EE/ACM 150 - Applications of Convex Optimization in Signal Processing and Communications
Lecture 14

Andre Tkacenko

Signal Processing Research Group
Jet Propulsion Laboratory

May 17, 2012

Caltech

Outline

Introduction to Stochastic Robust Approximation

For the approximation problem with basic objective $||\mathbf{A}\mathbf{x} - \mathbf{b}||$, it is often desirable to account for uncertainty in the data matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$. To design a model fit robust to variations in the data, it is natural to use the expected value $E[\cdot]$ of the basic objective.

Stochastic Robust Approximation Problem:

$$\text{minimize} \quad E[||\mathbf{A}\mathbf{x} - \mathbf{b}||] \quad .$$

- This problem is always a convex optimization problem.
- It is often not tractable as it is difficult to evaluate the objective and its derivatives.

Sum-of-Norms Problem: One special tractable case occurs when $\mathbf{A}$ assumes only a finite number of values for which

$$\Pr\{\mathbf{A} = \mathbf{A}_i\} = p_i \, , \; i = 1, \ldots, k \, .$$

In this case, the problem becomes

$$\text{minimize} \quad p_1 \, ||\mathbf{A}_1\mathbf{x} - \mathbf{b}|| + \cdots + p_k \, ||\mathbf{A}_k\mathbf{x} - \mathbf{b}|| \quad ,$$

which is equivalent to

$$\begin{array}{ll} \text{minimize} & \mathbf{p}^T\mathbf{t} \\ \text{subject to} & ||\mathbf{A}_i\mathbf{x} - \mathbf{b}|| \leq t_i \, , \; i = 1, \ldots, k \end{array} \quad .$$

Relation to Tikhonov Regularization

A common way to model the uncertainty in the data matrix $\mathbf{A}$ is to decompose it as

$$\mathbf{A} = \overline{\mathbf{A}} + \mathbf{U}\,,$$

where $\overline{\mathbf{A}}$ is the mean of $\mathbf{A}$ (i.e., $\overline{\mathbf{A}} = E[\mathbf{A}]$) and $\mathbf{U}$ is a random matrix with zero mean. Typically, we assume $\overline{\mathbf{A}}$ is known as well as other information such as the covariance matrix of $\mathbf{U}^T$ (and hence $\mathbf{A}^T$) given by $\mathbf{P} = E\left[\mathbf{U}^T\mathbf{U}\right] = E\left[\mathbf{A}^T\mathbf{A}\right]$.

Statistical Robust Least-Squares Problem: One variation of the above problem is

$$\text{minimize} \quad E\left[||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2\right] \quad .$$

This can be shown to be identical to the problem

$$\text{minimize} \quad \left|\left|\overline{\mathbf{A}}\mathbf{x} - \mathbf{b}\right|\right|_2^2 + \left|\left|\mathbf{P}^{1/2}\mathbf{x}\right|\right|_2^2 \quad ,$$

which has the solution

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

- This corresponds to *Tikhonov regularization* with data matrix $\overline{\mathbf{A}}$ and $\mathbf{\Gamma} = \mathbf{P}^{1/2}$.
- When $\mathbf{\Gamma} = \sqrt{\delta}\mathbf{I}_n$, the interpretation is that $U_{k,\ell}$ are zero mean, uncorrelated random variables with variance δ/m.

Introduction to Worst-Case Robust Approximation

Another way to account for variation in the data matrix $\mathbf{A}$ is to use a set-based, worst-case approach. Here, the uncertainty in $\mathbf{A}$ is described by a set $\mathcal{A} \subseteq \mathbb{R}^{m \times n}$ of possible values. The *worst-case error* of a candidate solution $\mathbf{x} \in \mathbb{R}^n$ is defined as

$$\xi_{\text{wc}}(\mathbf{x}) \triangleq \sup \{||\mathbf{Ax} - \mathbf{b}|| : \mathbf{A} \in \mathcal{A}\} ,$$

and is always a convex function of $\mathbf{x}$.

Worst-Case Robust Approximation Problem:

$$\text{minimize} \quad \xi_{\text{wc}}(\mathbf{x}) = \sup \{||\mathbf{Ax} - \mathbf{b}|| : \mathbf{A} \in \mathcal{A}\} \quad .$$

- This problem is always a convex optimization problem.
- Its tractability depends on the norm used and the uncertainty set $\mathcal{A}$.

