Software

The FBA Solution Space Kernel – Introduction and Illustrative Examples

Wynand S. Verwoerd[1,*] and Longfei Mao[2]

[1]Centre for Advanced Computational Solutions, Faculty of Ag & Life Sciences, Lincoln University, Lincoln 4647, New Zealand.

[2]Department of Pharmacy, College of Biology, Hunan University, Changsha, China.

*To whom correspondence should be addressed.

Abstract

Background: The solution space of an FBA-based model of cellular metabolism, can be characterised by extraction of a bounded, low dimensional kernel (the SSK) that facilitates perceiving it as a geometric object in multidimensional flux space. The aim is to produce an amenable description, intermediate between the single feasible extreme flux of FBA, and the intractable proliferation of extreme modes in conventional solution space descriptions. Fluxes that remain fixed are separated off while the focus of interest is put on the subset of variable fluxes that have a nonzero but finite range of values. For unbounded fluxes, a finite subrange that geometrically corresponds to the variable flux range is determined and is supplemented by a limited set of rays that encapsulates their unbounded aspects. In this way the kernel emphasises the realistic range of flux variation allowed in the interconnected biochemical network by e.g. limited nutrient uptake, an optimised objective and other model constraints. This work builds on the full presentation of the SSK approach in a research monograph.

Results: It is demonstrated how knowledge of the SSK and accompanying rays can be exploited to explore representative flux states of the metabolic network. Noting that bioengineering interventions such as gene knockouts modify the solution space, new tools based on the SSK analysis are presented here that predict the effects of such interventions on a target flux constructed to represent a desired metabolic output. A simple metabolic model is used first to demonstrate the special concepts and constructions needed to define and compute the SSK. The demonstration model is tweaked to produce typical behaviours of larger models, but with kernels in 1,2 or 3 dimensions that are explicitly displayed to visualise the concepts. General applicability to models where visualisation is inaccessible, is illustrated by showing evaluation of potential bioengineering strategies for a genome scale model.

Conclusions: SSKernel is a flexible interactive tool that facilitates an overview of the FBA solution space as a multidimensional geometric object, in terms of a manageable number of parameters. It allows exploration of effects on this solution space from metabolic interventions, and can be used to investigate bioengineering strategies to manipulate cellular metabolism.

Keywords: Metabolic fluxes, Flux Balance Analysis, solution space reduction, metabolomics, bioengineering, biological networks, network analysis.

1 Background

A fundamental premise of the constraint-based modelling of cellular metabolism is that stoichiometric, thermodynamic, and capacity constraints limit the vector of flux values for biochemical reactions in a cell to a feasible region (the solution space) within the flux space created by taking each reaction flux as an independent axis.

Flux balance analysis (FBA) (1,2)takes this one step further, assuming that the actual metabolic state of a cell is well represented by the flux vector that optimises an appropriately chosen objective such as the biomass growth rate. With linear constraints and objectives, solving for an optimal flux value is a standard linear programming (LP) (3) problem.

However, even though realistic models remain extensively underdetermined even after an objective is fixed, LP only delivers a single optimal flux value. In contrast, extreme pathway or elementary mode analysis (4–6) gives an exact mathematical description of all feasible flux values by finding a convex vector basis that completely spans the solution space.

Neither of these approaches provide a readily utilizable description of the flux or range of fluxes available to a cell. In the case of FBA, this is because the single flux it yields cannot be taken as typical; for the simplex LP algorithm, it is, in fact by construction located at a

vertex, i.e., at the very extreme of the solution space. For extreme pathways, the problem is that the sheer number of basis vectors becomes overwhelmingly large for realistic metabolic models.

A popular way to address this difficulty is an elaboration of FBA called flux variability analysis (FVA) (7). This consists of using LP to find the minimum and maximum value that each reaction flux can achieve within the model constraints. This fixes a cuboid bounding box in the multidimensional flux space, within which the solution space lies. Numerical difficulties are often encountered in performing FVA, but a more fundamental limitation is that in high-dimensional spaces, the solution space polytope occupies a negligible fraction of the bounding box, rendering the FVA box uninformative. This point is further discussed below.

The solution space kernel (SSK) approach (8) steers midway between the extremes and pitfalls described. It expands on FBA by specifying the range of feasible flux values rather than a single value. It characterises this region in a way that is far more specific and informative than FVA is but in terms of far fewer parameters than the extreme pathway/elementary mode approach. In direct contrast to the FVA bounding box, every flux value that falls in the SSK is an actual feasible flux.

This goal is achieved by identifying a compact, low-dimensional subspace of the SS from which most, if not all, feasible fluxes can be reached by the addition of a limited number of ray vectors that are also identified. This subspace is termed the *kernel* of the SS. The full SS is unbounded in some directions for virtually all metabolic models, but even so, the kernel remains bounded. It is specified by a small number of constraints in far fewer dimensions than the full SS.

To enhance comprehension, the SSK is treated as a geometrical object in multidimensional space with a shape described by a set of mutually orthogonal maximal chords and resulting aspect ratios. Even more concretely, a list of explicit flux vectors that specify the center and chosen periphery points of the SSK allows the expected values and ranges for flux-dependent attributes of cell metabolism to be evaluated.

This paper clarifies these concepts by detailing their application to a simple metabolic network. No attempt is made here to explain the extensive computation that is needed to construct an SSK; more details are available in the work referenced in the next section.

