Low-rank and sparse methods for high-dimensional approximation and model order reduction

---

Lecture 3

Tensors - Sparse tensors

Outline

Outline

Tensor product of vectors

For $I = \{1, \ldots, N\}$, an element v of the vector space $\mathbb{R}^I$ is identified with the set of its coefficients $(v_i)_{i \in I}$ on a certain basis $\{e_i\}_{i \in I}$ of $\mathbb{R}^I$,

$$v = \sum_{i \in I} v_i e_i.$$

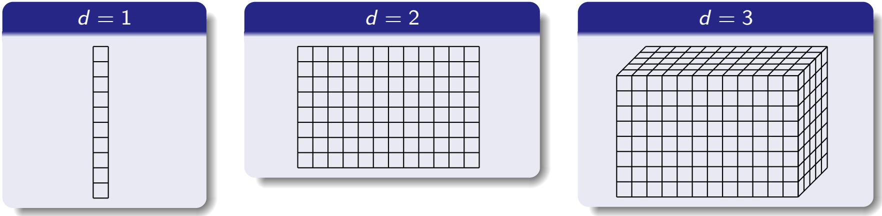
Given d index sets $I^\nu = \{1, \ldots, N_\nu\}$, $1 \leq \nu \leq d$, we introduce the multi-index set

$$I = I_1 \times \ldots \times I_d.$$

An element v of $\mathbb{R}^I$ is called a tensor of order d and is identified with a multidimensional array

$$(v_i)_{i \in I} = (v_{i_1,\ldots,i_d})_{i_1 \in I_1, \ldots, i_d \in I_d}$$

which represents the coefficients of v on a certain basis of $\mathbb{R}^I$.



Tensor product of vectors

The entries of the multidimensional array are equivalently denoted

$$v(i) = v(i_1, \ldots, i_d).$$

Given d vectors $v^{(\nu)} \in \mathbb{R}^{I_\nu}$, $1 \leq \nu \leq d$, the tensor product of these vectors

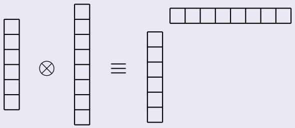
$$v := v^{(1)} \otimes \ldots \otimes v^{(d)}$$

is defined by

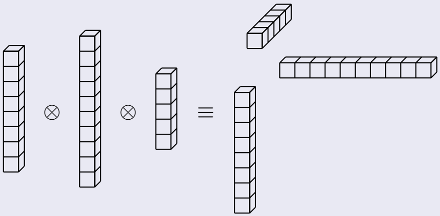
$$v(i) = v^{(1)}(i_1) \ldots v^{(d)}(i_d).$$

Such a tensor is called an elementary tensor. For $d = 2$, using matrix notations, $v \otimes w$ is identified with the matrix vw^T.

$d = 2$



$d = 3$



Tensor product of vectors

The tensor space $\mathbb{R}^I = \mathbb{R}^{I_1 \times \dots \times I_d}$, also denoted $\mathbb{R}^{I_1} \otimes \dots \otimes \mathbb{R}^{I_d}$, is defined by

$$\mathbb{R}^I = \mathbb{R}^{I_1} \otimes \dots \otimes \mathbb{R}^{I_d} = \text{span}\{v^{(1)} \otimes \dots \otimes v^{(d)} : v^{(\nu)} \in \mathbb{R}^{I_\nu}, 1 \leq \nu \leq d\}$$

A canonical basis for $\mathbb{R}^I$ is given by

$$e_i = e_{i_1} \otimes \dots \otimes e_{i_d}, \quad i = (i_1, \dots, i_d) \in I.$$

where

$$e_i(j) = e_{i_1}(j_1) \dots e_{i_d}(j_d) = \delta_{i_1,j_1} \dots \delta_{i_d,j_d} = \delta_{i,j}, \quad \forall j \in I.$$

A tensor $v \in \mathbb{R}^I$ can be written

$$v = \sum_{i \in I} v(i) e_i = \sum_{i_1 \in I_1} \dots \sum_{i_d \in I_d} v(i_1, \dots, i_d) e_{i_1} \otimes \dots \otimes e_{i_d}$$

A natural norm on $\mathbb{R}^I$ is given by

$$\|v\| = \sqrt{\sum_{i \in I} v(i)^2}$$

which makes $\mathbb{R}^I$ a Hilbert space.

