  If $\operatorname{rank}(\mathbf{A}) = n$, then $\mathbf{x}^\star = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}$.
- **Chebyshev approximation:** minimize $\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_\infty = \max\{|r_1|,\ldots,|r_m|\}$.
  This can be cast and solved as an LP

  $$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & -t\mathbf{1} \preceq \mathbf{A}\mathbf{x}-\mathbf{b} \preceq t\mathbf{1} \end{array},$$

  with variables $\mathbf{x} \in \mathbb{R}^n$ and $t \in \mathbb{R}$.
- **Sum of absolute residuals approximation:** minimize $\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_1 = |r_1| + \cdots + |r_m|$.
  This can be cast and solved as an LP

  $$\begin{array}{ll} \text{minimize} & \mathbf{1}^T\mathbf{y} \\ \text{subject to} & -\mathbf{y} \preceq \mathbf{A}\mathbf{x}-\mathbf{b} \preceq \mathbf{y} \end{array},$$

  with variables $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$.

Penalty Function Approximation Problem

For ℓ_p-norm approximation, the objective is

$$\left(|r_1|^p + \cdots + |r_m|^p\right)^{1/p} .$$

As with least-squares problems, we can consider an equivalent problem with objective

$$|r_1|^p + \cdots + |r_m|^p ,$$

which is a separable and symmetric function of the residuals.

A useful generalization of this appears in the *penalty function approximation problem*.

Penalty Function Approximation Problem:

$$\begin{array}{ll} \text{minimize} & \phi(r_1) + \cdots + \phi(r_m) \\ \text{subject to} & \mathbf{r} = \mathbf{A}\mathbf{x} - \mathbf{b} \end{array} ,$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$ are the problem data, $\mathbf{r} \in \mathbb{R}^m$ and $\mathbf{x} \in \mathbb{R}^n$ are the variables, and $\phi : \mathbb{R} \to \mathbb{R}$ is the *penalty function*.

- We will assume that ϕ is convex, so we have a convex optimization problem.
- Often ϕ is symmetric, nonnegative, and satisfies $\phi(0) = 0$.
- The problem can be viewed as minimizing the total cost or penalty incurred by the residuals $r_1, \ldots, r_m$ obtained for the approximation $\mathbf{A}\mathbf{x}$ of $\mathbf{b}$.

Examples of Penalty Functions

- **ℓ_p-norm:**

$$\phi(u) = |u|^p \ , \ p \geq 1 \, .$$

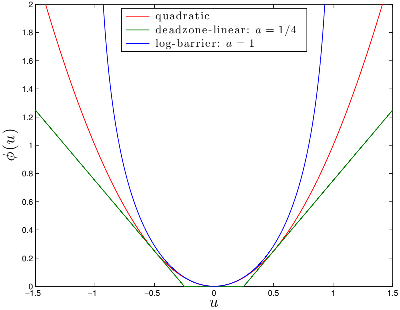
When $p = 2$, we get the *quadratic* penalty function $\phi(u) = u^2$.

- **Deadzone-linear:** (with width $a > 0$)

$$\phi(u) = \max\{0, |u| - a\} \ .$$

- **Log-barrier:** (with limit $a > 0$)

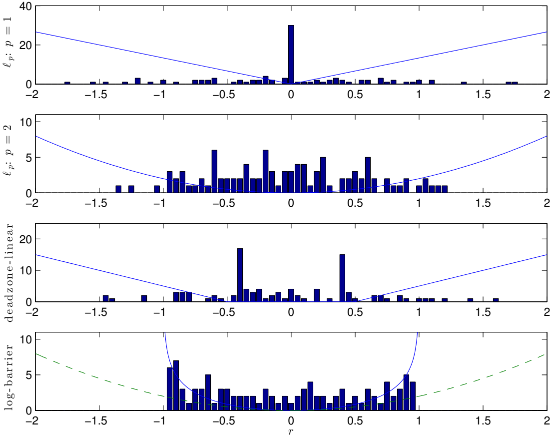
$$\phi(u) = \begin{cases} -a^2 \log\left(1 - (u/a)^2\right) \ , & |u| < a \\ \infty \, , & |u| \geq a \end{cases} \ .$$