However, in broad outline, the major stages are as follows:

1. All reaction fluxes that remain fixed over the entire SS are separated off.
2. Identify directions in flux space for which the SS is unbounded and separate them off by defining ray vectors.
3. The remaining SS kernel constraints and flux variables are obtained via vector projection, the introduction of capping constraints and other means.
4. The extent and shape of the SSK are delineated by finding a set of mutually orthogonal, maximal chords spanning the SSK.
5. The center and periphery of the compact kernel are determined, which further characterizes the SSK location, size and shape in a reduced-dimensional space.

Stage 5 allows an even more amenable approximation of the SSK for cases where it still has a sizable dimensionality *N*. A collection of representative periphery points is explicitly determined in this stage, and their convex hull delineates a central region that typically covers 80% or more of the full extent of the SSK. This is termed the peripheral point polytope (PPP) and has only in the order of 4 N explicitly listed vertices that can be used as a “poor man’s version” of extreme pathways to specify the most physically significant central part of the SSK.

The solution space (SS) concept needs further discussion. In constraint-based analysis, this is generally considered to be the flux subspace that satisfies stoichiometry and other external constraints. For FVA, the optimised objective is added as a further constraint, further reducing the subspace to what could be termed an objective subspace (OS). The distinction is ignored in most of what follows since kernel analysis can be applied to either the SS or OS. However, for most models, the upper limits for individual fluxes are unknown, and the SS is a simple unbounded hypercone whose apex at the flux space coordinate origin represents the zero-flux solution. In this case, only the apex point serves as the kernel since all other feasible fluxes can be reached from it by adding a suitable ray vector.

For this reason, the SSKernel analysis mostly comes into its own for the OS, which has a more complex shape. A linear objective defines a hyperplane, and the OS is the intersection between this hyperplane and the SS hypercone. This intersection generally retains some unbounded directions but also acquires bounded facets. These bounded facets embody the biologically significant flux constraints. Determining all feasible, bounded facets (FBF) is one of the major computational challenges of the SSK calculation. Once found, the kernel is constructed as the hypervolume bordered by all FBF’s, together with any additional capping constraints needed to partition it from the unbounded section of the OS.

This computational framework can also be exploited to characterise the dimensionality and shape of a simple cone SS more fully. To do this, its minimal single-point kernel is extended by capping the cone at a chosen radius. This allows the “cross-sectional” shape of the simple cone to be characterised by chords and aspect ratios in a similar way as is done for the OS.

In this sense, the SS kernel approach can be usefully applied to both the SS and OS, and to simplify terminology, both concepts are henceforth referred to as the solution space.

The brief introduction presented above merely touches upon the main SSK concepts, and the main purpose of this article is to further elucidate them by explicit, concrete examples in low dimensions that can be readily visualised. A more leisurely introduction to the solution space kernel approach may be found in the software manual included in the supplementary information attached to this article [See Additional File 3]. Also, a more extensive exposition of how this work relates to existing FBA approaches and literature, is provided in the first chapter of the monograph (8) in which the Solution Space Kernel (SSK) was first introduced.

2 Implementation

The computational processing is performed via the software package SSKernel (9), which is readily available as a GitHub release, and its Wolfram Language source code is also provided in the Supplementary Data [See Additional File 3]. This implements an interactive interface that allows the user to perform the major stages as outlined in the previous section, adjusting relevant parameters for each subsequent stage as needed. The software is accompanied by a user manual that describes the concepts and calculations in more detail than is possible here.

To process a metabolic model, it is specified in the format defined by the COBRA toolbox (10) as an *.mat file or alternatively in SBML format.

An interactive run of the SSKernel software produces two sets of outputs. First, there is a 2-page PDF file that summarises the main outcomes of the calculation, such as stage-by-stage dimension reduction and SSK shape information. Second, the full numerical results, including a constraint-based specification of the SSK, explicit ray vectors and peripheral flux points, are exported to an output file in the Data Interchange Format (*.dif). Examples of the input files and two output files for the toy model discussed in section 3 are provided in the supplementary data [See Additional File 2].

Also included in the SSK software are tools for exploring the solution space and how it is changed by bioengineering interventions. These are used independently of the interactive interface for postprocessing and take the abovementioned DIF file as input. The realistic network example discussed in section 3.2 is based on processing the DIF output file of an SSK run on the red blood cell model iAB_RBC_283 (Bordbar, Jamshidi, & Palsson, 2011).

The calculation of the SSK involves many special concepts and algorithms for dealing with polytopes in high-dimensional spaces, and a full exposition has been presented in a recent monograph (8). However, the exploratory tools are a subsequent development, and the formal details of the new algorithms they use are presented in the Supplementary Information [See Additional File 1].

3 Results and discussion

The results are presented in two main parts.

First, the main features of the SSK are illustrated on the basis of a simple model system. This is a toy model formulated in a 10-dimensional flux space and allows SSK reductions to 2 or 3 dimensions so that they can be graphically displayed. The emphasis here is on transparency. Such a simple model does not demonstrate the full power of simplification that SSK analysis is capable of. For that, the reader is referred to the dedicated monograph (8), which discusses applications to state-of-the-art models involving thousands of flux space dimensions.

Second, it is shown how the SSK and rays can be used to explore the solution space and its response to flux restrictions. The toy model illustrates the concepts, and an application to a realistic metabolic model is also shown to demonstrate its practical usefulness.