Tensor product of functions

Let $\mathcal{X}_\nu \subset \mathbb{R}$, $1 \leq \nu \leq d$, be an interval and V_ν be a space of functions defined on $\mathcal{X}_\nu$.

The tensor product of functions $v^{(\nu)} \in V_\nu$, denoted

$$v = v^{(1)} \otimes \ldots \otimes v^{(d)},$$

is a multivariate function defined on $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_d$ and such that

$$v(x) = v(x_1, \ldots, x_d) = v^{(1)}(x_1) \ldots v^{(d)}(x_d)$$

for $x = (x_1, \ldots, x_d) \in \mathcal{X}$. For example, for $i \in \mathbb{N}_0^d$, the monomial $x^i = x_1^{i_1} \ldots x_d^{i_d}$ is an elementary tensor.

Tensor product of functions

The algebraic tensor product of spaces V_ν is defined as

$$V_1 \otimes \ldots \otimes V_d = \text{span}\{v^{(1)} \otimes \ldots \otimes v^{(d)} : v^{(\nu)} \in V_\nu, 1 \leq \nu \leq d\}$$

which is the space of multivariate functions v which can be written as a finite linear combination of elementary (separated functions), i.e.

$$v(x) = \sum_{k=1}^{n} v_k^{(1)}(x_1) \ldots v_k^{(d)}(x_d).$$

Up to a formal definition of the tensor product $\otimes$, the above construction can be extended to arbitrary vector spaces V_ν (not only spaces of functions).

Infinite dimensional tensor spaces

For infinite dimensional spaces V_ν, a Hilbert (or Banach) tensor space equipped with a norm $\|\cdot\|$ is obtained by the completion (w.r.t. $\|\cdot\|$) of the algebraic tensor space

$$\overline{V}^{\|\cdot\|} = \overline{V_1 \otimes \ldots \otimes V_d}^{\|\cdot\|}.$$

Example 1 (L^p spaces)

Let $1 \leq p < \infty$. If $V_\nu = L^p_{\mu_\nu}(\mathcal{X}_\nu)$, then

$$L^p_{\mu_1}(\mathcal{X}_1) \otimes \ldots \otimes L^p_{\mu_d}(\mathcal{X}_d) \subset L^p_\mu(\mathcal{X}_1 \times \ldots \times \mathcal{X}_d)$$

with $\mu = \mu_1 \otimes \ldots \otimes \mu_d$, and

$$\overline{L^p_{\mu_1}(\mathcal{X}_1) \otimes \ldots \otimes L^p_{\mu_d}(\mathcal{X}_d)}^{\|\cdot\|} = L^p_\mu(\mathcal{X}_1 \times \ldots \times \mathcal{X}_d)$$

where $\|\cdot\|$ is the natural norm on $L^p_\mu(\mathcal{X}_1 \times \ldots \times \mathcal{X}_d)$.

Example 2 (Bochner spaces)

Let $\mathcal{X}$ be equipped with a finite measure μ, and let W be a Hilbert (or Banach) space. For $1 \leq p < \infty$, the Bochner space $L^p_\mu(\mathcal{X}; W)$ is the set of Bochner-measurable functions $u : \mathcal{X} \to W$ with bounded norm $\|u\|_p = (\int_{\mathcal{X}} \|u(x)\|^p_W \mu(dx))^{1/p}$, and

$$L^p_\mu(\mathcal{X}; W) = \overline{W \otimes L^p_\mu(\mathcal{X})}^{\|\cdot\|_p}.$$

Tensor product basis

If $\{\psi_i^{(\nu)}\}_{i\in I_\nu}$ is a basis of V_ν, then a basis of $V = V_1 \otimes \ldots \otimes V_d$ is given by

$$\left\{\psi_i = \psi_{i_1}^{(1)} \otimes \ldots \otimes \psi_{i_d}^{(d)} : i \in I = I_1 \times \ldots \times I_d\right\}.$$

A tensor $v \in V$ admits a decomposition

$$v = \sum_{i\in I} a(i)\psi_i = \sum_{i_1\in I_1} \ldots \sum_{i_d\in I_d} a(i_1,\ldots,i_d)\psi_{i_1}^{(1)} \otimes \ldots \otimes \psi_{i_d}^{(d)},$$

and v can be identified with the set of its coefficients

$$a \in \mathbb{R}^I.$$

Orthogonal tensor product basis

If the V_ν are Hilbert spaces with inner products $(\cdot,\cdot)_\nu$ and associated norms $\|\cdot\|_\nu$, a canonical inner product on V can be first defined for elementary tensors

$$(v^{(1)} \otimes \ldots \otimes v^{(d)}, w^{(1)} \otimes \ldots \otimes w^{(d)}) = (v^{(1)}, w^{(1)}) \ldots (v^{(d)}, w^{(d)})$$

and then extended by linearity to the whole space V. The associated norm $\|\cdot\|$ is called the canonical norm.