Nuances of Different Penalty Functions

For a random penalty function approximation problem with $m = 100$ and $n = 30$, we have the following histograms for the following penalty functions.

$$\phi(u) = |u| \ , \ \phi(u) = u^2 \ , \ \phi(u) = \max\{0, |u| - (1/2)\} \ , \ \phi(u) = -\log\big(1 - u^2\big) \ .$$



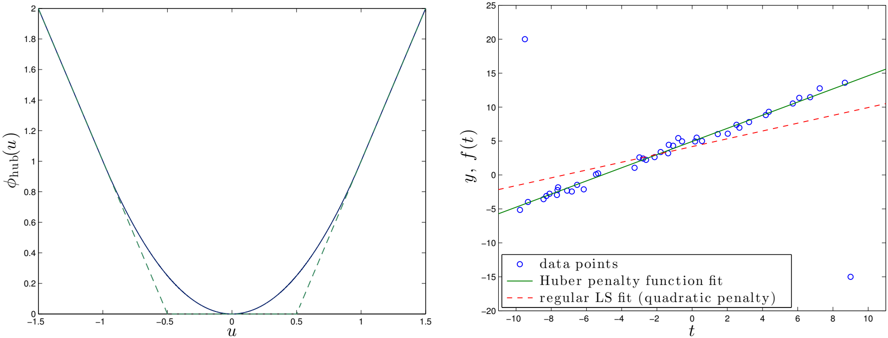
The *shape* of the penalty function has a strong influence on the distribution of the residuals.

Huber Penalty Function

The *Huber penalty function* (with parameter M) shares properties with the ℓ_1 and ℓ_2 penalties:

$$\phi_{\mathrm{hub}}(u) = \begin{cases} u^2 \,, & |u| \le M \\ M\left(2\,|u| - M\right) = 2M\left(|u| - (M/2)\right)\,, & |u| > M \end{cases} .$$

This is sometimes referred to as the *robust least-squares* penalty function, since the linear growth for large u makes the approximation less sensitive to outliers.



- Left: Huber penalty function for $M = 1$.
- Right: Affine function $f(t) = \alpha + \beta t$ fitted to 42 points (t_i, y_i) (circles) using Huber (solid) and quadratic (dashed) penalty functions.

Interpretations of Adding Constraints

As expected, it is possible to add constraints to the basic norm approximation problem. When the constraints are convex, the resulting problem is also convex. In a practical setting, constraints may arise for a variety of reasons.

- In an approximation problem setting, constraints can rule out certain unacceptable approximations of $\mathbf{b}$, or ensure that the approximator $\mathbf{Ax}$ satisfies certain properties.
- For an estimation problem, constraints arise as prior knowledge of the vector $\mathbf{x}$ to be estimated or the estimation error $\mathbf{v}$.
- Constraints arise in a geometric setting in determining the projection of $\mathbf{b}$ onto a more complicated set than a subspace, for example, a cone or a polyhedron.

Examples of Adding Constraints

- **Nonnegativity constraints on variables:**

$$\begin{array}{ll} \text{minimize} & \|\mathbf{Ax}-\mathbf{b}\| \\ \text{subject to} & \mathbf{x} \succeq \mathbf{0} \end{array}.$$

This is useful for estimating parameters known to be nonnegative, such as powers, intensities, or rates.

- **Variable bounds:**

$$\begin{array}{ll} \text{minimize} & \|\mathbf{Ax}-\mathbf{b}\| \\ \text{subject to} & \mathbf{l} \preceq \mathbf{x} \preceq \mathbf{u} \end{array}.$$

This formulation arises out of prior knowledge of the intervals in which each variable lies.

- **Probability distribution:**

$$\begin{array}{ll} \text{minimize} & \|\mathbf{Ax}-\mathbf{b}\| \\ \text{subject to} & \mathbf{x} \succeq \mathbf{0}\,,\ \mathbf{1}^T\mathbf{x} = 1 \end{array}.$$

This arises in the estimation of proportions or relative frequencies.