Finite/Polyhedral Set: If $\mathcal{A} = \{\mathbf{A}_1, \ldots, \mathbf{A}_k\}$ or $\mathcal{A} = \text{conv}(\{\mathbf{A}_1, \ldots, \mathbf{A}_k\})$, the problem is equivalent to

$$\text{minimize} \quad \max_{i=1,\ldots,k} \{||\mathbf{A}_i\mathbf{x} - \mathbf{b}||\} \quad .$$

Using an epigraph form, the problem can be in turn recast as

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & ||\mathbf{A}_i\mathbf{x} - \mathbf{b}|| \leq t \, , \; i = 1, \ldots, k \end{array} \quad .$$

Norm Bound Error

Suppose the uncertainty set $\mathcal{A}$ is a norm ball of the form

$$\mathcal{A} \triangleq \left\{ \overline{\mathbf{A}} + \mathbf{U} : ||\mathbf{U}|| \leq a \right\} ,$$

where $||\cdot||$ is a norm on $\mathbb{R}^{m \times n}$. In this case, we have

$$\xi_{\text{wc}}(\mathbf{x}) = \sup \left\{ \left|\left| \overline{\mathbf{A}}\mathbf{x} - \mathbf{b} + \mathbf{U}\mathbf{x} \right|\right| : ||\mathbf{U}|| \leq a \right\} .$$

Note that the first norm above is on $\mathbb{R}^m$, whereas the second one is on $\mathbb{R}^{m \times n}$.

Euclidean norm on $\mathbb{R}^m$ / maximum singular value norm on $\mathbb{R}^{m \times n}$: Here, the supremum is attained for

$$\mathbf{U} = a\mathbf{u}\mathbf{v}^T , \text{ where } \mathbf{u} = \frac{\overline{\mathbf{A}}\mathbf{x} - \mathbf{b}}{\left|\left| \overline{\mathbf{A}}\mathbf{x} - \mathbf{b} \right|\right|_2} , \ \mathbf{v} = \frac{\mathbf{x}}{||\mathbf{x}||_2} \Longrightarrow \xi_{\text{wc}}(\mathbf{x}) = \left|\left| \overline{\mathbf{A}}\mathbf{x} - \mathbf{b} \right|\right|_2 + a \, ||\mathbf{x}||_2 .$$

Thus, the robust approximation problem becomes

$$\text{minimize} \quad \left|\left| \overline{\mathbf{A}}\mathbf{x} - \mathbf{b} \right|\right|_2 + a \, ||\mathbf{x}||_2 \quad ,$$

which can be equivalently expressed as the SOCP

$$\begin{array}{ll} \text{minimize} & t_1 + a t_2 \\ \text{subject to} & \left|\left| \overline{\mathbf{A}}\mathbf{x} - \mathbf{b} \right|\right|_2 \leq t_1 , \ ||\mathbf{x}||_2 \leq t_2 \end{array} .$$

The solution of this problem is the same as that of the regularized least-squares problem

$$\text{minimize} \quad \left|\left| \overline{\mathbf{A}}\mathbf{x} - \mathbf{b} \right|\right|_2^2 + \delta \, ||\mathbf{x}||_2^2 \quad ,$$

for some value of the regularization parameter δ.

Uncertainty Ellipsoids

Suppose we describe the variation in $\mathbf{A}$ by giving an ellipsoid of possible values for each row:

$$\mathcal{A} \triangleq \left\{ \mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_m \end{bmatrix}^T : \mathbf{a}_i \in \mathcal{E}_i \, , \; i = 1, \ldots, m \right\} ,$$

where

$$\mathcal{E}_i = \left\{ \overline{\mathbf{a}}_i + \mathbf{P}_i \mathbf{u} : ||\mathbf{u}||_2 \leq 1 \right\} .$$

Here, $\overline{\mathbf{a}}_i$ denotes the nominal value of $\mathbf{a}_i$ and $\mathbf{P}_i$ describes its variation. In this case, we have

$$\sup_{\mathbf{a}_i \in \mathcal{E}_i} \left| \mathbf{a}_i^T \mathbf{x} - b_i \right| = \sup \left\{ \left| \overline{\mathbf{a}}_i^T \mathbf{x} - b_i + \left(\mathbf{P}_i^T \mathbf{x} \right)^T \mathbf{u} \right| : ||\mathbf{u}||_2 \leq 1 \right\} = \left| \overline{\mathbf{a}}_i^T \mathbf{x} - b_i \right| + \left|\left| \mathbf{P}_i^T \mathbf{x} \right|\right|_2 .$$