3.1 Elucidating SSK Concepts

The toy network used here originates from a textbook demonstration of extreme pathways (11) and is shown in Figure . Various slight elaborations of the toy model are made below to reproduce several scenarios commonly encountered in models of real cellular metabolism. Explicit input data sets to reproduce the results for all scenarios below are supplied in the Supplementary Information [See Additional File 2].

3.1.1 Scenario 1: A simple unbounded cone solution space

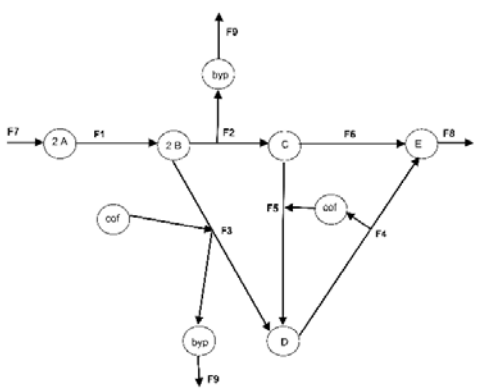

Figure 1. Toy network that produces a product *E* and byproduct *byp* from six internal reactions with fluxes F1 to F6. F7 is an external input flux and F8 and F9 are external output fluxes.

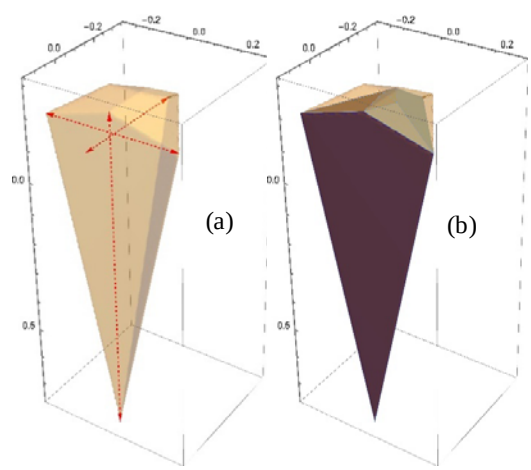

Figure 2. (a) The capped SSK for the toy model without objective. The main chords are shown as red arrows. (b) The periphery point polytope (PPP) superimposed on the SSK shows coverage of most but not all of its volume.

The bare-toy model is representative of realistic models that contain no objective. In the conventional description, the metabolic states of this network are superpositions of its three distinct extreme pathways (11), i.e., its SS is a 3-sided infinite cone in 9 flux dimensions.

The SSK analysis uses a default capping radius, as mentioned above, to reduce this to a finite kernel, which is explicitly specified by 6 constraints on 3 variables. The reduction from the 16 constraints on 9 variables of the original model merely hints at the large reductions in realistic models.

Figure 2 shows a 3D plot of the SSK. The SSK algorithm finds 3 peripheral rays aligned along the 3 long edges of the cone, each yielding one capping plane visible at the top in the figure. These ray directions are identical to the extreme pathways listed in Figure 13.1 of (11). However, that is generally not the case; for example, if the SS was a 5-sided cone in 3D, there would be 5 extreme pathways, but still only 3 peripheral rays, since they are constructed to be only a vectorially complete basis for the ray space.

The coordinate axes in Figure 2 represent the particular 3D subspace within the full 9D flux space, in which the model allows feasible metabolic fluxes. Specifically, they are by construction the directions of the main chords of the SSK, shown as red arrows in the figure. For the simple unbounded cone of this model, the arbitrary capping radius means that the actual chord lengths are not meaningful. However, their ratios are independent of the radius and characterise the shape of the cone. For this model, these aspect ratios are 1.2, 2.1 and 2.5, respectively. The lowest value reflects a fairly regular cross section, and the other values reflect a quite sharp cone, as shown in Figure 2.

3.1.2 Scenario 2: A fully bounded solution space

Next, the model is made slightly more realistic by introducing an upper limit for the input flux F7. FVA shows that this constrains all nine fluxes to finite, nonzero ranges. There are no fixed fluxes, and the FVA bounding box is a cuboid in 9 dimensions.

With this SS already bounded, no rays or capping hyperplanes need to be calculated, and the SSK is in fact identical to the SS for this case. The SSK shape analysis remains applicable, and the result is very similar to that shown in Figure 2; only the cone is now truncated by a single plane fixed by the upper limit on F7. Therefore, the SSK is, in this case, given by just 4 constraints on 3 variables.

Provided that the F7 cut-off value has been set at a physically meaningful value, the 3 main chords indicate the range of metabolic flux variation allowed by the model along the 3 available orthogonal directions. Note how this is far more specific than the 9D FVA cuboid within which this truncated cone is contained.

The aspect ratios derived from the chords are 1.5, 1.8 and 2.8, indicating a similarly sharp cone but with a slightly flattened cross section compared with Figure 2.

3.1.3 Scenario 3: Optimised objective in a bounded SS

Moving on to a case more appropriate for the purpose for which the SSK analysis is designed, we introduce an optimised objective into the model while keeping the input flux limitation. A biomass composition of 80% of metabolite *E* and 20% of the byproduct *byp* is assumed and incorporated into the model by adding a further output flux F10 that exports these metabolites in a 4:1 ratio. The objective is taken as maximising F10.

An FVA calculation to characterise the OS with F10 kept fixed to its optimised value, shows that a total of 5 fluxes become fixed, leaving a 5D cuboid as the bounding box for feasible metabolic states.

The OS is now bounded, so no rays or capping is needed, and the SSK is identical to the OS. The calculation algorithm shows that the SSK/OS is only 2-dimensional and is plotted in Figure 3.