- **Norm ball constraint:**

$$\begin{array}{ll} \text{minimize} & \|\mathbf{Ax}-\mathbf{b}\| \\ \text{subject to} & \|\mathbf{x}-\mathbf{x}_0\| \leq d \end{array}.$$

This arises in estimation, where $\mathbf{x}_0$ is a prior guess of what $\mathbf{x}$ is, and d is a maximum plausible deviation. The constraint can also represent a *trust region* in which the model leading to the norm approximation problem is valid.

Interpretations of Least-Norm Problems

A companion to the norm approximation problem is the *least-norm problem* given as follows.

Least-Norm Problem:

$$\begin{array}{ll} \text{minimize} & ||\mathbf{x}|| \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b} \end{array},$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$ are the problem data, and $\mathbf{x} \in \mathbb{R}^n$ is the variable.

Properties:

- The problem is only interesting when $\mathbf{A}$ is 'fat', meaning $m < n$.
- With $\mathbf{x}_0$ as any solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$ and $\mathbf{Z} \in \mathbb{R}^{n \times k}$ as any matrix whose columns are a basis for $\mathcal{N}(\mathbf{A})$, it can be reformulated as the norm approximation problem

$$\text{minimize} \quad ||\mathbf{x}_0 + \mathbf{Z}\mathbf{u}|| \quad ,$$

  with variable $\mathbf{u} \in \mathbb{R}^k$.

Interpretations of the solution $\mathbf{x}^\star = \underset{\mathbf{A}\mathbf{x}=\mathbf{b}}{\operatorname{argmin}} ||\mathbf{x}||$:

- **Geometric:** $\mathbf{x}^\star$ is the point in the affine set $\{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}\}$ with minimum distance to the origin $\mathbf{0}$.
- **Estimation:** $\mathbf{b} = \mathbf{A}\mathbf{x}$ are given measurements of $\mathbf{x}$; $\mathbf{x}^\star$ is the smallest (and hence 'most plausible') estimate consistent with the measurements.
- **Design:** $\mathbf{x}$ are design variables (inputs) and $\mathbf{b}$ are the known required results (outputs); $\mathbf{x}^\star$ is the smallest (and hence 'most efficient') design that satisfies the requirements.

Examples of Least-Norm Problems

- **Least-squares solution of linear equations:**

$$\begin{array}{ll} \text{minimize} & \|\mathbf{x}\|_2^2 \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b} \end{array} .$$

  This can be solved via the KKT optimality conditions

$$2\mathbf{x}^\star + \mathbf{A}^T \boldsymbol{\nu}^\star = \mathbf{0}\,,\ \mathbf{A}\mathbf{x}^\star = \mathbf{b} \Longleftrightarrow \boldsymbol{\nu}^\star = -2\left(\mathbf{A}\mathbf{A}^T\right)^{-1}\mathbf{b}\,,\ \mathbf{x}^\star = \mathbf{A}^T\left(\mathbf{A}\mathbf{A}^T\right)^{-1}\mathbf{b}\,.$$

- **Sparse solutions via least ℓ_1-norm:**

$$\begin{array}{ll} \text{minimize} & \|\mathbf{x}\|_1 \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b} \end{array} .$$

  This can be solved as an LP

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T \mathbf{y} \\ \text{subject to} & -\mathbf{y} \preceq \mathbf{x} \preceq \mathbf{y}\,,\ \mathbf{A}\mathbf{x} = \mathbf{b} \end{array} .$$

  The problem tends to produce a *sparse* solution $\mathbf{x}^\star$.
- **Least-penalty problems:** A common extension is to replace the norm objective with a sum of penalties

$$\begin{array}{ll} \text{minimize} & \phi(x_1) + \cdots + \phi(x_n) \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b} \end{array} .$$

  Here, $\phi : \mathbb{R} \rightarrow \mathbb{R}$ is typically a convex penalty function.

Bi-Criterion Formulation of Regularized Approximation

The goal behind regularized approximation is to find a vector $\mathbf{x}$ that is small (if possible), and also makes the residual $\mathbf{r} = \mathbf{A}\mathbf{x} - \mathbf{b}$ small. This leads to the following bi-criterion problem.