For the robust ℓ_2-norm approximation problem in which $\xi_{\text{wc}}(\mathbf{x}) = \sup_{\mathbf{A} \in \mathcal{A}} \left\{ ||\mathbf{A}\mathbf{x} - \mathbf{b}||_2 \right\}$, we have

$$\xi_{\text{wc}}(\mathbf{x}) = \sqrt{\sum_{i=1}^{m} \left(\sup_{\mathbf{a}_i \in \mathcal{E}_i} \left| \mathbf{a}_i^T \mathbf{x} - b_i \right| \right)^2} = \sqrt{\sum_{i=1}^{m} \left(\left| \overline{\mathbf{a}}_i^T \mathbf{x} - b_i \right| + \left|\left| \mathbf{P}_i^T \mathbf{x} \right|\right|_2 \right)^2} .$$

It follows that the robust ℓ_2-norm approximation problem can be expressed as the SOCP

$$\begin{array}{ll} \text{minimize} & s \\ \text{subject to} & ||\mathbf{t}||_2 \leq s \\ & \left|\left| \mathbf{P}_i^T \mathbf{x} \right|\right|_2 \leq t_i - \left(\overline{\mathbf{a}}_i^T \mathbf{x} - b_i \right) \, , \; i = 1, \ldots, m \\ & \left|\left| \mathbf{P}_i^T \mathbf{x} \right|\right|_2 \leq t_i + \left(\overline{\mathbf{a}}_i^T \mathbf{x} - b_i \right) \, , \; i = 1, \ldots, m \end{array} .$$

Norm Bounded Error with Linear Structure

As a generalization of the norm bound description $\mathcal{A} = \left\{\overline{\mathbf{A}} + \mathbf{U} : ||\mathbf{U}|| \leq a\right\}$, we can consider the image of the norm ball under an affine transformation:

$$\mathcal{A} \triangleq \left\{\overline{\mathbf{A}} + u_1\mathbf{A}_1 + \cdots + u_p\mathbf{A}_p : ||\mathbf{u}|| \leq 1\right\} ,$$

where $||\cdot||$ is a norm on $\mathbb{R}^p$ and the $p+1$ matrices $\overline{\mathbf{A}}, \mathbf{A}_1, \ldots, \mathbf{A}_p$ are given. Then, the worst-case error can be expressed as

$$\xi_{\mathrm{wc}}(\mathbf{x}) = \sup_{||\mathbf{u}|| \leq 1} \left|\left|\left(\overline{\mathbf{A}} + u_1\mathbf{A}_1 + \cdots + u_p\mathbf{A}_p\right)\mathbf{x} - \mathbf{b}\right|\right| = \sup_{||\mathbf{u}|| \leq 1} ||\mathbf{P}(\mathbf{x})\,\mathbf{u} + \mathbf{q}(\mathbf{x})|| ,$$

where $\mathbf{P}(\mathbf{x}) = \begin{bmatrix} \mathbf{A}_1\mathbf{x} & \cdots & \mathbf{A}_p\mathbf{x} \end{bmatrix} \in \mathbb{R}^{m \times p}$ and $\mathbf{q}(\mathbf{x}) = \overline{\mathbf{A}}\mathbf{x} - \mathbf{b} \in \mathbb{R}^m$.

Robust Chebyshev approximation problem: Here we have

$$\begin{aligned} \xi_{\mathrm{wc}}(\mathbf{x}) &= \sup_{||\mathbf{u}||_\infty \leq 1} ||\mathbf{P}(\mathbf{x})\,\mathbf{u} + \mathbf{q}(\mathbf{x})||_\infty = \max_{i=1,\ldots,m} \sup_{||\mathbf{u}||_\infty \leq 1} \left|\mathbf{p}_i(\mathbf{x})^T\,\mathbf{u} + q_i(\mathbf{x})\right| , \\ &= \max_{i=1,\ldots,m} \left\{||\mathbf{p}_i(\mathbf{x})||_1 + |q_i(\mathbf{x})|\right\} . \end{aligned}$$

Hence, the robust Chebyshev approximation problem can be cast as the following LP.

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & -\mathbf{y}_0 \preceq \overline{\mathbf{A}}\mathbf{x} - \mathbf{b} \preceq \mathbf{y}_0 \\ & -\mathbf{y}_k \preceq \mathbf{A}_k\mathbf{x} \preceq \mathbf{y}_k \,,\; k = 1, \ldots, p \\ & \mathbf{y}_0 + \sum_{k=1}^{p} \mathbf{y}_k \preceq t\mathbf{1} \end{array} .$$

Norm Bounded Error with Linear Structure (Continued)

Robust least-squares approximation problem: Here we have

$$\xi_{\rm wc}(\mathbf{x}) = \sup_{||\mathbf{u}||_2 \le 1} ||\mathbf{P}(\mathbf{x})\,\mathbf{u} + \mathbf{q}(\mathbf{x})||_2 \ ,$$

where $\mathbf{P}(\mathbf{x}) = \begin{bmatrix} \mathbf{A}_1\mathbf{x} & \cdots & \mathbf{A}_p\mathbf{x} \end{bmatrix} \in \mathbb{R}^{m \times p}$ and $\mathbf{q}(\mathbf{x}) = \overline{\mathbf{A}}\mathbf{x} - \mathbf{b} \in \mathbb{R}^m$.