If the $\{\psi_i^{(\nu)}\}_{i\in I_\nu}$ are orthonormal bases of spaces V_ν, then $\{\psi_i\}_{i\in I}$ is an orthonormal basis of $\overline{V}^{\|\cdot\|}$. A tensor

$$v = \sum_{i\in I} a_i \psi_i$$

is such that

$$\|v\| = \sqrt{\sum_{i\in I} a_i^2} := \|a\|.$$

Therefore, the map Ψ which associates to a tensor $a \in \mathbb{R}^I$ the tensor $v = \Psi(a) := \sum_{i\in I} a_i\psi_i$ defines a linear isometry from $\mathbb{R}^I$ to V for finite dimensional spaces, and between $\ell_2(I)$ and $\overline{V}^{\|\cdot\|}$ for infinite dimensional spaces.

Curse of dimensionality

A tensor $a \in \mathbb{R}^I = \mathbb{R}^{I_1 \times \dots \times I_d}$ or a corresponding tensor $v = \sum_{i \in I} a_i \psi_i$, when $\#I_\nu = O(n)$ for each ν, has a storage complexity

$$\#I = \#I_1 \dots \#I_d = O(n^d)$$

which grows exponentially with the dimension.

Manipulating tensors requires exploiting special properties of tensors such as sparsity or low rankness.

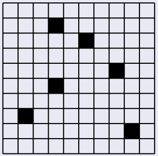
Outline

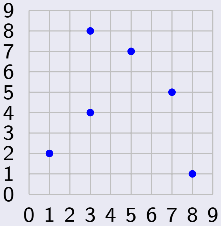
Sparse tensors

A tensor $a \in \mathbb{R}^I$ is said to be Λ-sparse, with Λ a subset of indices in I, if

$$a_i = 0 \quad \forall i \notin \Lambda.$$

A Λ-sparse tensor (left) and the set Λ (right)





A tensor

$$v = \sum_{i \in I} a_i \psi_i \in V$$

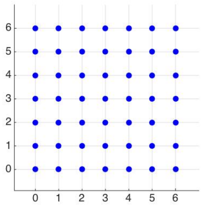
is said to be Λ-sparse relatively to the tensor product basis $\{\psi_i = \psi^{(1)}_{i_1} \otimes \ldots \otimes \psi^{(d)}_{i_d}\}_{i \in I}$ of V if the tensor $a = (a_i)_{i \in I}$ of its coefficients is Λ-sparse.

Structured subsets of multi-indices

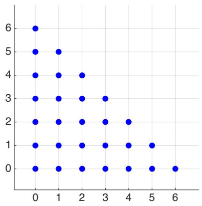
The set of multi-indices in $I = \mathbb{N}_0^d$ with p-norm bounded by m is

$$\Lambda = \{i \in I : \|i\|_p \leq m\}, \quad \|i\|_p = \begin{cases} \left(\sum_k i_k^p\right)^{1/p} & \text{for } 0 < p < \infty \\ \max_k i_k & \text{for } p = \infty \end{cases}$$

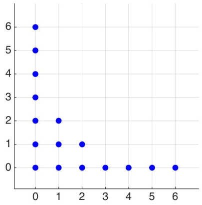
$m = 6,\ p = \infty$



$m = 6,\ p = 1$



$m = 6,\ p = 1/2$



Structured subsets of multi-indices

Example 3

Consider the tensor basis of monomials

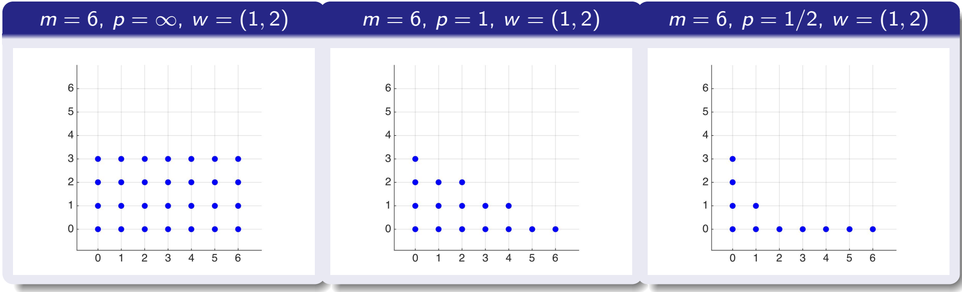
$$\psi_i(x) = x^i = x_1^{i_1} \dots x_d^{i_d}$$