Clearly, this SSK is simply the intersection of the objective plane with the triangular cone of Figure 2, and as expected, the aspect ratio of 1.4 is similar to that of the cone cross section.

The figure also illustrates the outcome of the centering procedure incorporated in the SSK calculation. The center is the green dot, which also serves as the coordinate origin in the SSK specification of 3 constraints on 2 variables. It also finds a semi-random series of periphery points shown as black dots. These points form a plausible sample of explicit feasible fluxes that encapsulates the range of variation of metabolic states allowed by the model. Their convex hull (the PPP) covers 100% of the SSK, but as shown in Figure 2(b), this is not generally the case.

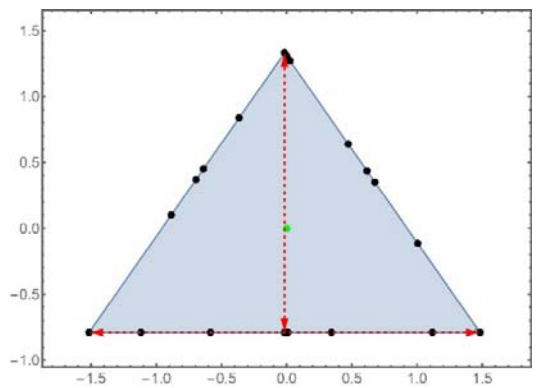

Figure 3. The OS and SSK for the toy model with a restricted input flux and a fixed objective. A centering algorithm determines the green center point and a semirandom selection of periphery points shown in black. The main chords are shown as red arrows.

3.1.4 Scenario 4: Comparing the SSK and FVA

In the first 3 scenarios, the FVA bounding box is in each case in a much higher-dimensional space than the SSK is, and cannot be drawn. To illuminate the relationship between the two constructs, we further restrict the model used in scenario 3 by keeping also F2 at a fixed value. This fixes F3 as well, reducing the bounding box from 5 to 3 dimensions. Removal of the F2 degree of freedom also reduces the SSK to one dimension. The results are shown in Figure 4.

Unlike the previous figures, for this case, the SSK subspace is shown relative to the 3D subspace that contains the variable fluxes F4, F5 and F6 which are taken as its coordinate axes. The FVA bounding box accurately depicts the range of variation along each axis, but provides a very poor rendering of the actual extent of feasible flux solutions. Only the SSK shows this exactly, namely, as just a single line, the diagonal of the cuboid box. This only occupies a negligible fraction of the FVA box volume; thus, random sampling from the cuboid is unlikely to reveal a single feasible flux combination.

Such dimensional discrepancy between the bounding cuboid and the SSK was also noted in the previous scenarios and is a universal feature of realistic metabolic models. For those the FVA overestimation is far worse, because even where the bounding box occupies the same dimensions as the SSK, the fraction of the box that the SSK occupies is still negligible. This is because of its typical high aspect ratio and the scaling effects in high dimensions mentioned in the introduction and discussed in more detail in (8).

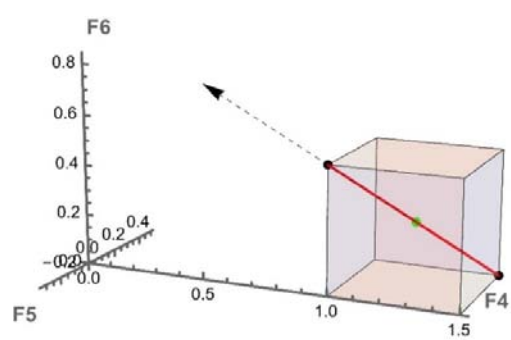

Figure 4. The 1D linear SSK (red line) and 3D FVA bounding box (in colour) for the toy model, with an optimised objective but constrained to leave only F4, F5 and F6 variable, while other fluxes assume fixed nonzero values.

In the discussion thus far, the FVA limits have been used as a familiar indication of flux ranges to discover fixed fluxes, but we note that the Hop, Skip and Jump algorithm incorporated in the SSK calculation performs this task faster by orders of magnitude and also more reliably. These advantages are not manifested in the small toy model.

3.1.5 Scenario 5: SS partially bounded; tangent capping.

The next scenario is chosen to resemble a realistic solution space as closely as possible within the size limitations of the toy model.

This is achieved by putting a lower limit on F2 and allowing F3 to be reversible. The same objective and input flux limitations as in scenario 3 are used. There are again 5 fixed fluxes and 5 variable fluxes, and the feasible solutions are restricted to a 2D plane within this 5D subspace. Unlike the previous scenarios, this *reduced solution space* or RSS (reduced by the exclusion of fixed fluxes) is no longer bounded. Therefore, there are ray directions in the solution space—in fact, only a single ray—the green arrow shown in Figure .

Despite the RSS being unbounded, some of its boundary facets remain bounded, emphasised as thick lines in the figure. In addition to the 2 bounded facets (FBF's), there are also 2 unbounded facets. To achieve a bounded kernel, the SSK procedure introduces a single *capping hyperplane*, the dotted line BC, that partitions the RSS.

Geometrically, any flux value in the unbounded section can be reached from at least one flux point in the bounded section (the SSK) by adding a multiple of the ray vector.

This is guaranteed by having chosen the capping line such that it does not bisect any of the bounded facets. The capping radius (the perpendicular distance from the origin to the capping line) is chosen to allow the capping line to intercept any bounded facet at most at a single point. This procedure is denoted as *tangential capping*.