Evaluating the worst-case error $\xi_{\rm wc}(\mathbf{x})$ can actually by simplified by noting that it is the square root of the optimal value of the (*nonconvex*) quadratic optimization problem

$$\begin{array}{ll} \text{maximize} & ||\mathbf{P}(\mathbf{x})\,\mathbf{u} + \mathbf{q}(\mathbf{x})||_2^2 \\ \text{subject to} & \mathbf{u}^T\mathbf{u} \le 1 \end{array} ,$$

with variable $\mathbf{u} \in \mathbb{R}^p$. This is a special nonconvex problem for which strong duality holds. The Lagrange dual of this problem can be expressed as the SDP

$$\begin{array}{ll} \text{minimize} & t + \lambda \\ \text{subject to} & \begin{bmatrix} \mathbf{I}_m & \mathbf{P}(\mathbf{x}) & \mathbf{q}(\mathbf{x}) \\ \mathbf{P}(\mathbf{x})^T & \lambda \mathbf{I}_p & \mathbf{0}_{p \times 1} \\ \mathbf{q}(\mathbf{x})^T & \mathbf{0}_{1 \times p} & t \end{bmatrix} \succeq \mathbf{0} \end{array} ,$$

with variables $t, \lambda \in \mathbb{R}$. So, for fixed $\mathbf{x}$, we can compute $(\xi_{\rm wc}(\mathbf{x}))^2$ by solving this SDP with variables t, λ. But optimizing jointly over t, λ, and $\mathbf{x}$ is equivalent to minimizing $(\xi_{\rm wc}(\mathbf{x}))^2$ over $\mathbf{x}$. Hence, we can solve the robust least-squares problem by solving the SDP with $\mathbf{x}, t, \lambda$ as variables. The problem is still an SDP when optimizing over $\mathbf{x}$ as $\mathbf{P}(\mathbf{x})$ and $\mathbf{q}(\mathbf{x})$ are *affine* in $\mathbf{x}$.

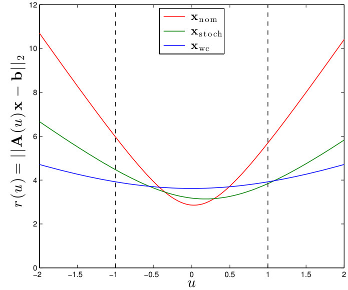
Parameterized Example

Here, our uncertainty in $\mathbf{A}$ is parameterized as

$$\mathbf{A}(u) = \mathbf{A}_0 + u\mathbf{A}_1 \,,$$

where u is an uncertainty parameter. For a specific instance with $\mathbf{A}(u) \in \mathbb{R}^{10\times 20}$, $||\mathbf{A}_0||_2 = 10$, $||\mathbf{A}_1||_2 = 1$, and $u \in [-1, 1]$ (yielding a variation of around $\pm 10\%$), we considered three solutions:

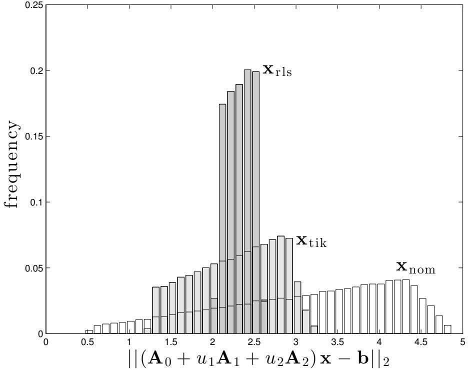
- *Nominal optimal:* $\mathbf{x}_{\mathrm{nom}}$ found by minimizing $||\mathbf{A}_0\mathbf{x} - \mathbf{b}||_2$.
- *Stochastic robust approximation:* $\mathbf{x}_{\mathrm{stoch}}$ found by minimizing $E\left[||\mathbf{A}_0\mathbf{x} - \mathbf{b}||_2^2\right]$, assuming $u \sim \mathcal{U}[-1, 1]$.
- *Worst-case robust approximation:* $\mathbf{x}_{\mathrm{wc}}$ found by minimizing $\sup_{-1\le u\le 1} ||\mathbf{A}(u)\,\mathbf{x} - \mathbf{b}||_2 = \max\left\{||(\mathbf{A}_0 - \mathbf{A}_1)\,\mathbf{x} - \mathbf{b}||_2 \,, ||(\mathbf{A}_0 + \mathbf{A}_1)\,\mathbf{x} - \mathbf{b}||_2\right\}$.