Bi-Criterion Regularized Approximation Problem:

$$\text{minimize (with respect to } \mathbb{R}^2_+) \quad (||\mathbf{A}\mathbf{x} - \mathbf{b}||, ||\mathbf{x}||) \quad ,$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$ are the problem data, $\mathbf{x} \in \mathbb{R}^n$ is the variable, and the norms on $\mathbb{R}^m$ and $\mathbb{R}^n$ can be different.

In summary, the idea is to find a good approximation $\mathbf{A}\mathbf{x} \approx \mathbf{b}$ with a small $\mathbf{x}$.

Interpretations:

- **Estimation:** We assume the linear measurement model $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$ holds, with the prior knowledge that $||\mathbf{x}||$ is small.
- **Optimal design:** A small $\mathbf{x}$ may be cheaper or more efficient than a larger one, or the linear model $\mathbf{y} = \mathbf{A}\mathbf{x}$ is only valid for small $\mathbf{x}$.
- **Robust approximation:** A good approximation $\mathbf{A}\mathbf{x} \approx \mathbf{b}$ with small $\mathbf{x}$ is less sensitive to errors in $\mathbf{A}$ than a good approximation with large $\mathbf{x}$.

Scalarization Form of Regularized Approximation

Solving the bi-criterion regularized approximation problem via scalarization leads to what is called *regularization*.

Regularization:

$$\text{minimize} \quad ||\mathbf{A}\mathbf{x} - \mathbf{b}|| + \gamma \, ||\mathbf{x}|| \quad ,$$

where $\gamma > 0$ is a problem parameter that traces out the optimal trade-off curve.

When the Euclidean norm is used for both objectives, another form of regularization is to minimize $||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2 + \delta \, ||\mathbf{x}||_2^2$ with $\delta > 0$.

An equivalent form of regularization can be obtained by taking the convex combination of objectives leading to the problem

$$\text{minimize} \quad \theta \, ||\mathbf{A}\mathbf{x} - \mathbf{b}|| + (1 - \theta) \, ||\mathbf{x}|| \quad ,$$

where $\theta \in (0, 1)$ is the trade-off parameter.

Uses:

- It can be used to select the proper trade-off between size $||\mathbf{x}||$ and fit $||\mathbf{A}\mathbf{x} - \mathbf{b}||$ for the application at hand.
- It can help approximately solve the linear equation system $\mathbf{A}\mathbf{x} = \mathbf{b}$ when $\mathbf{A}$ is ill-conditioned or even singular.

Tikhonov and Smoothing Regularization

The most common form of regularization is *Tikhonov regularization* in which both norms are Euclidean.

Tikhonov Regularization:

$$\text{minimize} \quad ||\mathbf{Ax} - \mathbf{b}||_2^2 + ||\mathbf{\Gamma x}||_2^2 \quad ,$$

where $\mathbf{\Gamma} \in \mathbb{R}^{p \times n}$ is the *Tikhonov matrix*. Typically, $\mathbf{\Gamma} = \sqrt{\delta}\mathbf{I}_n$. This is a least-squares problem with solution

$$\mathbf{x}^\star = \left(\mathbf{A}^T\mathbf{A} + \mathbf{\Gamma}^T\mathbf{\Gamma}\right)^{-1}\mathbf{A}^T\mathbf{b}\,.$$

A common extension is to add several regularization terms leading to the problem

$$\text{minimize} \quad ||\mathbf{Ax} - \mathbf{b}||_2^2 + ||\mathbf{\Gamma}_1\mathbf{x}||_2^2 + \cdots + ||\mathbf{\Gamma}_q\mathbf{x}||_2^2 \quad ,$$

where $\mathbf{\Gamma}_i \in \mathbb{R}^{p_i \times n}$ for $i = 1, \ldots, q$. For example, terms which satisfy

$$[\mathbf{\Gamma}_i\mathbf{x}]_k = \begin{cases} n\left(x_{k+1} - x_k\right), & k = 1, \ldots, n-1 \quad \text{(1st order difference)} \\ n^2\left(x_{k+1} - 2x_k + x_{k-1}\right), & k = 2, \ldots, n-1 \quad \text{(2nd order difference)} \end{cases},$$

penalize large variations in $\mathbf{x}$, and as such, result in *smoothing* regularization.