To grasp the biochemical meaning of the capping, note that the ray vector in fact represents the single cycle created in the network of Figure when reaction 3 was made reversible. This is shown by the calculated 10D ray vector {0, 0.57735, -0.57735, 0, 0.57735, 0, 0, 0, 0, 0}, i.e., only fluxes F2, F3 and F5 participate equally.

This cycle has no net consumption/production of metabolites, so it can support arbitrarily large flux values without violating limits such as the maximum uptake imposed in the scenario under discussion. This means that partitioning off the unbounded section of the RSS by capping simply excludes nonphysical cyclical fluxes. The remaining solutions that define the SSK embody the physically meaningful flux variation as limited by the model constraints.

Note that this partitioning is not unique. The line BD would be an alternative (and minimal) capping line. The key feature is that a viable SSK needs to include all bounded facets (FBF's), which is why FBF's are considered to encompass the physically meaningful constraints that result from the way that individual flux constraints combine across a network. In higher dimensions, minimal tangent capping is hard to compute, so merely capping all ray directions orthogonally (as line BC does in this example) is considered a sufficient strategy.

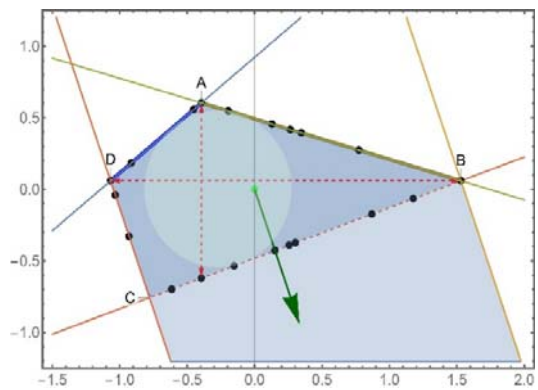

Figure 5. The RSS (light blue shading) and SSK (darker shading) for the toy model with a single reversible reaction. The RSS has two bounded facets, AB and AD. The other two parallel facets remain unbounded and leaves a single ray direction (green arrow). The SSK is bounded by the capping line BC, perpendicular to the ray and only intersecting the bounded facets tangentially at the single point B. The inscribed circle is shown in a lighter shade, but the (green) centroid of the chosen periphery points (black dots) is taken as the coordinate origin and the axes are aligned with the main chords shown as

3.1.6 Scenario 6: Ray space and coincidence capping.

This case is similar to scenario 5, but reaction 6 is now also allowed to be reversed. The result, shown in Figure 6, is that only the single bounded facet AB of scenario 5 survives. The RSS remains 2D, but there are now two peripheral ray directions aligned with the two unbounded facets. Intermediate rays can be formed by a linear combination of these two rays, i.e., they span the *ray space*.

It would be possible to define a 2D SSK for this case by taking a capping line perpendicular to each peripheral ray to form a triangle with AB on one side. However, a better alternative is to note that there exists a ray that is perpendicular to line AB. Suppose that this ray is used to define a capping constraint, resulting in a capping line parallel to AB. The capping radius is adjusted until the capping line first intersects the FBF (i.e., line AB itself), which occurs only when the two lines coincide. At that point, the entire shaded region of the RSS is separated off, and only the single, one-dimensional SSK defined by line AB remains.

This is an example of *coincidence capping*. In contrast to tangent capping, it gives an SSK of lower dimensions than the RSS.

With respect to the biochemical interpretation, inspection of the two 10D numerical ray vectors reveals that they represent fluxes that involve (F2, F3, F5) and (F4, F5, F6) respectively. . These paths are easily recognizable from Figure as overlapping but independent cyclical paths formed when F3 and F6 are reversed.

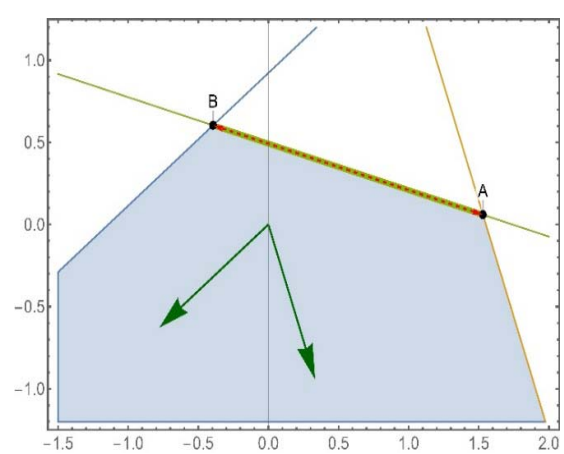

Figure 6. The toy model RSS (unbounded 2D shaded region) and 1D SSK (the line AB) for scenario 6. There are two peripheral ray directions shown as green arrows. The coordinate system of scenario 5 is chosen to facilitate comparison.

Rays are, however, not necessarily associated with cycles as happens for the examples shown; e.g., if the toy network contained a linear network path that does not involve the constrained uptake flux F2 or the metabolites that make up the objective, this might also have no upper limit on its flux and thus constitute a ray.