Monte Carlo Simulation Example

$$\text{minimize} \quad \sup_{||\mathbf{u}||_2 \leq 1} ||(\mathbf{A}_0 + u_1 \mathbf{A}_1 + u_2 \mathbf{A}_2)\,\mathbf{x} - \mathbf{b}||_2 \quad .$$

- $m = 50$, $n = 20$, $||\mathbf{A}_0||_2 = 10$, $||\mathbf{A}_1||_2 = ||\mathbf{A}_2||_2 = 1$, u_1, u_2 in unit disk in $\mathbb{R}^2$.
- Compared nominal least-squares $\mathbf{x}_{\text{nom}}$ ($\mathbf{u} = \mathbf{0}$), Tikhonov regularized $\mathbf{x}_{\text{tik}}$ ($\delta = 1$), and worst-case robust least-squares $\mathbf{x}_{\text{rls}}$ solutions.
- Simulated 10^5 parameter vectors $\mathbf{u}$ and generated histograms of the residual $||(\mathbf{A}_0 + u_1 \mathbf{A}_1 + u_2 \mathbf{A}_2)\,\mathbf{x} - \mathbf{b}||_2$ for each parameter value.



Introduction and Examples of Function Families

Consider a family of functions (often called *basis functions*) $f_1, \ldots, f_n : \mathbb{R}^k \to \mathbb{R}$ with common domain $\mathrm{dom}(f_i) = \mathcal{D}$. For some $\mathbf{x} \in \mathbb{R}^n$, we are interested in a function $f : \mathbb{R}^k \to \mathbb{R}$ of the form

$$f(\mathbf{u}) = x_1 f_1(\mathbf{u}) + \cdots + x_n f_n(\mathbf{u}) \ ,$$

where $\mathbf{u} \in \mathbb{R}^k$. Here, $\mathbf{x}$ is a *coefficient vector* applied to the family of basis functions $\{f_1, \ldots, f_n\}$ to produce the function f. As $\mathrm{dom}(f) = \mathcal{D}$, the basis functions generate a subspace $\mathcal{F}$ of functions on $\mathcal{D}$.

Examples:

- *Polynomials:*

$$f_\ell(t) = t^{\ell-1}$$ (powers)

$$\int f_\ell(t)\, f_m(t)\, \phi(t)\ dt = \delta[\ell - m]$$ (orthonormal, $\phi \geq 0$, $\deg(f_i) \leq n$)

$$f_\ell(t) = \prod_{\substack{i=1 \\ i \neq \ell}}^{n} \frac{t - t_i}{t_\ell - t_i} \ , \ f_\ell(t_m) = \delta[\ell - m]$$ (Lagrange interpolation polynomial) (for distinct points $\{t_1, \ldots, t_m\}$)

- *Piecewise polynomials and splines:* The family of functions can be piecewise polynomial over $t \in \mathbb{R}$ with knots at $t_1 \leq t_2 \leq \cdots \leq t_m$. If the functions satisfy differentiability conditions at the knot points, they are referred to as *B-splines*.

Examples of Convex Constraints on the Coefficients

In terms of the coefficient vector $\mathbf{x}$, there are several relevant constraints which will be useful to enforce which are also convex. This follows from the fact that we have

$$f(\mathbf{u}) = \mathbf{f}(\mathbf{u})^T \mathbf{x}\,, \text{ where } \mathbf{f}(\mathbf{u}) \triangleq \begin{bmatrix} f_1(\mathbf{u}) & \cdots & f_n(\mathbf{u}) \end{bmatrix}^T .$$

- **Function value interpolation and inequalities:**

$$f(\mathbf{u}_\ell) = z_\ell \iff \mathbf{f}(\mathbf{u}_\ell)^T \mathbf{x} = z_\ell \qquad \text{(Interpolation condition)}$$

$$|f(\mathbf{u}_\ell) - f(\mathbf{u}_m)| \le L\,||\mathbf{u}_\ell - \mathbf{u}_m|| \iff$$
$$-L\,||\mathbf{u}_\ell - \mathbf{u}_m|| \le \left(f(\mathbf{u}_\ell) - f(\mathbf{u}_m)\right)^T \mathbf{x} \le L\,||\mathbf{u}_\ell - \mathbf{u}_m|| \qquad \text{(Lipschitz continuity)}$$

$$f(\mathbf{u}) \ge 0 \;\; \forall \mathbf{u} \in \mathcal{D} \qquad \text{(Nonnegativity constraint)}$$