- $\Lambda = \{i : \|i\|_\infty \leq m\}$ corresponds to the set of polynomials with partial degree less than m, with
$$\#\Lambda = (m+1)^d$$
- $\Lambda = \{i : \|i\|_1 \leq m\}$ corresponds to the set of polynomials with total degree less than m, with
$$\#\Lambda = \frac{(m+d)!}{m!d!}$$

Structured subsets of multi-indices

The set of multi-indices in $I = \mathbb{N}_0^d$ with weighted p-norm bounded by m is

$$\Lambda = \{i : \|i\|_{p,w} \le m\}, \quad \|i\|_{p,w} = \begin{cases} \left(\sum_k {w_k}^p i_k^p\right)^{1/p} & \text{for } 0 < p < \infty \\ \max_k w_k i_k & \text{for } p = \infty \end{cases}$$



The choice of anisotropic sets is the key for high-dimensional approximation.

Greedy algorithms

Standard greedy algorithms can be applied for sparse tensor approximation by considering as the dictionary $\mathcal{D}$ the tensor product basis

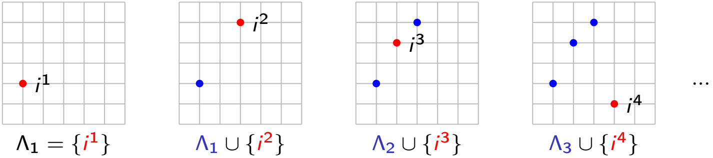
$$\mathcal{D} = \{\psi_i = \psi^{(1)}_{i_1} \otimes \ldots \otimes \psi^{(d)}_{i_d} : i \in I\}.$$

Greedy algorithms construct a sequence of approximations $(u_n)_{n\geq 1}$ such that

$$u_n \in \text{span}\{\psi_i : i \in \Lambda_n\}$$

where $(\Lambda_n)_{n\geq 1}$ is an increasing sequence of subsets such that

$$\Lambda_n = \Lambda_{n-1} \cup \{i^n\}, \quad i^n \in I.$$



Convex relaxation methods

Convex relaxation methods for sparse tensor approximation consists in solving problems of the form

$$\min_{a\in\mathbb{R}^I} J(a) + \lambda\Omega(a)$$

where J is the cost functional and where $\Omega(a)$ is a convex function which induces sparsity in the resulting tensor $a \in \mathbb{R}^I$.

Note that in practice, for the problem to be tractable, the tensor is constrained to be Λ-sparse for a given Λ with moderate cardinality.

Convex relaxation methods

The standard approach uses the 1-norm

$$\Omega(a) = \|a\|_1 = \sum_{i \in I} |a_i|,$$

which is the convex relaxation of the natural measure of sparsity $\|a\|_0 = \#\{i \in I : a_i \neq 0\}$.

A weighted version of the 1-norm can be defined as

$$\Omega(a) = \|a\|_{1,\omega} = \sum_{i \in \Lambda} \omega_i |a_i|,$$

where the weights can be deduced from a sequence of weights $\omega^{(\nu)}_{i_\nu}$ associated with univariate functions $\psi^{(\nu)}_{i_\nu}$ according to $\omega_i = \omega^{(1)}_{i_1} \dots \omega^{(d)}_{i_d}$.

Also, by introducing a partition $\{\Lambda_k\}_{k=1}^K$ of Λ (e.g. associated with a hierarchy of tensor spaces), we can consider

$$\Omega(a) = \sum_{k=1}^{K} \omega_k \|a_{\Lambda_k}\|_2$$

for inducing group sparsity.

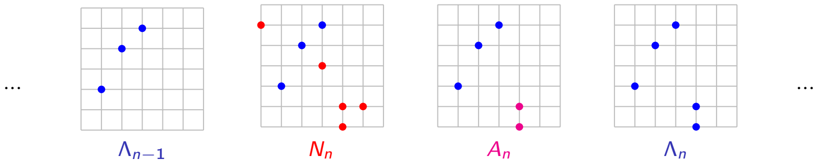
Working set algorithms

Working set algorithms for sparse tensor approximation construct an increasing sequence of subsets $(\Lambda_n)_{n\geq 1}$ in I and a sequence of approximations $u_n \in V_{\Lambda_n}$ computed through interpolation, regression or other projection methods.