3.1.7 Scenario 7: Coincidence capping of prismatic rays

The RSSs in scenarios 5 and 6 are examples of facetted cones, and the rays are described as *conical rays*. However, for the SSK calculation, several additional types of rays are distinguished (8), and the most common type, typically approximately half of all rays for realistic metabolic models, are *prismatic rays*. This type can be induced into the toy model by inserting an additional unidirectional reaction, which involves only some newly added metabolites, into the network. This would add an additional flux axis F11 to the flux space. If the same constraints are used as in scenario 3, the OS shown in Figure 3 is modified by the addition of a vertical axis perpendicular to the plane of the paper. The assumed direction of the reaction limits its feasible fluxes to positive values. Therefore, instead of the 2D triangle shown in the figure, the OS is now a semi-infinite triangular prism that stretches vertically to positive infinity and is bordered by one horizontal constraint plane (the plane of the paper, which constrains F11 to positive values) and three vertical constraint planes.

This modification adds a single ray direction to the OS, which is clearly described as a *prismatic ray*. The distinguishing feature of a prismatic ray is that it is antiparallel to one constraint vector (the one defining the horizontal plane in this example) and orthogonal to all others. Prismatic rays are particularly easy to deal with, as they can be coincidence capped. In this case, that reduces the 3D OS back to the 2D triangular SSK shown in Figure 3.

Similarly, if the constraints of scenarios 5 or 6 are applied to the 11D extension of the toy model, the RSSs in both cases become three-dimensional semi-infinite prisms, which are partially unbounded in both the horizontal and vertical directions. They each acquire an additional prismatic ray as a result. Once more, this can be coincidence capped, and then tangent or coincidence capping in the horizontal plane respectively produces the same 2D or 1D SSK's as illustrated in Figures 5 and 6, respectively.

3.1.8 Discussion.

The toy model scenarios that were sketched reproduce in isolation some of the behaviours encountered in full-scale metabolic models. However, in a realistic model, multiple instances are usually combined, such as the simultaneous occurrence of multiple ray types and capping methods along different dimensions.

Moreover, in a realistic higher-dimensional RSS, there are hierarchies of facet levels, similar to the boundary planes, edges and vertices of a 3D polytope.

For realistic models, the boundaries formed by the original model constraint hyperplanes are invariably unbounded, and bounded facets occur only at much higher facet levels. This requires an extensive computational framework for their detection and utilisation to construct appropriate tangent capping. An exponential proliferation of facets at higher levels makes this challenging, but it turns out that with appropriate algorithms, it is possible to reduce most models to a fairly low-dimensional SSK with a moderately sized accompanying set of ray vectors that suffice to specify the FBA solution space. That is, for example, the case for the red blood cell model discussed in the next section.

3.2 Exploring and Manipulating the SS.

Knowledge of the SSK provides insight into the location, extent and shape of the solution space and can plausibly be exploited to direct cellular metabolism along desired channels. This section introduces some new computational tools for that and demonstrates their use.

Although realistic models are not amenable to visualisation, the SSK periphery point calculation yields a representative sample of feasible metabolic states that similarly serves to display the range of variation allowed by the model.

The key idea pursued here is to resample a particular subregion of the SS that corresponds to a metabolic intervention of interest, such as manipulating nutrient availability, gene regulation, etc. Inspection of the resulting new sample yields information about the cellular responses predicted by the model.

Resampling a section of the SSK could be done by mere interpolation of the SSK periphery point set, but the *HyperSampler* function provided in the SSKernel software goes beyond that by also using the ray basis to extrapolate sample points located in SS regions outside the SSK. Execution of the combined interpolation and extrapolation is not entirely straightforward, and details of the algorithm used are provided in the Supplementary Information [See Additional File 1].

Given a target metabolite or flux on which some intended bioengineering effort focusses, a further function, *TargetVariation*, tabulates the mean value and range of variation of the target, observed in the new sample, for each of a chosen list of interventions such as gene knockouts. Examples are shown and interpreted below.

3.2.1 Toy model with *cof* as a target metabolite

Consider the toy model modified in scenario 5 above, and for illustration, suppose that we wish to influence the production of the metabolite *cof* (see Figure) via an external intervention. Two tweaks of the model are needed. First, in the model as stated, the stoichiometry guarantees net zero production of *cof* (as appropriate for its intended role as a cofactor). Increasing the coefficient of the metabolite for reactions 3 and 4 from 1.0 to 2.0 results in excess production. Second, a sink reaction to absorb this excess is added to the model, and its flux is defined as the target.

The effects on the target flux of various interventions, which were chosen to reflect typical cases and include some single gene knockouts, are shown in Table 1.

For comparison, the first line shows values for the SSK.

Starting at the bottom of the table, blocking the input flux F7 makes the model infeasible, as indicated by 0 sample points. Blocking F5 allows a range of feasible points, but the target flux remains fixed to zero. Fixing F3 and F4 to 0.5 units each restricts feasibility to a single flux point still with a zero target flux value. In contrast, blocking F8 has no discernible effect: the sample count and target flux values remain the same as those for the SSK. Fixing F2 to 1.0 produces almost as many sample points, but now, the maximum target flux is reduced to a finite value while the mean value is increased. Blocking either F6 or F3 respectively restricts the lower and upper target flux limits, increasing the mean value in both cases. The strongest boosting effect is observed by simultaneously blocking F3 and restraining F5 to a range of values above 0.8 units.