Optimal Input Design Example

Consider the following *linear dynamical system* with impulse response h:

$$y(t) = \sum_{\tau=0}^{t} h(\tau)\, u(t-\tau)\ ,\ t = 0, 1, \ldots, N\,.$$

Input design problem: multicriterion problem with 3 objectives

1. Tracking error with desired output y_{des}: $J_{\text{track}} = \frac{1}{N+1} \sum_{t=0}^{N} \left(y(t) - y_{\text{des}}(t)\right)^2$.
2. Input magnitude: $J_{\text{mag}} = \frac{1}{N+1} \sum_{t=0}^{N} \left(u(t)\right)^2$.
3. Input variation: $J_{\text{der}} = \frac{1}{N} \sum_{t=0}^{N-1} \left(u(t+1) - u(t)\right)^2$.

The objective is to track the desired output using a small and slowly varying input signal.

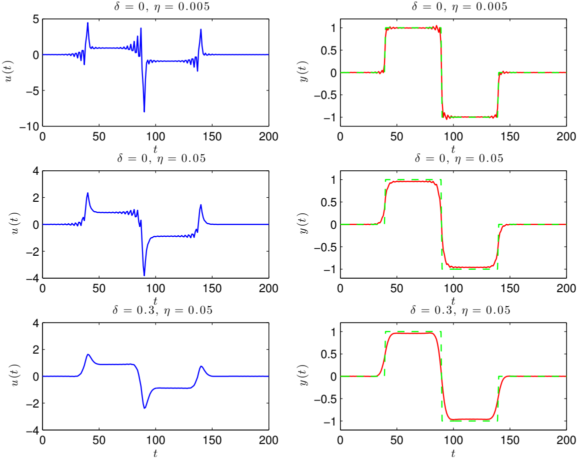
Regularized least-squares formulation:

$$\text{minimize} \quad J_{\text{track}} + \delta J_{\text{der}} + \eta J_{\text{mag}} \quad .$$

For fixed $\delta > 0$ and $\eta > 0$, this is a least-squares problem in $u(0), \ldots, u(N)$.

Optimal Input Design Example (Continued)

Three solutions on the optimal trade-off surface are shown below.
(top) $\delta = 0$, $\eta = 0.005$; (middle) $\delta = 0$, $\eta = 0.05$; (bottom) $\delta = 0.3$, $\eta = 0.05$.



As can be seen from the bottom plots, a large amount of smoothness can be obtained for only a marginal increase in tracking error.

ℓ_1-Norm Regularization

Regularization with an ℓ_1-norm can be used as a heuristic for finding a *sparse* solution. This leads to the problem

$$\text{minimize} \quad \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 + \gamma \|\mathbf{x}\|_1 \quad .$$

By varying γ, the optimal trade-off curve obtained is an approximation to the optimal trade-off between $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2$ and the *sparsity* or *cardinality* $\mathrm{card}(\mathbf{x})$ of the vector $\mathbf{x}$.

Regressor Selection Problem:

Given a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ whose columns are potential regressors, and a vector $\mathbf{b} \in \mathbb{R}^m$, we wish to fit $\mathbf{b}$ by a linear combination of $k < n$ columns of $\mathbf{A}$.

$$\begin{array}{ll} \text{minimize} & \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \\ \text{subject to} & \mathrm{card}(\mathbf{x}) \leq k \end{array} \quad .$$

- This is an NP-hard combinatorial problem.
- A straightforward way to solve it is to fix a sparsity pattern, minimize $\left\|\widetilde{\mathbf{A}}\widetilde{\mathbf{x}} - \mathbf{b}\right\|_2$, where $\widetilde{\mathbf{A}}$ is the submatrix obtained by keeping the columns of $\mathbf{A}$ corresponding to the sparsity pattern and $\widetilde{\mathbf{x}}$ is the subvector with the nonzero components of $\mathbf{x}$, and compare the optimal value for all $C(n,k) = n!/\left(k!\,(n-k)!\right)$ sparsity patterns with k nonzeros.
- A good heuristic is to use ℓ_1-norm regularization and find the smallest γ which approximately yields $\mathrm{card}(\mathbf{x}) = k$.