The sequence of subsets is defined by

$$\Lambda_n = \Lambda_{n-1} \cup A_n$$

where A_n is a subset of a candidate set N_n.



The definition of N_n requires to define a strategy for the exploration of the set of multi-indices I.

Hierarchy of tensor spaces

A Λ-sparse tensor v is an element of the space

$$V_\Lambda = \text{span}\{\psi_i : i \in \Lambda\}.$$

A basis $\{\psi_i^\nu\}_{i \in I_\nu}$ in V_ν defines a hierarchy of spaces $H_m^\nu = \text{span}\{\psi_0^\nu, \ldots, \psi_m^\nu\}$ such that

$$H_0^\nu \subset \ldots \subset H_m^\nu \subset \ldots \tag{1}$$

For example, when considering canonical polynomial bases on $\mathcal{X}_\nu$, $H_m^\nu = \mathbb{P}_m(\mathcal{X}_\nu)$.

For $i \in I$, the product set

$$R_i = \{0, \ldots i_1\} \times \ldots \times \{0, \ldots, i_d\} = \{j \in I : j \leq i\}$$

defines a tensor space

$$H_i := V_{R_i} = H^1_{i_1} \otimes \ldots \otimes H^d_{i_d}$$

which only depends on the spaces $H^\nu_{i_\nu}$ and not on the choice of bases $\{\psi_0^\nu, \ldots, \psi_{i_\nu}^\nu\}$ for $H^\nu_{i_\nu}$.

For two multi-indices $i, j \in I$,

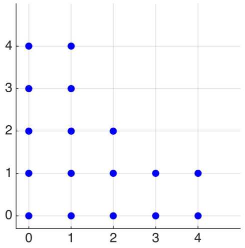
$$i \leq j \quad \text{implies} \quad R_i \subset R_j \quad \text{and} \quad H_i \subset H_j$$

Downward closed sets

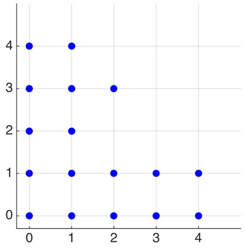
A set $\Lambda \subset I$ is downward closed if for $j \in I$,

$$(i \in \Lambda \text{ and } j \leq i) \quad \Rightarrow \quad j \in \Lambda$$

Downward closed



Not downward closed



Downward closed sets

For Λ downward closed, the space

$$V_\Lambda = \text{span}\{\psi_i : i \in \Lambda\}$$

only depends on the hierarchy of spaces H_m^ν and not on the particular choice of bases $\{\psi_0^\nu, \ldots, \psi_m^\nu\}$ for these spaces.

Example 4 (Polynomial spaces)

If $H_m^k = \mathbb{P}_m$ and Λ is downward closed, then

$$V_\Lambda = \text{span}\{x^i = x_1^{i_1} \ldots x_d^{i_d} : i \in \Lambda\}$$

whatever the choice of polynomial basis $\{\psi_0^k, \ldots, \psi_m^k\}$ of $\mathbb{P}_m$.

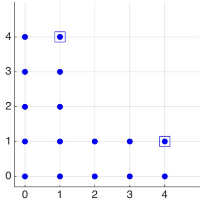
Downward closed sets

An element i of a set Λ is said maximal if and only if there is no $j \in \Lambda$ such that $i \leq j$ and $i \neq j$.

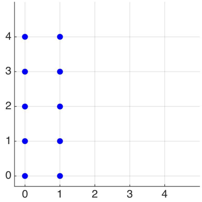
A downward closed set Λ is completely determined by the set of its maximal elements

$$\Lambda = \bigcup_{\substack{i \in \Lambda \\ i \text{ maximal}}} R_i.$$

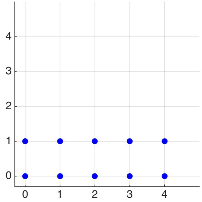
Λ



$R_{(1,4)}$



$R_{(4,1)}$



The resulting space V_Λ is such that

$$V_\Lambda = \sum_{\substack{i \in \Lambda \\ i \text{ maximal}}} H_i = \sum_{\substack{i \in \Lambda \\ i \text{ maximal}}} H^1_{i_1} \otimes \ldots \otimes H^d_{i_d}.$$