ASSIGN		SAMPLE			
Flux	Value	Count	Mean	Lower	Upper
SSK	---	23	0.265	0	∞
F3, F5	0, ≥0.8	12	1.05	0.8	1.5
F3	0.	18	0.869	0	1.5
F6	0.	5	0.621	0.5	∞
F2	1.	20	0.396	0.	1.0
F8	0.	23	0.265	0.	∞
F3, F4	0.5, 0.5	1	0.	0.	0.
F5	0.	9	0.	0.	0.
F7	0.	0	---	---	---

Table 1. Target flux variation in response to flux assignments

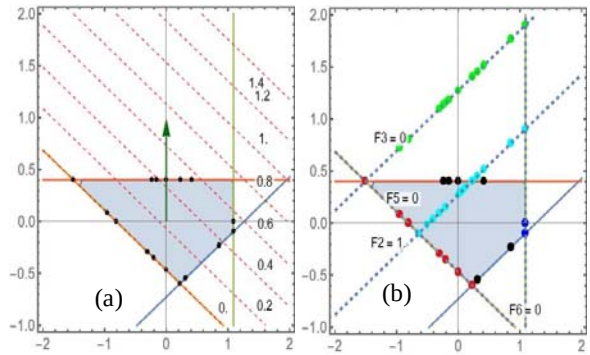

Figure 7: a) SSK with sampled periphery points shown as black dots, and contours of fixed target flux values as dashed lines. b) Fixed flux lines (dotted) and resampled points for some flux assignments: F3 = 0. (green), F2 = 1. (cyan), F5 = 0. (red) and F6 = 0 (blue).

All of these behaviors are easily interpreted by graphically displaying the SS and various flux contours, as shown in Figure 7. The SSK is the same as that in Figure and is only reoriented for better display, and the unbounded SS now stretches to infinity along the vertical direction. The actual sample points generated are shown as coloured dots for a few of the single fixed flux cases included in the table. Note how these extend beyond the SSK where appropriate and cover the full range of variation fairly uniformly.

The case of assigning a value range is not explicitly shown, but for example, resampling F2 ≤ 1 produces a sample corresponding to the cyan, blue and some of the red points, in addition to some points on the lower right SSK boundary. The points represent all finite SS boundaries but may also include a few interior points inherited from the SSK sample.

Comparing a fixed flux line in Figure 7b with the target flux contours in Figure 7a, the value range shown in Table 1 for the corresponding case is easily reconciled. Such graphic interpretation is not possible for realistic models, but the table still conveys the relevant information.

3.2.2 Resampling a realistic model

The red blood cell (RBC) metabolic model (12) included as a demonstration example with the SSKernel software is used to illustrate that resampling is also viable for realistic models. This model has 469 fluxes, of which 408 are found to remain fixed, leaving 61 fluxes that vary across the SS but are restricted to an 8-dimensional bounded SSK with 4 ray directions.

Simply inspecting points sampled from the SS can reveal aspects of metabolism. For example, the flux for the water transport sink reaction EX_h2O_e varies between -6.3 and 4.8 units, allowing both the import and export of water. What would happen if either of these is ruled out or if no net water transport is allowed?

This is investigated by resampling from the subregions where for that particular reaction, either F ≤ 0, ≥ 0 or the intersection of hyperplane F = 0 with the SS. In the latter case, for example, all changes in the remaining mean fluxes from their values over the SSK are plotted in Figure 8. There is a large contrast in the observed responses. For reference, the seven reactions where the magnitude of the change exceeded 0.1 are labelled EX_ade_e, EX_adn_e, EX_bilglcur_e, EX_fum_e, EX_h2o2_e, EX_hco3_e and EX_man_e. The biochemical significance of this observation, if any, could perhaps be further explored by sampling from SS subregions that focus attention on subsets of these particular fluxes.

Figure 8: Mean value changes from SSK values for the variable fluxes, in response to requiring net zero water exchange in RBC model.

3.2.3 Maximising RBC oxygen production

To show an example of how choosing a target metabolite could suggest a bioengineering target, consider that the function of the red blood cell is to transport oxygen. Is it possible to enhance this function by knocking out any single gene? To address that question, we identify the oxygen export reaction EX_O2_e in the existing model as the target flux. The broader SSK analysis discussed above has already narrowed the field from 469 fluxes to 61 variable fluxes, for which variability is compatible with the RBC model constraints.

The *TargetVariation* function illustrated above for the toy model provides a table of predicted outcomes for a systematic trial of each of the 61 knockouts. An extract of the results is shown in Table 2.

There are again some cases in which an infeasible model is produced or the target flux is restricted to zero. Most knockouts leave the range of the target flux unchanged while either increasing or reducing the mean target flux relative to the SSK value. Two cases stand out, however: knocking out either the H2Ot or EX-H2o_e reaction restricts the target EX_O2 flux to a fairly narrow, positive value range. Although there are other knockouts, such as ADEt, that produce a higher average value, this is arguably a more potent intervention since it guarantees a substantial oxygen export flux.

This result is not claimed to be a profound biochemical insight but clearly demonstrates that interrogation of the SSK using resampling tools can be useful for identifying bioengineering interventions for further study.

ASSIGN		SAMPLE			
Flux	Value	Count	Mean	Lower	Upper
SSK	---	64	2.61	0.	5.
ADEt	0.	64	3.13	0.	5.
UREAt	0.	10	2.77	0.	5.
SBTR_D2	0.	14	2.76	0.	5.
H2Ot	0.	64	2.75	2.32	3.14
EX_H2O_e	0.	64	2.75	2.32	3.14
NH4t3r	0.	17	2.68	0.	5.
PYRt2	0.	5	2.21	0.	5.
O2t	0.	2	0.	0.	0.
L_LACt2r	0.	0	---	---	---

Table 2. Red blood cell oxygen export flux for selected single-gene knockouts.