- **Derivative constraints:** As $\nabla f(v) = \sum_{i=1}^n x_i \nabla f_i(\mathbf{v})$ and $\nabla^2 f(v) = \sum_{i=1}^n x_i \nabla^2 f_i(\mathbf{v})$, both of which are linear in $\mathbf{x}$, constraints of the form

$$||\nabla f(\mathbf{v})|| \le M\,,\; \boldsymbol{\ell}\mathbf{I} \preceq \nabla^2 f(\mathbf{v}) \preceq \mathbf{u}\mathbf{I}\,,$$

  are convex in $\mathbf{x}$.

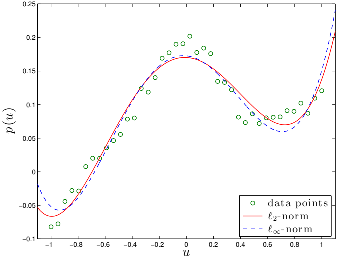
- **Integral constraints:**

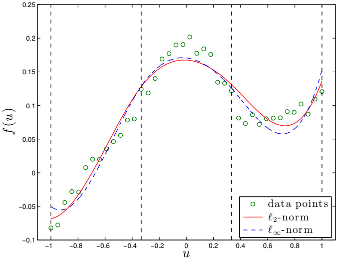
$$\int_{\mathcal{D}} \phi(\mathbf{u})\, f(\mathbf{u}) \begin{cases} = a \\ \ge a \\ \le a \end{cases} \Longrightarrow \mathbf{c}^T\mathbf{x} \begin{cases} = a \\ \ge a \\ \le a \end{cases},\ \text{where } c_i = \int_{\mathcal{D}} \phi(\mathbf{u})\, f_i(\mathbf{u}) \,.$$

- **Linear functional constraints:** Any linear functional $\mathcal{L}$ on the subspace of functions is a linear function of $\mathbf{x}$, i.e., $\mathcal{L}(f) = \mathbf{c}^T\mathbf{x}$ for some $\mathbf{c} \in \mathbb{R}^n$. The derivative and integral constraints are special cases of this.

Types of Fitting Problems

- **Minimum norm function fitting:** Given data $(\mathbf{u}_1, y_1), \ldots, (\mathbf{u}_m, y_m)$ find a function f that minimizes $||\mathbf{f} - \mathbf{y}||$, where $\mathbf{f} \triangleq \begin{bmatrix} f(\mathbf{u}_1) & \cdots & f(\mathbf{u}_m) \end{bmatrix}^T$. Here, $m \geq n$ and typically $m \gg n$.
- **Least-norm interpolation:** In this case, we have $m < n$. We must satisfy the interpolation conditions $f(\mathbf{u}_\ell) = y_\ell$ for $\ell = 1, \ldots, m$ and we may seek a function that is smoothest or smallest in some sense.
- **Interpolation, extrapolation, and bounding:** An optimal function fit $\widehat{\mathbf{f}}$ allows us to guess the value of the function at some point $\mathbf{v}$ not in the original data set. This is *interpolation* if $\mathbf{v} \in \text{conv}(\{\mathbf{u}_1, \ldots, \mathbf{u}_m\})$ and *extrapolation* otherwise. From this, we can produce bounds on $f(\mathbf{v})$ by minimizing and maximizing $f(\mathbf{v})$ subject to the constraints. This can be useful for identifying outliers.





Introduction to Basis Pursuit

For many applications, the function family consists of a large number of linearly dependent functions referred to as an *over-complete basis* or *dictionary*). In *basis pursuit* (BP), we seek to model the data of interest $(\mathbf{u}_i, y_i)$ for $i = 1, \ldots, m$ using only a small number of basis functions. Specifically, we seek a function $f(\mathbf{u}) = \mathbf{f}(\mathbf{u})^T \mathbf{x} \in \mathcal{F}$ that fits the given data well,

$$f(\mathbf{u}_i) \approx y_i \,, \; i = 1, \ldots, m \,,$$

with a *sparse* coefficient vector $\mathbf{x}$, i.e., one for which $\mathrm{card}(\mathbf{x})$ is small. In this case, $f = \sum_{i \in \mathcal{B}} x_i f_i$, where $\mathcal{B} = \{i : x_i \neq 0\}$, is a *sparse description* of the data.

Uses:

- *Data compression:* The transmitter and receiver both know the dictionary and so the signal of interest can be conveyed using only the information present in the sparse vector $\mathbf{x}$.
- *De-noising or smoothing:* A parsimonious representation of the signal in terms of the basis functions can effect a removal of noise and a smoothing of the data.