Signal Reconstruction Problem

In *signal reconstruction*, we are given a corrupted version of signal which we would like to recover. The source of the corruption is typically an additive noise that is unknown, small, and, unlike the desired signal, rapidly varying. The process of recovering the original signal is also called *de-noising* or *smoothing*.

Signal Reconstruction Problem:

$$\text{minimize (with respect to } \mathbb{R}_+^2\text{)} \quad \left(\left|\left| \widehat{\mathbf{x}} - \mathbf{x}_{\text{cor}} \right|\right|, \phi(\widehat{\mathbf{x}}) \right) \quad .$$

- $\mathbf{x} \in \mathbb{R}^n$ is the unknown signal.
- $\mathbf{x}_{\text{cor}} = \mathbf{x} + \mathbf{v}$ is the known corrupted version of $\mathbf{x}$, with additive noise $\mathbf{v}$.
- The variable $\widehat{\mathbf{x}}$ (reconstructed signal) is the estimate of $\mathbf{x}$.
- Typically, the norm used is the ℓ_2-norm.
- $\phi : \mathbb{R}^n \to \mathbb{R}$ is the regularization function or smoothing objective.

Examples of smoothing functions: quadratic and total variation

$$\phi_{\text{quad}}(\widehat{\mathbf{x}}) = \sum_{i=1}^{n-1} \left(\widehat{x}_{i+1} - \widehat{x}_i \right)^2 \ , \ \phi_{\text{tv}}(\widehat{\mathbf{x}}) = \sum_{i=1}^{n-1} \left| \widehat{x}_{i+1} - \widehat{x}_i \right| \ .$$

Smoothing Operators

Suppose that the vector $\mathbf{x} \in \mathbb{R}^n$ represents the value of some continuous parameter along the interval $[0, 1]$. A simple approximation to the gradient or first derivative near i/n is given by the *first order difference*

$$n\left(x_{i+1} - x_i\right), \; i = 1, \ldots, n-1,$$

which can be computed as $\mathbf{\Delta}_1 \mathbf{x}$, where $\mathbf{\Delta}_1 \in \mathbb{R}^{(n-1)\times n}$ is the bidiagonal Toeplitz matrix

$$\mathbf{\Delta}_1 = n \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}.$$

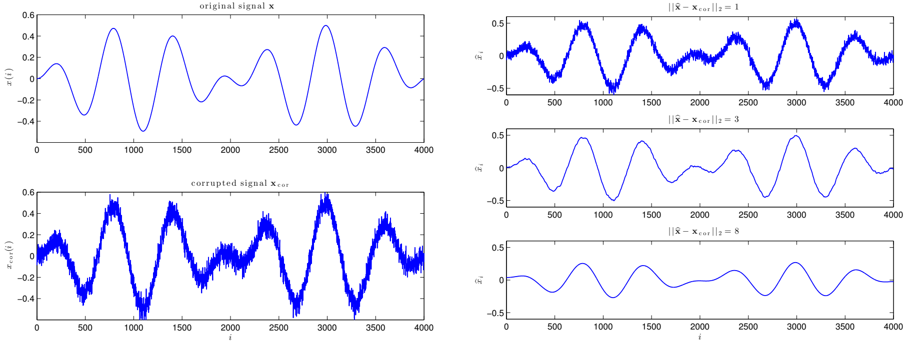
Similarly, a simple approximation to the second derivative near i/n is given by the *second order difference*

$$n\left(n\left(x_{i+1} - x_i\right) - n\left(x_i - x_{i-1}\right)\right) = n^2\left(x_{i+1} - 2x_i + x_{i-1}\right), \; i = 2, \ldots, n-1,$$

which can be computed as $\mathbf{\Delta}_2 \mathbf{x}$, where $\mathbf{\Delta}_2 \in \mathbb{R}^{(n-2)\times n}$ is the tridiagonal Toeplitz matrix

$$\mathbf{\Delta}_2 = n^2 \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{bmatrix}.$$