Margins

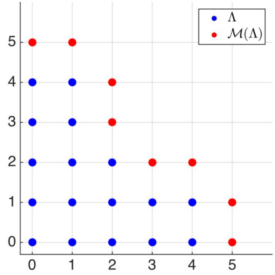
For a given set Λ, a natural neighborhood is the margin of Λ, defined by

$$\mathcal{M}(\Lambda) = \{i \in I \setminus \Lambda : \exists j \in \Lambda \text{ s.t. } \|i - j\|_1 = 1\}$$

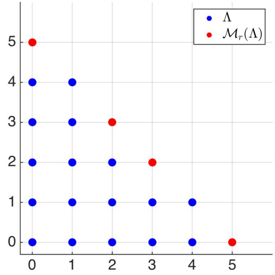
Also, we define the reduced margin of Λ as

$$\mathcal{M}_r(\Lambda) = \{i \in I \setminus \Lambda : i - e_k \in \Lambda \text{ for all } k \text{ s.t. } i_k > 1\}$$

A set Λ and its margin $\mathcal{M}(\Lambda)$



A set Λ and its reduced margin $\mathcal{M}_r(\Lambda)$



For a downward closed set Λ, an interesting property of the reduced margin $\mathcal{M}_r(\Lambda)$ is that for any subset $A \subset \mathcal{M}_r(\Lambda)$, $\Lambda \cup A$ is downward closed.

Outline

Sparse interpolation

Let $\{t_i^\nu\}_{i\geq 0}$ be a sequence of points such that $\Gamma_m^\nu = \{t_0^\nu, \dots, t_m^\nu\}$ is unisolvent for H_m^ν, that means for all $y \in \mathbb{R}^{m+1}$, there exists a unique function $v \in H_m^\nu$ such that $v(t_i^\nu) = y_i$ for all $0 \leq i \leq m$. Let $\{\ell_i^{\nu,m}(t)\}_{i=0}^m$ be the interpolation functions in H_m^ν.

We define the interpolation operator $\mathcal{I}_m^\nu$ which associates to a function v the unique function $\mathcal{I}_m^\nu(v) \in H_m^\nu$ such that $\mathcal{I}_m^\nu(v)(t_i^\nu) = v(t_i^\nu)$ for $0 \leq i \leq m$,

$$\mathcal{I}_m^\nu(v)(t) = \sum_{i=0}^m v(t_i^\nu)\ell_i^{\nu,m}(t).$$

The interpolation operator can be written as the sum of difference operators

$$\mathcal{I}_m^\nu = \sum_{i=0}^m \Delta_i^\nu$$

with

$$\Delta_i^\nu = \mathcal{I}_i^\nu - \mathcal{I}_{i-1}^\nu, \quad \mathcal{I}_0^\nu = 0.$$

Sparse interpolation

To a multi-index i corresponds a point

$$x_i = (t^1_{i_1}, \dots, t^d_{i_d}).$$

For the product set of indices $R_i = \{j \in I : j \leq i\}$, we define the tensor product grid

$$\Gamma_i = \{x_i : i \in R_i\},$$

which is unisolvent for the tensor space $V_{R_i} = H_i = H^1_{i_1} \otimes \dots \otimes H^d_{i_d}$, and the associated interpolation operator $\mathcal{I}_i = \mathcal{I}^1_{i_1} \otimes \dots \otimes \mathcal{I}^d_{i_d}$ such that

$$\mathcal{I}_i(v)(x) = \sum_{j_1=0}^{i_1} \dots \sum_{j_d=0}^{i_d} v(t^1_{j_1}, \dots, t^d_{j_d}) \ell^{1,i_1}_{j_1}(x_1) \dots \ell^{d,i_d}_{j_d}(x_d) := \sum_{j \in \Gamma_i} v(x_j) \ell^i_j(x).$$

$\mathcal{I}_i$ can be expressed as

$$\mathcal{I}_i = \sum_{j \leq i} \Delta_j$$

where

$$\Delta_j = \Delta^1_{j_1} \otimes \dots \otimes \Delta^d_{j_d}.$$

Sparse interpolation

The above construction can be generalized to an arbitrary downward closed set Λ, with the definition of an interpolation operator

$$\mathcal{I}_\Lambda = \sum_{i \in \Lambda} \Delta_i$$

onto the space V_Λ, associated with the sparse grid

$$\Gamma_\Lambda = \{x_i : i \in \Lambda\}.$$

The grid Γ_Λ is unisolvent for the space V_Λ.