Common Approach to Basis Pursuit (BP):

1. *ℓ_1-norm regularization heuristic:* Minimize $\displaystyle\sum_{i=1}^{m} \left(\mathbf{f}(\mathbf{u}_i)^T \mathbf{x} - y_i \right)^2 + \gamma \, ||\mathbf{x}||_1$, for some $\gamma > 0$.
2. *Determine sparsity pattern:* Identify $\mathcal{B} = \{i : \widehat{x}_i \neq 0\}$, where $\widehat{\mathbf{x}}$ is the solution from Step 1.
3. *Refinement step:* Solve the least-squares problem: minimize $\displaystyle\sum_{i=1}^{m} \left(\mathbf{f}(\mathbf{u}_i)^T \mathbf{x} - y_i \right)^2$, with variables x_i for $i \in \mathcal{B}$ and $x_i = 0$ for $i \notin \mathcal{B}$.

Time-Frequency Analysis Via BP

When the dictionary functions are localized in time and frequency, basis pursuit can be used to perform a time-frequency analysis of the signal of interest, in addition to obtaining a sparse representation in terms of the dictionary.

Example: Suppose each basis function is formed from the *Gabor function* given by

$$g_{\tau,\omega}(t) = e^{-\frac{(t-\tau)^2}{\sigma^2}} \cos(\omega t + \phi) \ .$$

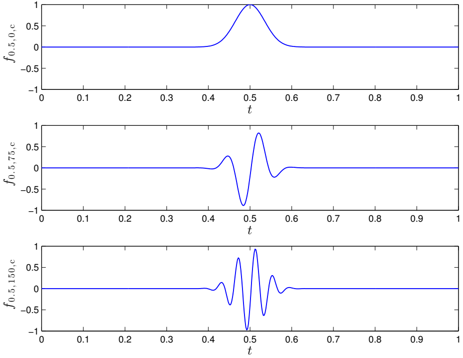
Note that $g_{\tau,\omega}(t)$ is *localized* at time τ and frequency ω.

- Range of interest is $t \in [0, 1]$.
- All basis functions have width $\sigma = 0.05$.
- Dictionary consists of pulse times and frequencies

$$\tau = 0.002k \ , \ k = 0, \ldots, 500 \ ,$$
$$\omega = 5k \ , \ k = 0, \ldots, 30 \ ,$$
$$\phi = 0, \pi/2 \ \text{(cosine and sine)} \ .$$

- The $501 \times 61 = 30561$ basis elements are denoted as

$$f_{\tau,\omega,\mathrm{c}} \ , \ \tau = 0, 0.002, \ldots, 1 \ , \ \omega = 0, 5, \ldots, 150 \ ,$$
$$f_{\tau,\omega,\mathrm{s}} \ , \ \tau = 0, 0.002, \ldots, 1 \ , \ \omega = 5, \ldots, 150 \ .$$



Time-Frequency Analysis Via BP (Continued)

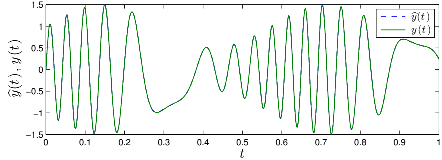
We will use BP to find a sparse representation of the signal

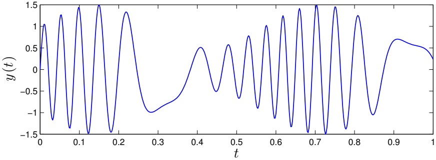
$$y(t) = a(t)\sin\theta(t) \ , \text{ where } a(t) = 1 + 0.5\sin(11t) \ , \ \theta(t) = 30\sin(5t) \ .$$

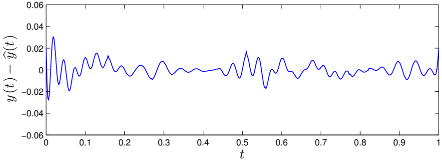
Here, $a(t)$ is the signal amplitude and $\theta(t)$ is the total phase. We can also interpret $\omega(t) = |d\theta/dt| = 150\,|\cos(5t)|$, as the *instantaneous frequency* of the signal at time t. Given the 501 data points (t_k, y_k) with

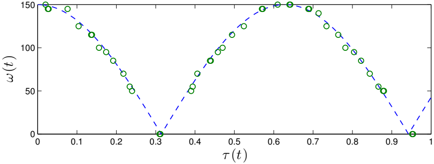
$$t_k = 0.005k \, , \ y_k = y(t_k) \ , \ k = 0, \ldots, 500 \, ,$$

we obtained the following by performing BP, which yielded a very sparse coefficient vector with only 42 nonzero coefficients out of 30561.