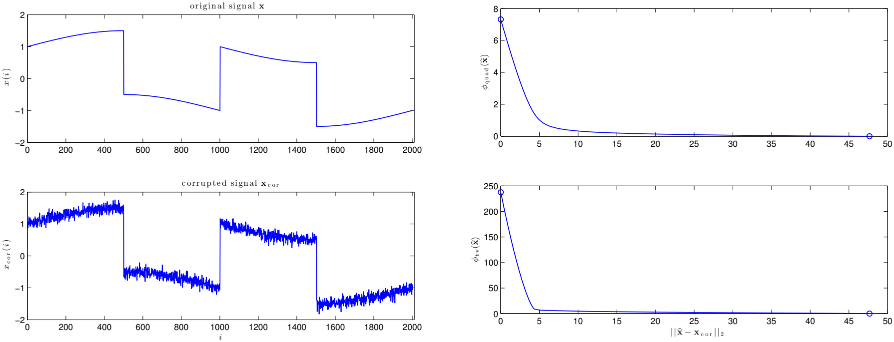
Quadratic Smoothing Reconstruction Example



- Left: Original signal $\mathbf{x}$ and noisy signal $\mathbf{x}_{\mathrm{cor}}$.
- Right: Three solutions on the optimal trade-off curve between $\phi_{\mathrm{quad}}(\widehat{\mathbf{x}})$ and $\|\widehat{\mathbf{x}} - \mathbf{x}_{\mathrm{cor}}\|_2$.

Comparing the original signal with the reconstructions, the trade-off between signal fidelity and smoothness is clear.

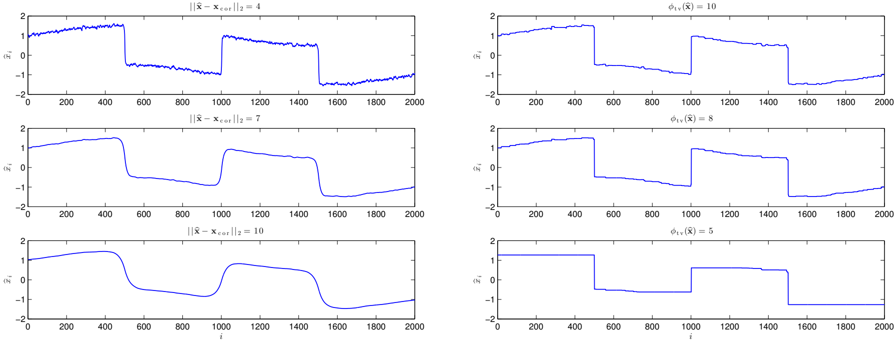
Comparing Quadratic and Total Variation



- Left: Original signal $\mathbf{x}$ and noisy signal $\mathbf{x}_{\text{cor}}$.
- Right: Optimal trade-off curves (between $\phi(\widehat{\mathbf{x}})$ and $||\widehat{\mathbf{x}} - \mathbf{x}_{\text{cor}}||_2$) for quadratic smoothing ϕ_{quad} (top) and total variation smoothing ϕ_{tv} (bottom).

As can be seen, both trade-off curves exhibit a knee, which likely represents a good trade-off between smoothing and signal fidelity.

Comparing Quadratic and Total Variation (Continued)



- Left: Quadratic smoothing - three solutions on the optimal trade-off curve between $\phi_{\text{quad}}(\widehat{\mathbf{x}})$ and $||\widehat{\mathbf{x}} - \mathbf{x}_{\text{cor}}||_2$.
- Right: Total variation smoothing - three solutions on the optimal trade-off curve between $\phi_{\text{tv}}(\widehat{\mathbf{x}})$ and $||\widehat{\mathbf{x}} - \mathbf{x}_{\text{cor}}||_2$.

As can be seen, quadratic smoothing will smooth out both noise and sharp transitions in the signal. Total variation smoothing, on the other hand, preserves sharp transitions in the signal.