For this realistic but still rather small model, it is practical to use a brute force approach of testing all available gene knockouts since less than 10 s of computing time was required to trial all 61 cases. For larger and more complex models, this may well become intractable. However, since the method merely requires a list of candidate interventions as input, the application of biochemical knowledge to set up such a list should be able to reduce the task to manageable proportions even for large models.

4 Conclusions

An understanding of the range of metabolic states available to the metabolism of a cell, as delineated by available FBA metabolic models, is a powerful tool to design bioengineering interventions that produce desired metabolic outcomes. Equally such understanding can contribute to explain observed effects of the cellular environment, such as nutrient availability, on the metabolism.

Computing the solution space kernel (SSK) facilitates such understanding, by reducing the vast complexity induced by the millions of extreme pathways needed to give an exact description of the solution space of genome-scale metabolic models. At the same time, it is far more informative than the reduction to a cuboid bounding box produced by flux variability analysis (FVA).

From a solution space dimension count of thousands, the SSK reduction focusses attention on a kernel subspace with a dimension count usually well below 100 and characterized by a correspondingly small set of constraint equations. It allows geometrical perception of this kernel by explicit determination of its centre and an orthogonal set of chords. At an even more practical level, a semi-random but representative sample of feasible fluxes are produced that can be used for statistical analysis or as the basis for experimenting with the effects of imposing chosen genetic regulation or other manipulations of the metabolism.

Availability and Requirements

Project name: SolutionSpaceKernel
Project home page: https://github.com/wynand-verwoerd
Current Release: doi:10.5281/zenodo.11483838
Operating system(s): Platform independent, provided Mathematica implementation available.
Programming language: Wolfram Language
Other requirements: Wolfram Mathematica or Wolfram Engine
License: GNU GPL 3 .
Any restrictions to use by non-academics: None

Computing environment: Full utilization of the SSKernel software including its interactive control interface, requires installation of the proprietary *Mathematica* system. Work is in progress to produce a stand-alone version of the software and make this available in due course as part of the GitHub release.

However, for reproducing the results in Section 3, a command line implementation is perfectly adequate and for that only the installation of the Wolfram Engine is required. This can be downloaded for free, though subject to use restrictions that do allow testing applications. It is also available for online use via a Wolfram cloud server.

Abbreviations

DIF – Digital Interchange Format
FBA – Flux Balance Analysis
FBF – Feasible Bounded Facet
FVA – Flux Variability Analysis
LP – Linear Programming
OS – Objective Space
PPP – Periphery Point Polytope
SBML – Systems Biology Markup Language
SS – Solution Space
SSK – Solution Space Kernel

5 Declarations

Ethics approval and consent to participate: Not Applicable
Consent for publication: Not Applicable
Availability of data and materials: Input data sets are explicitly listed in the Supplementary Information.
Competing interests: Not Applicable
Funding: This work was supported by the provision of publication charges by Hunan University and is gratefully acknowledged by the authors.
Authors' contributions: WSV (Conceptualization, Methodology, Formal Analysis, Software, Writing – original draft preparation), LM (Conceptualization, Validation, Funding Acquisition, Writing – review and editing,)
Acknowledgements: Not Applicable

References

1. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nat Biotechnol. 2010;28(3):245–8.
2. Palsson BO. Systems biology: constraint based reconstruction and analysis. Cambridge, UK: Cambridge University Press; 2015. 550 p.
3. Jeter MW. Mathematical Programming□: an Introduction to Optimization. 1 st. New York: Taylor & Francis, Routledge; 1986. 360 p.
4. Schuster S, Dandekar T, Fell DA. Detection of elementary flux modes in biochemical networks: A promising tool for pathway analysis and metabolic engineering. Vol. 17, Trends in Biotechnology. Elsevier Ltd; 1999. p. 53–60.
5. Papin J, Stelling J, Price N, Klamt S, Schuster S, Palsson BØ. Comparison of network-based pathway analysis methods. Trends Biotechnol. 2004;22(8):400–5.
6. Schilling CH, Letscher D, Palsson BO. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. J Theor Biol. 2000 Apr 7;203(3):229–48.
7. Mahadevan R, Schilling CH. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng. 2003;5(4):264–76.

8. Verwoerd WS. The Compact Kernel of a Metabolic Flux Balance Solution Space. Singapore: World Scientific; 2022. 250 p.
9. Verwoerd WS. SSKernel; Mathematica implementation. GitHub.com; 2024.
10. Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc. 2011;6(9):1290–307.
11. Palsson BØ. Systems biology properties reconstructed networks. Cambridge, UK: Cambridge University Press; 2006. 336 p.
12. Bordbar A, Jamshidi N, Palsson BØ. iAB-RBC-283: A proteomically derived knowledge-base of erythrocyte metabolism that can be used to simulate its physiological and patho-physiological states. BMC Syst Biol. 2011 Jul 12;5.

Supplementary Information

The online version contains supplementary material in the form of three additional files, as follows:

- Additional File 1.docx: The mathematical formulation and computational procedure to resample points in a subregion of the solutions space, for use in exploring metabolic interventions.
- Additional File 2.docx: Explicit input data sets for producing the various Toy Model scenarios described in Section 3 of the article.
- Additional File 3.zip: The compressed file directory SSKernel Implementation , which contains the current release of the SSKernel software. This includes a 55 page software manual, all source code and test data.