Using Subgradients to Fit a Convex Function to Data

In many applications, it may be desirable to attempt to fit a convex function to a given set of data. For some special cases, we can solve interpolation problems involving an infinite-dimensional set of functions using finite-dimensional convex optimization.

Convex Interpolation and Subgradients:

A convex function $f : \mathbb{R}^k \to \mathbb{R}$, with $\mathrm{dom}(f) = \mathbb{R}^k$ satisfies the interpolation conditions

$$f(\mathbf{u}_i) = y_i \,,\ i = 1, \ldots, m \,,$$

at given points $\mathbf{u}_i \in \mathbb{R}^k$ if and only if there exist $\mathbf{g}_i \in \mathbb{R}^k$ for $i = 1, \ldots, m$ such that

$$y_j \geq y_i + \mathbf{g}_i^T (\mathbf{u}_j - \mathbf{u}_i) \ ,\ i, j = 1, \ldots, m \,.$$

- Because of the direction of each inequality, the vector $\mathbf{g}_i$ is called a *subgradient*.
- If f is differentiable, we can take $\mathbf{g}_i = \nabla f(\mathbf{u}_i)$.

Proof:

- *Necessity:* Suppose f is convex with $\mathrm{dom}(f) = \mathbb{R}^k$ and $f(\mathbf{u}_i) = y_i$. Then, at each $\mathbf{u}_i$, we can find a vector $\mathbf{g}_i$ such that
$$f(\mathbf{z}) \geq f(\mathbf{u}_i) + \mathbf{g}_i^T (\mathbf{z} - \mathbf{u}_i) \ .$$
Setting $\mathbf{z} = \mathbf{u}_j$, we obtain the desired inequality.
- *Sufficiency:* Conversely suppose $\mathbf{g}_1, \ldots, \mathbf{g}_m$ satisfy the above inequality. Defining f as
$$f(\mathbf{z}) \triangleq \max_{i=1,\ldots,m} \left\{ y_i + \mathbf{g}_i^T (\mathbf{z} - \mathbf{u}_i) \right\} \ ,$$
we note that f is convex and that $f(\mathbf{u}_i) = y_i$ for $i = 1, \ldots, m$.

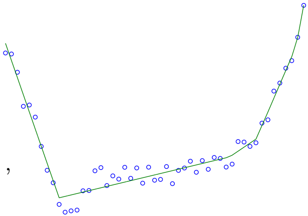
Fitting a Convex Function to Given Data

One simple application of the above result is to compute the least-squares fit of a convex function to given data $(\mathbf{u}_i, y_i)$ for $i = 1, \ldots, m$:

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{m} (y_i - f(\mathbf{u}_i))^2 \\ \text{subject to} & f : \mathbb{R}^k \to \mathbb{R} \text{ is convex}, \ \mathrm{dom}(f) = \mathbb{R}^k \end{array} .$$

This infinite-dimensional problem can be converted to the finite-dimensional QP

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{m} (y_i - \widehat{y}_i)^2 \\ \text{subject to} & \widehat{y}_j \geq \widehat{y}_i + \mathbf{g}_i^T(\mathbf{u}_j - \mathbf{u}_i) \ , \ i, j = 1, \ldots, m \end{array} ,$$



with variables $\widehat{\mathbf{y}} \in \mathbb{R}^m$ and $\mathbf{g}_1, \ldots, \mathbf{g}_m \in \mathbb{R}^k$.

Variations and Extensions:

- *Bounding values of an interpolating convex function:* To determine bounds on $f(\mathbf{u}_0)$ given data $(\mathbf{u}_i, y_i)$ that can be interpolated by a convex function, we solve the two LPs

$$\begin{array}{ll} \text{minimize/maximize} & y_0 \\ \text{subject to} & y_j \geq y_i + \mathbf{g}_i^T(\mathbf{u}_j - \mathbf{u}_i) \ , \ i, j = 0, \ldots, m \end{array} ,$$

  with variables $y_0 \in \mathbb{R}$, $\mathbf{g}_0, \ldots, \mathbf{g}_m \in \mathbb{R}^k$.
- *Interpolation with monotone convex functions:* For f to also by monotone nondecreasing (i.e., $f(\mathbf{u}) \geq f(\mathbf{v})$ whenever $\mathbf{u} \succeq \mathbf{v}$), each subgradient $\mathbf{g}_i$ must also satisfy $\mathbf{g}_i \succeq \mathbf{0}$.